The effects of computer-elicited structural and group knowledge on complex problem solving performance

An application of two computer-based tools for knowledge elicitation

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Psychologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Herr Dipl.-Psych. Bertolt Meyer
geboren am 04.04.1977 in Hamburg

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Wolfgang Coy

Gutachter:

1. Prof. Dr. Wolfgang Scholl
2. Prof. Oliver Günther, Ph.D
3. Prof. Dr. Hans-Jürgen Rothe

eingereicht am:	14. Dezember 2007
Tag der mündlichen Prüfung:	16. Mai 2008

Abstract

This thesis analyzes the influence of structural knowledge on the individual level and the influence of knowledge heterogeneity on the group level on complex problem solving (CPS) performance. For the elicitation of structural knowledge, a computer based method, the association structure test (AST), is developed. Through term associations, measurement of thinking times, and through pairwise concept comparisons, the AST elicits a knowledge graph that can be described with graph-theoretical coefficients. The AST is put to test in the domain of CPS.

As complex problems are usually addressed by groups in practice, a group setting is chosen. Scholl's model of team effectiveness postulates a curvilinear n-shaped connection between the group's knowledge heterogeneity and its CPS performance.

In a laboratory experiment, 150 participants were divided into dyads. The participants each received an instructional text with seven knowledge elements on the control of a complex scenario. The overlap of knowledge elements inside a dyad was varied (small overlap, partial overlap, full overlap). After learning, dyad members self-assessed their knowledge. This assessment entered was fed into a computer-based knowledge management system (the skillMap), for which an algorithm for knowledge similarity calculation had been devised. The knowledge similarity was also used for performance prediction. A discussion followed, during which dyad members taught each other what they had learned. Their structural knowledge was then assessed with the AST. In the following complex scenario task, dyads from the partially shared condition exhibited a significant superior performance in comparison with the other two conditions. Knowledge heterogeneity exhibited a curvilinear relationship with the dyad's CPS performance. The weighted density of AST-elicited knowledge graphs weakly correlated with CPS performance and explained a small but unique fraction of its variance. The skillMap similarity measure correlated significantly with CPS performance.

Computer-based knowledge elicitation tools are thus potentially suited for performance prediction. CPS performance of groups is partially determined by the way in which knowledge is distributed inside the group.

Keywords:
knowledge, complex problem solving, group performance, structural knowledge

Zusammenfassung

Die Arbeit untersucht den Einfluss von strukturellem Wissen auf individueller Ebene sowie den Einfluss der Wissensheterogenität auf Gruppenebene auf komplexes Problemlösen. Zur Erhebung von strukturellem Wissen wird ein computerbasiertes Verfahren, der Assoziations- Strukturtest (AST), entwickelt. Für die Berechnung der Wissensheterogenität in Gruppen wird das Wissensmanagementsystem skillMap um einen entsprechenden Algorithmus erweitert. Der AST erhebt mittels freier Assoziation, Denkzeitmessung und Paarvergleichen zu einer Wissensdomäne einen Graphen, der mit graphentheoretischen Kennwerten beschreiben wird. Für Leistungsvorhersagen mit dem AST werden komplexe Problemlöseaufgaben gewählt. Da komplexe Probleme i.d.R. von Gruppen bearbeitet werden, wird ein Gruppensetting gewählt. Scholls Modell der Teameffektivität postuliert einen umgekehrt-u-förmigen Zusammenhang zwischen der Wissensheterogenität in der Gruppe und ihrer Problemlöseleistung. In einem Laborexperiment wurden 150 Versuchspersonen in Dyaden eingeteilt. Die Teilnehmer erhielten je einen Lerntext mit sieben Wissenselementen zur Steuerung eines komplexen Problemlöseszenarios. Die Heterogenität des Wissens der beiden Gruppenmitglieder wurde variiert (geringe Heterogenität, mittlere Heterogenität, große Heterogenität). Nach der Lernphase gaben die Dyadenmitglieder eine Einschätzung über ihr Wissen ab. Diese Einschätzungen wurden mittels des skillMap-Algorithmus zu einem Ähnlichkeitskoeffizienten verrechnet, der auch zur Vorhersage der Gruppenleistung genutzt wird. Es folgte eine Diskussionsphase, in der die Dyadenmitglieder sich gegenseitig das Erlernte beibrachten. Ihr strukturelles Wissen wurde dann mit dem AST getestet. In der anschließenden Bearbeitung des Szenarios waren die Dyaden mit mittlerer Heterogenität den beiden anderen signifikant überlegen: Die Wissensheterogenität zeigt einen kurvenlinearen Zusammenhang mit der Gruppenleistung. Die gewichtete Dichte der AST-erhobenen Wissensgraphen korreliert gering mit der Problemlöseleistung und erklärt einen eigenständigen kleinen Anteil ihrer Gesamtvarianz. Das Ähnlichkeitsmaß korreliert signifikant mit der Problemlöseleistung. Computerbasierte Wissensdiagnoseverfahren sind somit potentiell dazu geeignet, Leistungsvorhersagen zu treffen. Die Problemlöseleistung von Gruppen ist zum Teil durch die Verteilung des Wissens innerhalb der Gruppe determiniert.

Schlagwörter:
Wissen, Komplexes Problemlösen, Gruppenleistung, Strukturelles Wissen

For Daniel

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The May 2007 issue of e.conomy, the economics magazine published by the German government, is entitled "Using knowledge as a resource" (Presse- und Informationsamt der Bundesregierung, 2007). The article implies that for countries lacking substantial natural resources such as Germany, knowledge is the most important resource, because it fuels innovation and economic success. Similar claims have been put forward by several scholars (Davenport & Prusak, 1998; Drucker, 1993; Nonaka & Takeuchi, 1995; Scholl, 2004). Arguing along the same lines, Stehr (1994) has coined the term 'knowledge society' for modern post-capitalist societies. The significance of knowledge as a vital resource for the world's economies has also been underlined by the European Union in the Lisbon Process (European Council, 2000).

The importance being accredited to the construct of knowledge and its role in the economy of the 21st century is contrasted by the blurred conceptualizations of the term: The field of knowledge management, defined as "the process of continuously creating new knowledge, disseminating it widely through the organization, and embodying it quickly in new products/services, technologies and systems" (Takeuchi, Nonaka, & Hitotsubashi-Daigaku, 2004, p. ix), is unstructured and scattered; Despres and Chauvel (2000) refer to it as a "patchwork" (p. 57). The concept of tacit knowledge (Polanyi, 1966) – credited with a key role in organizational performance (Nonaka & Konno, 1998; Takeuchi et al., 2004) – is one of the most blurred concepts in management literature (Busch, Richards, & Dampney, 2001) and there is an argument whether Polanyi, who coined the concept of tacit knowledge and Nonaka, who introduced it into knowledge management (Nonaka & Takeuchi, 1995), are referring to the same thing (Li & Gao, 2003).

This uncertainty with regard to the definition of the construct is contrasted by the agreement that "an increased focus on the handling of non-explicit knowledge might pose a considerable value-creating factor" (Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW), 2001, p.7, own translation).

In order to close the gap between the assumed importance of the construct on one side and its ambiguity on the other, a thorough investigation of the concept of knowledge and its different forms and the development of a sound conceptual framework appear necessary. Such a framework should allow the derivation of hypotheses with regard to the connection between knowledge and (economically relevant) performance, because one of the main claims of the KM literature is that an increase of (non-explicit) knowledge improves organizational performance.

Knowledge also plays an important part in modern forms of collaboration. In today's complex information environments, assigning tasks to groups is seen as a possibility to overcome information processing limitations of individuals (J. H. Davis, 1980; Kahnemann, Slovic, & Tversky, 1982). Project-oriented team work is one of the most common forms of collaboration in today's knowledge-intensive businesses (Scholl, 1997). However, "groups can also fall prey to information processing biases and limitations" (Tindale & Sheffey, 2002, p. 5). In order to investigate optimal conditions under which groups can surpass individual limitations without falling prey to limiting processes, the effect of knowledge distribution inside groups on their performance has been the subject of extensive research, for example in the field of cognitive diversity (Sauer, Felsing, Franke, & Rüttinger, 2006; Mohammed & Dumville, 2001). Yet, many studies do not model an information-rich environment or complex tasks, but employ rather simple and deterministic problems (Endres & Putz-Osterloh, 1994). Only few studies deal with the effect of knowledge distribution on group performance in complex problem solving scenarios, which can be viewed as an operationalization of real-world problems (Endres & Putz-Osterloh, 1994). The existing studies do not clearly identify whether the amount of knowledge in complex problem solving groups or the amount of knowledge overlap inside a group is the key determinant for group complex problem solving performance. Based on Scholl's theory of team effectiveness, (Scholl, 1996, 2005), this thesis seeks to predict the effect of knowledge (after the concept has been clarified) and its distribution in groups on problem solving quality. The aim of this thesis is thus fourfold:

1. The development of a sound conceptual framework of knowledge on the individual level combining conceptualizations of knowledge from business with psychological findings and allowing psychometric approaches for its measurement.

2. The establishment of a theory on how knowledge on the group level influences complex and ill-structured tasks of today's complex business

environments.

3. The development of (computer-based) tools for operationalization of the hypotheses derived from 1. and 2.

4. The test of the hypotheses from 1. and 2. with a set of empirical laboratory experiments.

In order to achieve these aims, the second chapter will introduce and define the concept of knowledge and will present a model of knowledge types. The model is developed by linking concepts from the field of individual implicit, explicit and tacit knowledge with findings from memory, cognitive and knowledge research. In this way, the concepts are not only sharpened but possibilities for their measurement are discussed. The model introduces structural knowledge as an important sub-category of human knowledge. The third chapter critically reviews possibilities of assessing structural knowledge and ends with the proposition for a new computer-based structural knowledge elicitation tool which may be capable of eliciting a part of individual implicit knowledge, unconscious access to structural knowledge, also referred to as structural implicitness. The domain in which this knowledge elicitation tool will be put to test, complex problem solving (CPS), is introduced in the subsequent fourth chapter. It also explains the importance of knowledge in today's complex and ill-structured tasks, which belong to the category of complex problems and develops hypotheses with regard to the connection between structural knowledge and CPS performance.

Before these hypothesis are put to an empirical test, the fifth and sixth chapter shift the focus from the individual level to the group level: A theory on the influence of knowledge distribution in groups on the group's complex problem solving ability is devised in the fifth chapter. Based on this theory, the sixth chapter introduces a possibility of augmenting a knowledge management system for the prediction of group performance. Thus, at the end of the sixth chapter, two computer-based methods for knowledge elicitation will have been introduced: A test for individual structural knowledge (from chapter three) and a system for group performance prediction (from chapter six). The research methodology that will put the developed hypotheses and systems to test is reported in the seventh chapter, the eighth chapter reports the results on the individual and on the group level. The thesis closes with the ninth chapter that discusses the findings, summarizes the presented work and provides an outlook on future research.

Chapter 2

The Concept of Knowledge

This chapter[1] will present a general broad definition of the term knowledge before introducing sub-categories of the concept, such as declarative, procedural, and implicit knowledge. Viewpoints from different scientific fields such as memory science, cognitive psychology, organizational psychology, and economics will contribute to their discussion. It leads to a dimensional model of knowledge types that encompasses all introduced concepts and leads to certain hypotheses with regard to the role of a specific type of knowledge, structural knowledge.

2.1 A Broad Definition of Knowledge

The difficulties that epistemology, the philosophy of the nature and scope of knowledge, has with defining the term become visible in the discussion on the definition of knowledge (compare Güldenberg, 1999, for an overview). The classic definition of knowledge in western philosophy in Plato's dialogue Theaetetus 201 (Eigler, 1990) describes knowledge as justified true belief, i. e. a subject knows something (referred to as P) if the subject believes in P, the subject is justified to believe in P and if P holds true. This statement has been rephrased in various ways, e.g. by Chisholm (1957, p. 16, cited according to Gettier, 1963) who states that S knows that P if three conditions are met:

1. S accepts P,

2. S has adequate evidence for P, and

[1]An earlier version of this chapter has been published together with Kozo Sugiyama (Meyer & Sugiyama, 2007). His contributions to that paper, namely the application of the introduced model in the evaluation of a research project, are not included in this thesis.

3. P is true.

Note that in both accounts, truth is a required feature of knowledge in order to distinguish it from errors.

In his classic paper, Gettier (1963) constructs an example where all three conditions are met without the subject knowing P: Smith and Jones apply for a job and Smith has evidence for and strongly believes that Jones will get the job and that Jones has ten coins in his pocket. This proposition entails that the one who will get the job has ten coins in his pocket and Smith is justified to believe that this proposition is true. If, unknown to him, Smith gets the job and Smith has, unknown to him, ten coins in his pocket, the proposition still is true, Smith is justified to believe in it and still Smith does not know that he himself gets the job.

Cognitive psychology avoids those pitfalls of the requirements of truth as a feature of knowledge by referring to it – simply speaking – as the content of human long-term memory (Strube & Schlieder, 1998). This pragmatic view appears too broad, because it des not distinguish knowledge from other forms of memory content, e. g. errors and assumptions. One way of introducing some form of assessment in order to distinguish knowledge from errors is to combine the views from psychology and philosophy by including constructivist aspects. Constructivism assumes that every individual mentally constructs his or her own environment based on his or her sensory input. Thus, there is no such thing as objectivity or absolute truth, because there is no objective depiction of reality (Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW), 2001). If there is no such thing as an objective truth, another characteristic of knowledge has to be introduced in order to distinguish it from errors. Therefore, the constructivist term ‘viability’ is introduced: “Actions, concepts and conceptual operations are viable if they fit to the intentions or descriptions for which they are used” (Glasersfeld, 1996, p. 43, own translation). This allows the inclusion of an assessment in the concept of knowledge that does not require objective truth: “Knowledge is not a picture or representation of reality; it is much more a map of those actions that reality permits. It is a repertoire of concepts, semantic relationships and actions or operations that have proven to be viable for the attainment of our goals” (Glasersfeld, 1997, p. 202, own translation).

From this perspective, knowledge contains an assessment in the way that it contains maps of certain aspects of the world that proved to be viable.

2.1.1 Data, information, and knowledge

According to Güldenberg (1999), knowledge is the product of a learning processes during which data is perceived as information (p. 161). This view requires a distinction between the three concepts. According to Aamodt and Nygard (1995), data is observed, uninterpreted symbols, for example characters. Information is interpreted symbols and symbol structures, and knowledge is interpreted symbol structures used within a decision process (compare Figure 2.1).

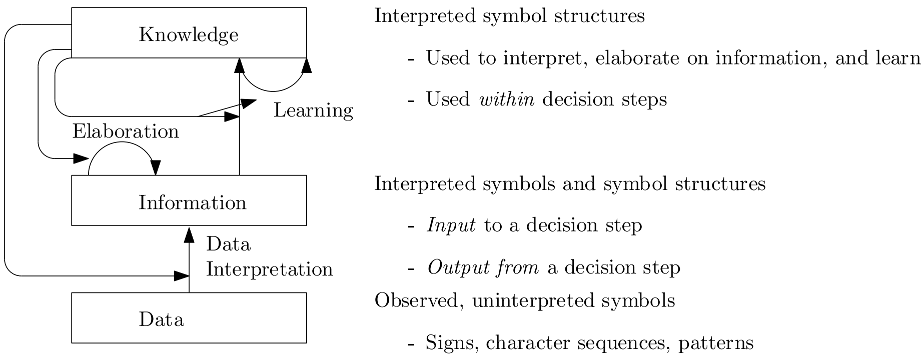


Figure 2.1: Data-Information-Knowledge model (adapted with permission from Aamodt & Nygard, 1995, p. 195) © 2008 Elsevier B. V. All rights reserved. http://www.sciencedirect.com/science/journal/0169023X

This process of interpretation is understood as the process of embedding information into existing knowledge, symbolized by the arrow pointing from knowledge to the arrow between information and knowledge.

Güldenberg (1999) further states that in knowledge-based systems, knowledge is stored in knowledge repositories as structural connectivity patterns. He states: "Knowledge is the entirety of all products of learning, in which data is perceived as information and is stored as structural connectivity patterns" (Güldenberg, 1999, p. 161, own translation). The distintion between data, information, and knowledge is also elaborated by the German Philosopher Ulf von Rauchhaupt (2005). In his view, knowledge creation is an act of information organization, which leads to the sometimes synonymous use of the terms knowledge and information:

> The interpretation of information that leads to knowledge by an individual can be seen as an organization process: The process of interpreting data as information is already an act of organization: we perceive the data, order it and link it with other information

> from our previous knowledge (...). In this way, the new information becomes a part of our knowledge for further acts of interpretation. What we know then has a higher degree of organization than the barely obtained information. [...]
>
> Contrariwise, knowledge becomes information again if it is expressed. In order to express knowledge, an individual cannot supply his or her entire network of previous knowledge, which consists of his or her entire history of experiences, his or her cognitive biography. In order to share and exchange knowledge, humans have to partially reduce it to information. The possibility of such a reduction, the possibility to encode knowledge into information, is the reason for the sometimes synonymous use of the terms knowledge and information. Information is a condensed form of knowledge, knowledge is information whose organization exists only for the knowing person (Rauchhaupt, 2005, p. 98f, own translation).

Note that in von Rauchhaupt's view, knowledge is represented as a network of organized information. This network view is in line with Güldenberg's perception of knowledge as structural connectivity pattern (see above).

2.1.2 Knowledge definition

Güldenberg's definition mentions the aspect of assessment or viability only indirectly by referring to the process of learning. Thus, the above-mentioned constructivist view, von Rauchhaupts conceptualization of knowledge as networked information and Güldenberg's (1999) definition of knowledge as structural connectivity patterns are combined into a general definition of knowledge that includes an aspect of assessment:

> *Knowledge is defined as a set of structural connectivity patterns. Its contents have proven to be viable for the achievement of goals.*

Based on constructivist assumptions, this definition avoids the term 'representation of reality'. It pays tribute to the fact that mental models of an individual are the result of a construction of environment, which can be very different from one individual to another (Opwis & Lüer, 1996). The term 'structural connectivity patterns' includes von Rauchhaupt's view on knowledge as a network of organized information (see above) and allows the inclusion of knowledge on different collective levels (individual and organizational), since organizational knowledge is embedded in the system or structure of the organization (see Chapter 6). The stress on the fact that knowledge has proven to be viable underlines assessment as a feature of human knowledge.

2.2 Explicit and Non-explicit Knowledge

Individuals can perform actions without being able to explain them and they can explain actions without being capable of performing them (Dick & Wehner, 2002). A person can know something without being able to articulate it. Consider Wittgenstein's two famous expressions (Wittgenstein, 2001, section 78, quoted according to von Rauchhaupt, 2005, p. 91), as an example:

- I know the height of the Montblanc.
- I know how a clarinet sounds.

Long descriptions of the sound of a clarinet cannot replace the actual experience of hearing it (Rauchhaupt, 2005, p. 91). From such observations, Polanyi (1966) concluded the existence of a silent dimension of knowledge that cannot be articulated: tacit knowledge. Spender (1996) uses a similar typology, differentiating between implicit (produced through action) and explicit (produced through communication) knowledge. Polanyi himself distinguished between explicit knowledge and tacit knowledge according to the differentiation between *Können* (being able to do sth.) and *Wissen* (knowing) in the German language (Polanyi, 1983, p. 16). Similarly, articulable knowledge is referred to as explicit knowledge by Nonaka and Takeuchi (1995), knowledge that is difficult to articulate or cannot be articulated at all is referred to as tacit knowledge, too. However, Li and Gao (2003) argue that Nonaka's understanding of tacit knowledge differs from Polanyi's concept and criticise their synonymous use in literature. The authors stress that Nonaka and Takeuchi and Polanyi referred to two different observations in two fundamentally different cultural contexts. Polanyi studied European scientists, where Nonaka and Takeuchi studied factory workers in Japan. Li and Gao state:

> It is out of Polanyi's argumentation for a careful differentiation between tacitness and implicitness, but from his terminology, tacitness is evidently different from impicitness [sic]. Implicitness, an other [sic] form of expressing knowing, does exist. It implies that one can articulate it but is unwilling to do that [...]. [...] When Nonaka and Takeuchi used Polanyi's dichotomy [...] we

> can see that actually what they mean by 'tacitness' includes 'implicitness' (Li & Gao, 2003, p. 8).

The fact that implicit and tacit knowledge are described as two separate things and the hint at different levels of codifiability points towards a dimensional character of non-explicit knowledge (see also Kogut & Zander, 1992). The dimension spans between the poles explicit knowledge and tacit knowledge. Following Li and Gao, implicit knowledge lies somewhere in between. Knowledge elements can be classified into this continuum based on the degree of their codifiability (European Foundation for the Improvement of Living and Working Conditions, 2004), compare Figure 2.2.

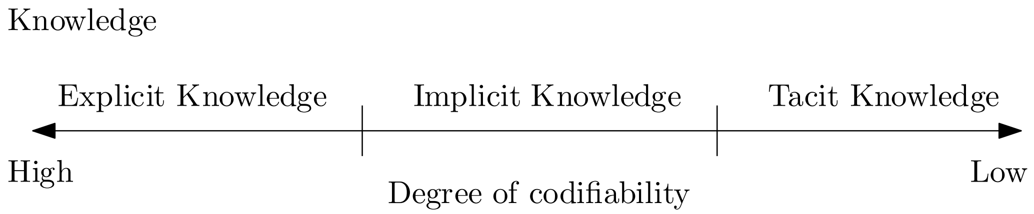


Figure 2.2: Dimensional classification of knowledge (based on European Foundation for the Improvement of Living and Working Conditions, 2004; Kogut & Zander, 1992; Li & Gao, 2003)

In their theory of implicit and explicit knowledge, Dienes and Scott (1999) also arrive at different levels of explicitness from a behavioral science perspective. They define knowledge as an attitude towards a proposition (predicating a property to some entity) which is true. Despite the issue of introducing the concept of truth to a knowledge definition (see above), their approach also leads to different levels of knowledge explicitness:

The representation of a fact and its internal functional use constitute knowledge. For example, the sentence "I know that this is a cat" consists of a person (I), a proposition (this is a cat) and an attitude relation (knowing), whereby the attitude results from how the proposition is used by the person (Dienes & Perner, 1999, p. 737). Functional use of a proposition without its representation, or, more precisely, without the representation of self and attitude towards that proposition, constitutes implicit knowledge. Knowledge is explicit if it is represented by analogies with explicit verbal communication, i.e. if the functional use of a proposition is represented by representational distinctions. "That is, there is an internal state whose function is to indicate the content of the knowledge" (ibid, p. 737). This leads to three types or stages of implicitness, depending on how many of the elements that constitute knowledge are explicitly available: content, attitude, and self. For example, “under subliminal conditions only the properties of a stimulus (the kind of stimulus) get explicitly represented (...), not the fact that there is a particular stimulus event that is of that kind” (ibid, p. 738).

In summary, the notion of a dimensional transition from explicit knowledge to implicit knowledge is supported by theories from different fields of research. However, the inconsistencies in the usage of the terms implicit and tacit knowledge and Li’s criticism of Nonaka underline the difficulties in clearly defining the constructs implicit and tacit knowledge. The following sections will therefore be used to elaborate on psychological and cognitive findings from the fields of memory research and cognitive science. This appears appropriate as Li and Gao, Polanyi, and Dienes and Perner refer to individual knowledge. These findings will then be integrated into the dimensional classification of knowledge, allowing a clearer definition and specification of non-explicit knowledge.

2.3 Models from Memory Research

In this section, findings from memory science are outlined that will each be connected to the concepts of implicit and explicit knowledge. For the following descriptions, the definition for memory employed by Sinz (1979) is used:

> The term memory describes the storage that depends on the learning of ontogenetically acquired information that selectively inserts itself into phylogenetical neuronal structures and can be recalled at any given point in time, e.g. that can be made available for situationally appropriate behavior (Sinz, 1979, quoted according to Markowitsch, 2002, p. 74, own translation).

It can thus be argued that individual knowledge is stored in memory (Strube & Schlieder, 1998). Generally speaking, memory models either describe the structure of memory or the processes that are active in memory (Tulving, 2002). In the following, two common and extensive memory models are presented: Tulving’s content-related memory model (Markowitsch, 1999, 2002; Squire & Frambach, 1990; Squire, Knowlton, & Musen, 1993; Tulving, 1972; Tulving & Schacter, 1990; Tulving, 1995, 2002) and the multimodal theory of memory (Engelkamp, 1991; Engelkamp & Pechmann, 1993; Engelkamp & Zimmer, 1994; Engelkamp, 1998). The latter includes both structure and processes; the former is a classification approach. After these two models are laid out, knowledge representation models from the field of cognitive science are described.

2.3.1 Content-related memory model

This memory model is based primarily on work by Tulving (1972; 1990; 1995, 2002) and is also supported by neuroanatomical findings by Markowitsch (1992, 1999, 2002), and Squire and his colleagues (1990; 1993).

Working memory

Firstly, the model postulates a memory with a short memory span of a few minutes that all information needs to pass through in order to be permanently stored in the long-term memory (Markowitsch, 2002, p. 85). It can be understood as that part of memory that is active at a certain point of time (ibid) and is therefore referred to as working memory.

Several findings indicate that the working memory is made up of several modality-specific subsystems, e.g. for verbally and visually coded information, that are coordinated by a central entity (see Squire et al., 1993, for an overview). The capacity of the verbal working memory is five (plus/minus two) informational units (chunks) (Markowitsch, 2002), the capacity of the visual working memory is assumed to be four objects that can have up to 16 memorable features (Vogel, Woodman, & Luck, 2001). On a neural level, network theories are most popular for describing memory processes (Markowitsch, 1999, 2002).

Long-term memory

Within long-term memory, where the maximum length of storage is practically unlimited, several different memory systems can be differentiated according to their content (long refers to a span beyond a minute, cf. Markowitsch, 1999). The youngest part of memory from an evolutional biological point of view is episodic memory (Tulving, 2002). “It consists of singular events that can be specified according to time and place” (Markowitsch, 2002, p. 88, own translation). Together with semantic memory that stores general facts about the world, it belongs to the declarative memory system. Episodic memory builds on semantic memory.

According to Squire et. al, “Declarative Memory is fast, it is not always reliable (i.e. forgetting and retrieval failure can occur), and it is flexible in the sense that it is accessible to multiple response systems” (Squire et al., 1993, p. 458). The content-related memory model states that humans also possess reflexive or non-declarative memory. “Non-declarative memory is slow [...] reliable and inflexible” (ibid). Reflexive memory is differentiated into three subsystems: procedural memory, the priming system, and the part of memory that is responsible for conditioning. For the non-declarative memory systems, Squire synonymously employs the term "implicit memory" (Squire et al., 1993, p. 471). The procedural memory system contains skills and habits: "Skills are procedures (motor, perceptual and cognitive) for operating in the world; habits are dispositions and tendencies that are specific to a set of stimuli and guide behaviour" (ibid). Under certain conditions, these can be acquired unconsciously. It should be noted that procedural memory does also contain skills that are not on a motor level, but on a perceptive and/or a cognitive level (Squire et al., 1993, p. 471). Non-declarative memory can be acquired independently of declarative memory (Squire & Frambach, 1990). Figure 2.3 illustrates the model.

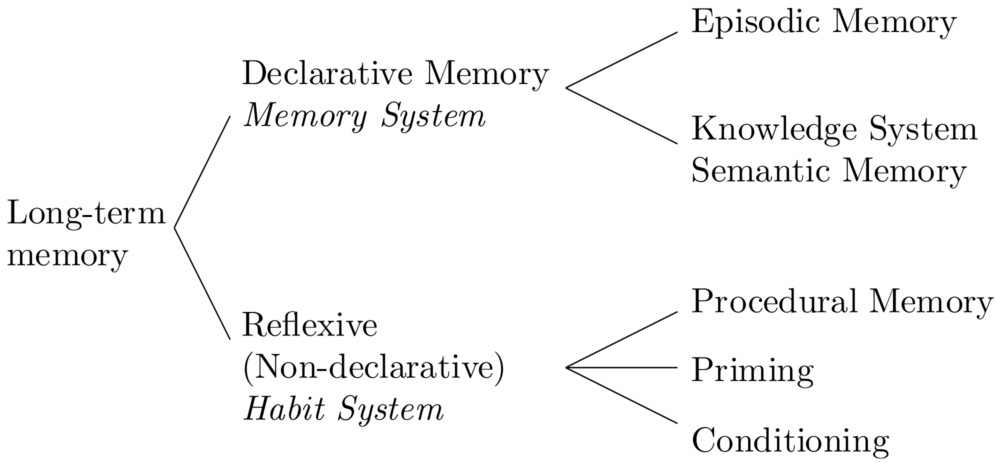


Figure 2.3: Overview of the content-related memory system (based on Tulving, 1995)

Regarding the distinction between implicit and explicit knowledge, findings from artificial grammar learning are of special importance. In these experiments, participants receive lists of meaningless words, being told that the syntax of these words does follow a set of rules, which remain undisclosed. After the participants are presented with 'valid' meaningless words in a trial period, they are asked to decide whether previously unknown meaningless words obey the rules or not. Although participants are typically unable to explain the grammatical rules on which their judgment is based, the number of correct decisions is above coincidence (Squire et al., 1993, p. 473f). It has been established that the knowledge base resulting from the trial period is tacit in nature (Reber, Walkenfeld, & Hernstadt, 1991). In order to empirically test whether implicit memory processes have "phyletic primacy" (Reber et al., 1991, p. 185) over explicit processes, Reber, Walkenfield, and Hernstadt (1991) conducted a laboratory experiment. If, they argued, implicit memory processes have "phyletic primacy", they should be largely independent of explicit memory processes, i.e. they should be (a) more robust against insults, (b) they should display a tighter distribution and should (c) operate largely independent of standard measures of cognitive capability, e.g. psychometric IQ-tests. As assumption (a) appears to be established in clinical literature, the authors attempted to validate hypotheses (b) and (c) in a laboratory experiment. Therefore, they used an artificial grammar-learning task and a series-solution task. The grammar-learning task was designed as described above. In the series solution task, they presented participants with sets of ordered sequences of letters. Participants had to determine the correct continuation of the sequence. Although this task appears to be similar to the grammar-learning task on a superficial level (strings of letters that are made up by a complex set of rules), the necessary cognitive functions for the series-completion tasks are different from the recognition-based grammar-learning task, as explicit rule-identification and application is required. Participants also completed an IQ-Test (WAIS-R). In the result, the mean percent of correct judgments for both grammar learning and series solution was .61 and significantly above chance. The distribution of scores from the explicit task showed greater variance than the implicit task, which was also visible in the ranges of percentage correct: It ranged from .33 to .92 in the explicit task and from .46 to .73 in the implicit task. Assumption (b) thus receives support from this experiment. The same applies to hypothesis (c): IQ correlated significantly with the percentage of correct answers in the explicit series task ($r = .69, p < .01$), but non-significantly with the artificial grammar well-formdness test ($r = .32, p > .05$). These results support the proposed differentiation between implicit and explicit memory processes. However, According to Squire et al. (1993), it is impossible to determine whether participants employ implicit knowledge in terms of procedural knowledge, or whether they employ incomplete or weak declarative knowledge in artificial grammar learning (1993, p. 474). This statement implies that non-verbalizable knowledge can have two causes: It is either procedural or weak declarative. Both have in common that knowledge elements are accessed subconsciously. For this reason, Markowitsch rejects the synonymous use of procedural and implicit memory:

> Implicit and explicit memory are not two different kinds of memory, they are different forms of expressing memory or phenomenologically different ways of retrieving specific events or experiences. Implicit means without making the actual content and its meaning conscious, explicit means including the associated connotations (time-spatial coordinate structure, the how, when and where

> of the encoding process). Explicit recall manifests the recalled information as an episode that can be personally experienced. The neural structural combinations that are responsible for implicit and explicit memory processing do differ (Markowitsch, 1999, p. 25, own translation).

Note the similarity between this statement and the different levels of explicit knowledge as put forward by Dienes and Perner (1999, see above) in their theory of implicit and explicit knowledge. Implicit memory thus describes an unconscious processing of memory contents, whereas the term explicit memory refers to a conscious mode of processing. Kluwe (2006) arrives at the same conclusion when he describes implicit knowledge as "superior performance in cognitive tasks based on an unconscious use of previously perceived and not intentionally stored information" (p. 40, own translation). In an analogous way, Kluwe defines explicit knowledge as conscious recall of previously encoded information.

In addition to conscious and unconscious use of memory contents, there exists the phenomenon that previously acquired knowledge is not used at all. This so-called inert knowledge (Renkl, 1996) is used to explain the discrepancy between knowledge and behaviour in pedagogy.

The above discussion on different forms of memory usage illustrates that a model needs to be presented that spans not only different types of memory but also different memory processes. The multimodal memory model (Engelkamp, 1991; Engelkamp & Pechmann, 1993; Engelkamp & Zimmer, 1994; Engelkamp, 1998) is such a model and is introduced in the next section.

2.3.2 The multimodal memory model

This model includes both process and structural assumptions. In accordance with Tulving's differentiation between semantic and episodic memory, Engelkamp and colleagues introduce a multimodal memory model for episodic memory processes.

The actual memory model is based on the assumption of the existence of two orthogonal dimensions: sensory – motor and verbal – nonverbal. Within these two dimensions, the authors postulate a conceptual system linked to modality-specific entry and output systems (Engelkamp, 1998, p. 35), the so-called sensory-motor systems. Throughout interaction between the systems, information is represented on two different levels: the conceptual system operates independently from the modality of the input; on the sensory motor level, encoding is specific for the modality of input and output (see Figure 2.4).

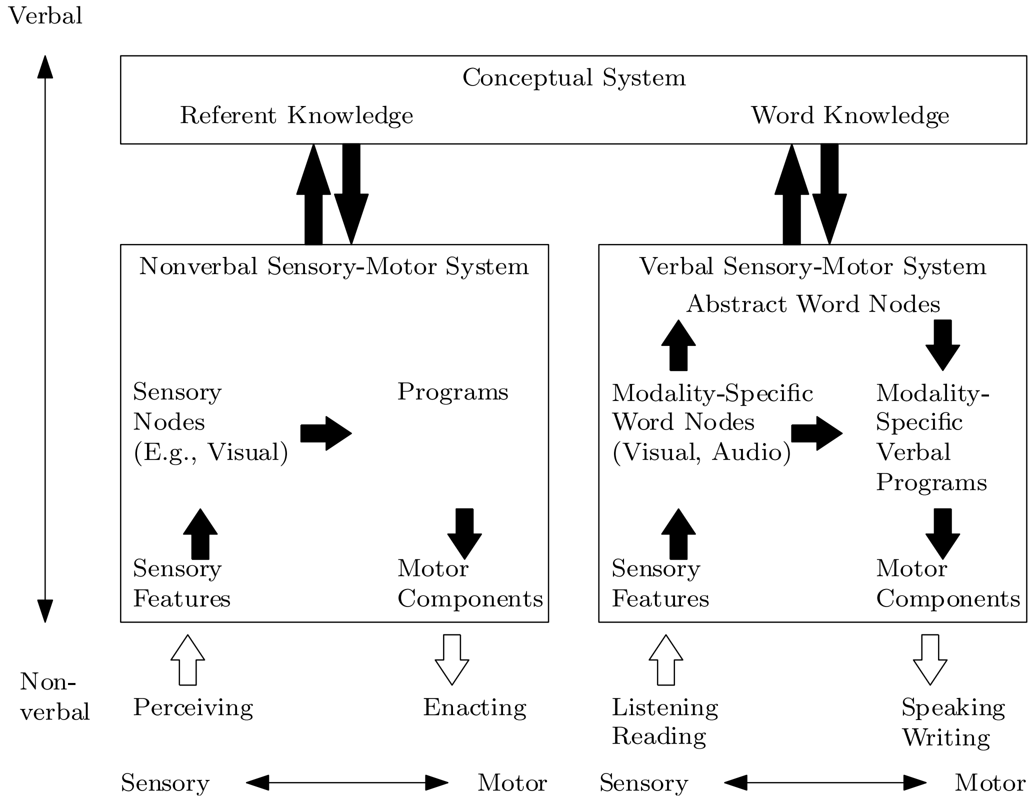


Figure 2.4: General architecture of the multimodal memory model (adapted from Engelkamp & Zimmer, 1994, p. 35). Reproduced with permission from *The Human Memory: A Multimodal Approach* by Johannes Engelkamp & Hubert D. Zimmer, ISBN 0-88937-116-4 , p. 35 © 1994 by Hogrefe & Huber Publishers Seattle Toronto Bern Göttingen

Sensory motor systems are differentiated into sensor and motor systems. For simple items such as single concepts or actions, the authors assume a strict separation of memory content into different subsystems. Complex material is based on several modalities and thus on several subsystems. Sensory motor knowledge preserves experiences close to perception and behaviour. Referent knowledge combines concepts into propositions (compare next section).

One speciality of Engelkamp's model is the assumed lack of connection between verbal sensory motor system and nonverbal sensory motor systems. The authors assume that an access from the verbal sensory motor system to the nonverbal sensory motor system requires participation of the conceptual system (Engelkamp, 1991, p. 8).

In order to make an action verbally explicit, a reference to a motor program must be present in the conceptual system that must be connected to a word node in the verbal sensory-motor system. Only if this word node is connected to a modality-specific verbal program, verbalization can take place. Explicit knowledge about acting requires a connection of all three systems with regard to a specific content. If someone is capable of performing an action without being able to verbalize it, this can have two reasons:

The sensory nodes, e.g. visual nodes, are directly connected to motor programs in the nonverbal sensory motor system. Such content that is related to a single subsystem can only occur for simple stimuli and actions (see above). An example would be turning the head towards a face we recognize in a crowd. The face is the visual sensory node that was activated by perceiving the face; it is directly connected to the motor program for turning the head. Knowledge regarding face recognition is accessed without any use of the conceptual system and is thus not codifiable. This type of knowledge can be labeled embodied knowledge.

The performance of complex acquired tasks that are difficult to verbalize includes both the nonverbal sensory-motor system and connected referent knowledge within the conceptual system. Conceptual knowledge does not include knowledge of words, or known words are not connected to the word nodes in the verbal sensory-motor system. The connection between the conceptual system and the verbal system, which was established during learning, may have faded over time since the verbal sensory-motor system is no longer required after learning in order to perform the action. Actual performance of the action requires only the conceptual and the nonverbal sensory-motor system. Due to the inclusion of actions and behaviour in the memory model, the multimodal memory model is capable of explaining differing levels of verbalization of behaviour that can be observed. Non-explicit knowledge acquired over time that was compiled into automated actions, such as the expert mastery of a musical instrument, can be explained as a disassociation of the verbal sensory-motor system for that particular action or concept.

All in all, Tulving's and Engelkamp's models can explain the existence of memory content which is not consciously accessible. In Tulving's model, this can either be procedural knowledge or unconscious access to (weak) declarative knowledge. In Engelkamp and Zimmer's model, sensory-motor systems without connection to the conceptual system or to the verbal sensory-motor system are active. The phenomenon of non-explicit knowledge thus finds its correspondence in memory psychology. However, the cognitive dimension of procedural knowledge and the reference to semantic network structures in previous sections do require the introduction of higher-level concepts of representation. These will be outlined in the following section, prior to introducing the model of knowledge types.

2.4 Cognitive Models for Knowledge Representation

Until now, different memory structures and their relationships have been described. Knowledge organization goes beyond this level; it deals with how semantic structures and productions are actually organized. Two forms of knowledge representations are propositional representation systems and rule-based representation systems (Opwis & Lüer, 1996). They will be briefly described in this section.

2.4.1 Propositional representation systems

Propositional representation systems represent "verbally articulable information with the help of special symbol structures, so-called propositions" (Opwis & Lüer, 1996, p. 349, own translation). Two famous propositional systems are semantic networks and cognitive schemes. Semantic networks are formally depicted as graphs in which nodes represent linguistic units and edges represent linguistic relations (compare Definition 1 in Section 3.1). This approach is primarily based on Quillian (1968, 1988), who assumed a networked organization of individual semantic knowledge. A problem with semantic networks is their limited expressiveness and the fact that they do not include methods for dealing with objects in memory. This criticism can be met with an advancement of semantic networks: cognitive schemata. They refer to a heterogeneous group of pre-structured representational formats. The two most popular types of cognitive schemata are frames and scripts (Strube & Schlieder, 1998). Frames are data structures in which experiences are generalized and that represent circumstances and expected coherences from a certain realm of reality (Schnotz, 1994). These representations contain constants and vacancies that store probabilities for other schemata that can be inserted. In this way, a schema is an instantiable class of a situation. A proposition is a structure that is created on instantiation of a schema.

A script is a frame for a situation involving several actions, much like a film script for standard situations. The most famous example is the script for a restaurant visit, in which certain behaviour such as waiting to be seated, being seated, receiving the menu, ordering, eating, paying and leaving are organized in a sequential manner. Scripts allow economic information processing that is steered by expectations. Note that scripts and schemata do not distinguish between declarative and procedural knowledge and extend beyond the scope of semantic network models that are assumed to be valid for semantic memory only

2.4.2 Rule-based representational models: ACT-R

This form of knowledge representation assumes concurring processes within a production system. Contrary to cognitive schemata, production systems claim separate storages for declarative and procedural knowledge. A production rule or production connects a condition to an action. Declarative knowledge consists of data structures processed in working memory. Processing takes place by applying production rules to the content of the working memory (Schnotz, 1994). Declarative knowledge is represented as a semantic network with edges and nodes and is stored in the declarative long-term memory. The nodes of the network are knowledge units; the edges between them correspond to certain relations between these units (Schnotz, 1994, p. 96).

A prominent rule-based representational model that has developed towards a unified theory of the mind is Anderson's Adaptive Control of Thought - Rational model (ACT-R). It was first outlined in 1976 (J. R. Anderson, 1976), substantially altered to version 4.0 (J. R. Anderson & Lebiere, 1998) and was again altered significantly in its transition to the fifth version (J. R. Anderson, Byrne, Douglass, Lebiere, & Qin, 2004). It is now a unified theory of cognition explaining the acquisition, representation, and use of knowledge. It is also a cognitive architecture that has evolved into a formal (computer) language for simulating human cognition. Because the ACT-R theory encompasses many of the concepts already introduced and, in its most recent form, combines research from a brain-region-level to the level of cognitive models, it is presented in its two most recent forms as an understanding of the model in the version 5.0 is difficult without knowledge of the prior version.

ACT-R 4.0 (J. R. Anderson & Lebiere, 1998) assumes two types of knowledge: declarative and procedural knowledge. "Declarative knowledge corresponds to things we are aware we know and can usually describe to others" (J. R. Anderson & Lebiere, 1998, p. 5), while "procedural knowledge is knowledge that we display in our behavior but that we are not conscious of. Procedural knowledge basically specifies how to bring declarative memory to bear in solving problems" (ibid). Declarative knowledge resides in declarative memory and is represented as memory units called chunks, "configurations of various things we know" (ibid). Chunks have been referred to as productions in earlier versions of the model. Chunks have slots and associated values; a chunk interassociates these values with certain associative strengths. For example, the chunk representing the fact that the sum of three and four equals seven can be textually represented as follows:

Chunk

is a	ADDITION FACT
addend1	Three
addend2	Four
sum	Seven

In this case, the chunk has four slots (is a, addend1, addend2 and sum) that hold the corresponding values.

Procedural knowledge resides in procedural memory and is represented as production rules. Each production is an IF-THEN pair where the IF parts specifies the condition in which the production becomes active and the THEN part specifies what to do. Apart from declarative and procedural memory, the ACT-R 4.0 model also includes a goal stack from which the current (sub-) goal of the cognitive process can be obtained. This current goal directs the actions of the procedural memory. The flow of information through the modules is presented in Figure 2.5; compare Anderson and Lebiere (1998) for further information.

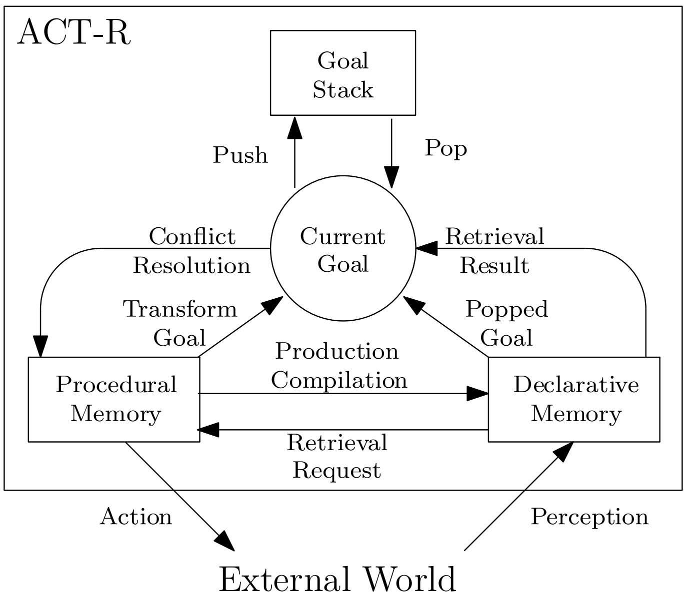


Figure 2.5: Flow of information among the various modules of the ACT-R 4.0 model (adapted with permission from J. R. Anderson & Lebiere, 1998, p. 11). Copyright 1998 From The atomic components of thought by John Robert Anderson and Christian Lebiere. Reproduced by permission of Routledge, Inc., a division of Informa plc.

While the fourth version of the model focuses on the cognitive processes, the fifth version of the architecture (J. R. Anderson et al., 2004) is heavily linked to neurobiological findings and brain region activity. It is intended to explain human cognition in general, i.e. knowledge acquisition, knowledge application, goal setting, and acting. It postulates the visual module as the key sensory module, although aural modules are also mentioned (note the similarity to Engelkamp's model in Section 2.3.2, which also assumes independent sensory and motor modules). Sensory modules take in information from the environment. Manual modules control hand movements and allow interacting with the environment. Apart from these modules, there are two other modules that interact with the environment without direct access to it: the intentional module and the declarative module. The intentional module or goal module keeps track of current goals and intentions; the declarative module retrieves information from memory. The declarative module can thus be seen as the long-term declarative memory. Each of the four introduced modules possess a buffer that makes one informational unit from the module (a chunk, see above) available to the central production system, which is at the centre of the ACT-R 5.0 model. The interaction of these buffers through application of production rules determines cognition (J. R. Anderson et al., 2004, p. 1038). The central production system is where the production rules reside and is thus the procedural memory of the suggested architecture. The central production system holds the production rules and further consists of three sub-sections or functions:

- Pattern matching: match the declarative chunks in the buffers with a production rule in the central production system
- Action selection: A form of conflict resolution: a "winner-lose-all"-system inhibits neurons representing the selected action, "which then no longer inhibit the thalamus from producing the action" (J. R. Anderson et al., 2004, p. 1038).
- Controlling the execution of production rules

All of the buffers, all of the components of the production system and all modules except for the intentional module (so far unidentified) are mapped to specific brain regions according to current neuroscientific findings: The declarative module is located in the temporal regions of the hippocampus, its buffer is associated with the ventrolateral prefrontal cortex (VLPFC). The goal buffer, responsible for keeping track of of the state of a problem solution, is associated with the dorsolateral prefrontal cortex (DLPFC), whereas the manual module and buffer are associated with the motor areas of the cerebellum. The visual module is associated with the occipital areas of the brain, its buffer with the parietal areas. "The basal ganglia and associated connections are thought to implement production rules in ACT-R" (J. R. Anderson et al., 2004, p. 1038), pattern matching is associated with the striatum, action selection with the pallidum and the control of execution with the thalamus. Figure 2.6 presents an overview of ACT-R 5.0.

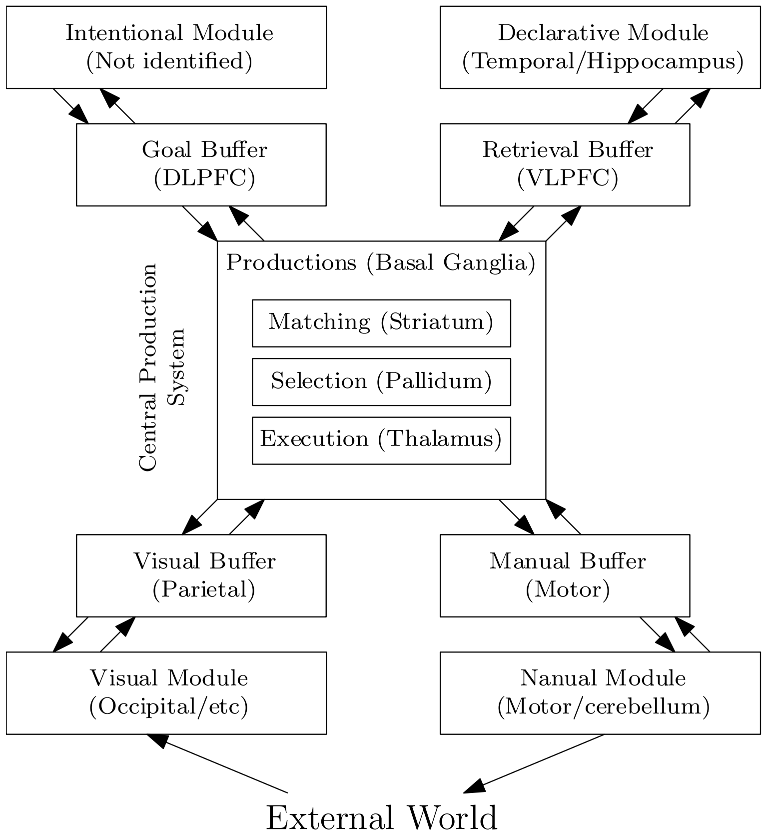


Figure 2.6: The organization of information in ACT-R 5.0 (adapted with permission from J. R. Anderson et al., 2004, p. 1037). © 2008 American Psycholoical Association. DLPFC = dorsolateral prefrontal cortex, VLPFC = ventrolateral prefrontal cortex.

ACT-R 5.0 is capable of combining a unified theory of cognition with current neurobiological findings with regard to brain functions. Furthermore, it led to the improvement of the ACT-R framework formal computer computer language that allows the creation of specific ACT-R models for specific cognitive tasks and produces predictions which can than be compared to empiric data. This has been successfully done for various domains (J. R. Anderson et al., 2004), but the creation of such models is still restricted to tasks that are clearly structured and have a predictable course of action.

One important aspect in the context of this thesis is the fact that in the

ACT-R model, declarative knowledge elements (or the declrative chunks provided by the declarative buffer) are associated with other declarative knowledge elements and goals through productions that fire on recognizing fitting chunks. This leads to the assumption that a larger repertoire of production rules or differing patterns of behaviour leads to a higher degree of connectedness between declarative knowledge elements.

All in all, propositional representation systems and rule-based representational models describe individual knowledge representations and knowledge processes on a higher level than the models in the previous section. Note that a network organization of knowledge is central to all models that have been described so far.

2.5 Structural Knowledge

Since network organization is a central characteristic underlying the models discussed, its organization is introduced as an independent characteristic of knowledge: Structural Knowledge. It is "[...] the knowledge of how concepts are interrelated" (Jonassen, Beissner, & Yacci, 1993, p. 4). It is a "[...] hypothetical construct referring to the organization of the relationships of concepts in long-term memory" (Shavelson, 1972, p. 226-227, quoted after Jonassen et al., 1993). The authors further state:

> "Structural Knowledge is also known as cognitive structure, the pattern of relationships among concepts in memory (Preece, 1976) [...]. Structural knowledge has also been referred to as internal connectedness, integrative understanding or conceptual knowledge" (Jonassen et al., 1993, p. 4f).

This conceptual structure facilitates between declarative and procedural knowledge and thus conditions the acquisition of procedural knowledge. "Declarative knowledge is comprised of the knowledge of the terms within a domain, but does not imply understanding. Procedural knowledge is knowledge application characterized by cognitive activities such as solving problems, formulating plans, and making arguments. Structural knowledge provides the integration and organization of declarative knowledge in such a manner to be accessible to procedural knowledge applications" (Hoole, 2006, p. 2). Thus, structural knowledge is seen as a link between knowing that (declarative knowledge) and knowing how (procedural knowledge). It has also been referred to as knowing why (Hoole, 2006; Jonassen et al., 1993). "Structural knowledge is thought to mediate the transition in learning from knowing about a domain, to application of knowledge in the domain" (Hoole, 2006,

p. 2). According to Jonassen et al. (1993), blanks within cognitive schemata are references to other schemata. In this way, the interrelations between cognitive schemata can be seen as a semantic network with schemata as nodes. This view is consistent with Quillian's concept of semantic memory that represents concepts as nodes and the relationship between concepts as links (Quillian, 1988). It is also consistent with the ACT-R theory that was introduced in the previous section.

Jonassen et al. (1993) assume that structural knowledge is always explicit, i.e. the connections between concepts can always be expressed. However, there are empirical findings indicating structural knowledge can be non-explicit. Rothe and Warning (1991) tried to elicitate the structural knowledge of experts in a limited specified knowledge domain through the structure-laying technique (Scheele & Groeben, 1984, compare Section 3.5.2). It turned out that the number of nodes and their labels were similar among participants while the labelling of the edges with Klix's standard semantic relations (Klix, 1984), differed to a great extent. Rothe and Warning concluded that their participants generally had substantial difficulties in naming the edges between knowledge nodes. This leads to the assumption that access to structural knowledge can be implicit in Tulving's sense, i.e. present but not consciously accessed.

In their theory of implicit and explicit knowledge, Dienes and Perner (1999) also support this notion by introducing the term "structure implicitness" (p. 740). It refers to a representation where "...concepts acquired incidentally and nonstrategically may have nondecomposable atomic representations in which the property structure is represented implicitly" (ibid, p. 740). "Explicit knowledge can be a representation of compounds (typically: compound properties) that leaves the structure of its components implicit" (ibid).

This assumption is supported by Davis, Curtis and Tschetter (2003) who assume that the elicitation of structural knowledge (structural assessment) also captures non-explicit knowledge. The authors state that tacit knowledge is comprised of the subtle interrelations between concepts and explicitly indicate the possibility of measuring at least a part of tacit knowledge by structural assessment. Lee, Choi and Choe (2002) follow this approach by attempting to capture the organizational members' tacit knowledge through knowledge structure elicitation techniques.

To sum up, the connections in semantic memory can be interpreted as an independent type of knowledge (structural knowledge). They can be accessed either consciously or unconsciously and can thus be non-explicit. Structural knowledge can be elicited through several different methods, which are described in detail in Chapter 3.

2.6 Dimensional Model of Knowledge Types

After all different conceptions of implicit and explicit memory, and implicit, explicit and tacit knowledge have been introduced, they are connected to each other in a single model. In order to do so, the dimensional model of knowledge codifiability presented earlier is replaced by a more detailed model by the end of this section, which is based on Schindler (2002) who states that the transition between explicit and non-explicit knowledge is fluid. In his model (compare Figure 2.7), the right pole of the dimension is assigned to non-articulable tacit knowledge that includes capabilities such as maintaining balance and face recognition. This corresponds to purely sensory-motor memory contents in Engelkamp and Zimmer's model and some motor outputs in the ACT-R model (see above). Schindler takes the next two sections of his dimension from Nonaka and Takeuchi (1995):

> "[...] tacit knowledge can be segmented into two dimensions. The first is the technical dimension, which encompasses the kind of informal and hard-to-pin-down skills or crafts captured in the term 'know-how'. [...] At the same time tacit knowledge contains an important cognitive dimension. It consists of schemata, mental models, beliefs, and perceptions so ingrained that we take them for granted" (Nonaka & Takeuchi, 1995, p. 8).

Thus the technical dimension corresponds to procedural knowledge in the non-reflexive memory system in Tulving's model (compare Figure 2.3). It is acquired through motor skill learning for which well-established theories exist (Fitts & Posner, 1979). The third part of Schindler's dimensional model of non-explicit knowledge model corresponds to the cognitive dimension of tacit knowledge (compare Figure 2.4), which can be associated with different information flows in the ACT-R 4.0 and 5.0 models without actions into the external world. This includes procedure compilation from declarative to production memory, goal transformations, and retrieval requests to the declarative memory module.

As noted earlier, implicit and explicit memory can be seen as different ways of retreival: conscious and unconscious. This is also formulated in the ACT-R model: "procedural knowledge is knowledge that we display in our behavior but that we are not conscious of" (J. R. Anderson & Lebiere, 1998). At the same time, it was stated that knowledge can be articulated in varying degrees (see above). Both features can be seen as dimensions that span an area, onto which the different knowledge types can be mapped. This knowledge map also allows the inclusion of Nonaka and Polanyi's concepts that span several knowledge types.

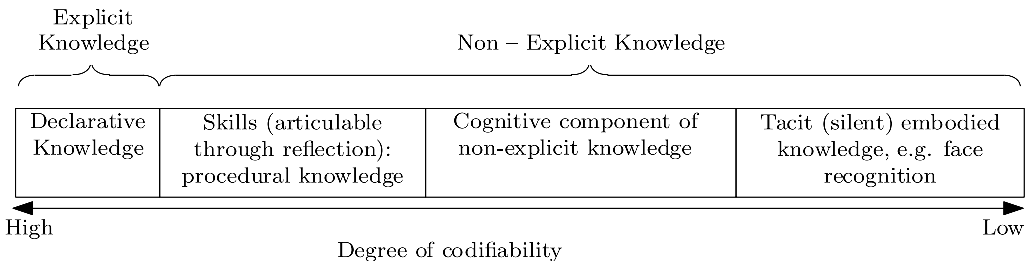


Figure 2.7: Dimensional knowledge classification according to Schindler (2002)

With reference to the different types of knowledge introduced in previous sections, Polanyi's tacit knowledge can be equated with embodied knowledge. Polanyi made no reference to conscious or unconscious use; he only referred to knowledge that cannot be articulated.

Li and Gao (2003) stress that Nonaka's concept of tacit knowledge extends beyond Polanyi's view (see above). It is thus assumed that it is Nonaka's cognitive part of tacit knowledge that surpasses Polanyi and that it does include all unconscious uses of memory, independent of the fact that they could possibly be verbalized. In this way, unconscious access to structural knowledge and weak declarative knowledge would be mapped to Nonaka's concept of tacit knowledge, but not to that of Polanyi. This assumption is supported by the fact that Nonaka and Takeuchi (1995) assume that non-explicit knowledge elements can be made explicit through appropriate techniques.

According to Nonaka and Takeuchi (1995), explicit knowledge can always be verbalized. Since verbalization requires conscious access to memory contents, declarative knowledge in Tulving's model (both semantic and episodic), declarative knowledge in the ACT-R model, and the conscious use of structural knowledge can be connected with this concept. The model is summarized in Figure 2.8. It turns out that constructs from non-psychological fields such as economics and organization theory literature are not contradictory to findings from memory research and other fields of psychological research. They integrate several constructs in a way that does seem suitable in practice. Both Nonaka and Takeuchi's constructs of tacit and explicit knowledge can be linked to an empirically founded basis.

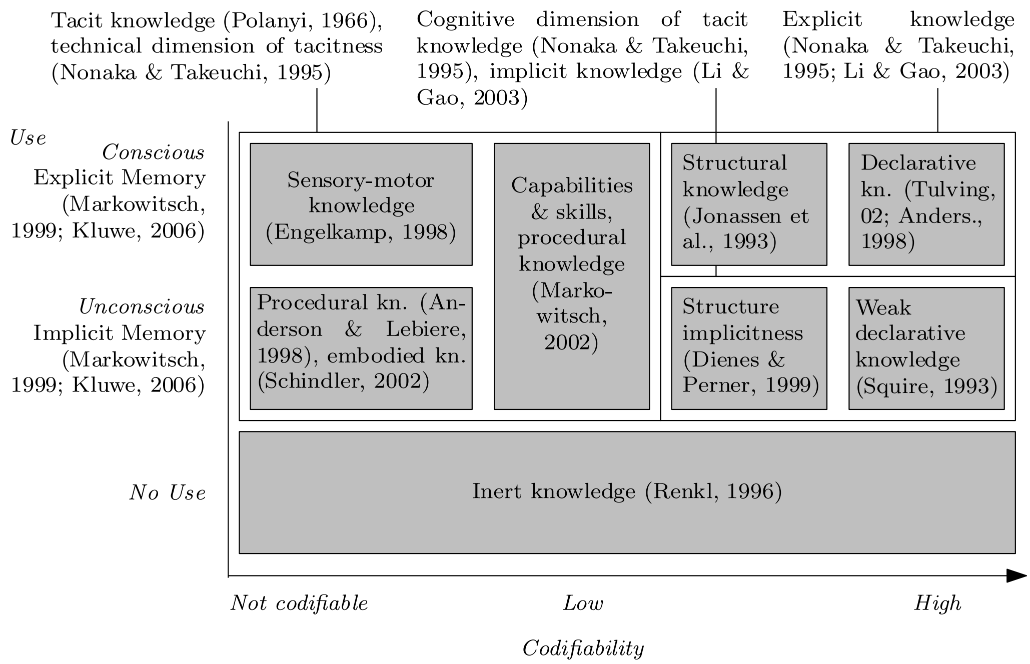


Figure 2.8: Extended dimensional model of knowledge types

2.7 Knowledge and Intelligence

As the terms knowledge and intelligence are used differently in this work, the concepts of knowledge and intelligence need to be circumscribed. Because intelligence is a debated construct (Herrnstein & Murray, 1994), this section does not and cannot claim to report the entirety of intelligence research but will rather outline the mainstream findings, sketch a recent model and report the interconnection between knowledge and intelligence.

A task force of the American Psychologists Association (APA) consisting of eleven scholars from the field of intelligence research developed a definition of intelligence that was intended to capture the common denominator of the concept in science:

> Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought. Although these individual differences can be substantial, they are never entirely consistent: a given person's intellectual performance will vary on different occasions, in different domains, as judged by different criteria. Concepts of "intelligence" are attempts to clarify and organize

this complex set of phenomena (Neisser et al., 1996, p. 77).

In a similar attempt, 52 scientists signed the definition of intelligence that was published in *Intelligence* in 1997:

> "A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – 'catching on', 'making sense' of things, or 'figuring out what to do' " (Gottfredson, 1997, p. 13).

Apart from these rather general specifications, the APA task force agrees that there are different subtests or scales of different intelligence tests that display different levels of positive intercorrelations. Spearman (1927) showed that in such sets of positive correlations, some portion of the variance of scores on each test can be mathematically attributed to a general factor (g). However, there is disagreement on what this finding means: "It has been described as a mere statistical regularity (Thomson, 1939), a kind of mental energy (Spearman, 1927), a generalized abstract reasoning ability (Gustafsson, 1984), or an index measure of neural processing speed (Reed & Jensen, 1992)" (Neisser et al., 1996, p. 78).

Thus, different models of intelligence have postulated different kinds of relationships among specific and general abilities that constitute intelligence. Some models assume hierarchical relationship between them; others assume multiple facets or factors of equal importance. An example for a hierarchical model is Vernon's hierarchical group factor theory of the structure of human intellectual abilities (Vernon, 1950). At the top of this hierarchy is Spearman's general factor (g), which accounts for the largest source of the variance in intelligence. Below g, Vernon placed several major, minor and specific group factors. Because Vernon's theory accounted for a general factor and group factors, it was seen as a reconciliation between Spearman's two factor theory (1927), postulating a general factor g and other task-specific factors (s) that did not have group factors, and Thurstone's multiple factor theory (1938) that did not have a general factor.

Based on the factor analysis method that he developed, Thurstone claimed that g was an artefact resulting from the early factorial approaches that had been used to study it by Spearman (Ruzgis, 1994) and claimed that intelligent behaviour is not based on a general factor g, but on seven orthogonal constructs that he referred to as primary abilities: word fluency, verbal comprehension, spatial visualization, number facility, associative memory, reasoning, and perceptual speed.

Hierarchical and primary factor models can be integrated into models conceptualizing intelligence as a hierarchy of different abilities on differing levels of generality that can be mapped to different modalities (Süß, 1996). According to Sternberg (1982), such models represent the highest state of evolution in intelligence research. An example of such an integrated model is the Berlin Intelligence Structure model as applied in the Berlin Intelligence Structure (BIS) Test (Jäger, Süß, & Beauducel, 1997). The BIS is based on three core assumptions:

1. Multifactorial determination: all intellectual capabilities are responsible for all intelligent performances but with varying levels of contributions.

2. Multiple modalities: Intelligent performances and abilities can be classified under certain aspects, referred to as modalities. So far, two modalities are specified: Content and operations.

3. Hierarchy: Intelligence can be differentiated into abilities of different degrees of generality.

At the most general level, general intelligence (GI) is postulated as an integral of all ability components. Below, there are the two categories Operations and Content. Operations are differentiated into four subcategories, Contents in two. Ability constructs are classified as follows (The labeling capital letters stem from the German titles of the scales, compare Jäger et al., 1997, p. 5):

K Reasoning: Computation of complex information in tasks that are not solvable at first go, but require the creation of multiple connections and analogies, exact logical thinking and appropriate assessment of information

E Imaginativeness: flexible production of ideas, availability of diverse information, a vast scope of imaginations and the conideration of many different sides, aspects, reasons, and possibilities of objects or problems

M Retentiveness: active memorization and short-term reproduction or recognition of material

B Processing speed: speed of mental processing, speed of comprehension and power of concentration in the solution of simple structured tasks of low difficulty

These abilities can be applied to three different kinds of contents: Numeric, verbal and pictographic. This could theoretically lead to twelve different abilities, but at the current state of model development, the performances in the four operations is assigned to the cells as some research issues are still in progress (Süß, 1996). Figure 2.9 illustrates the model.

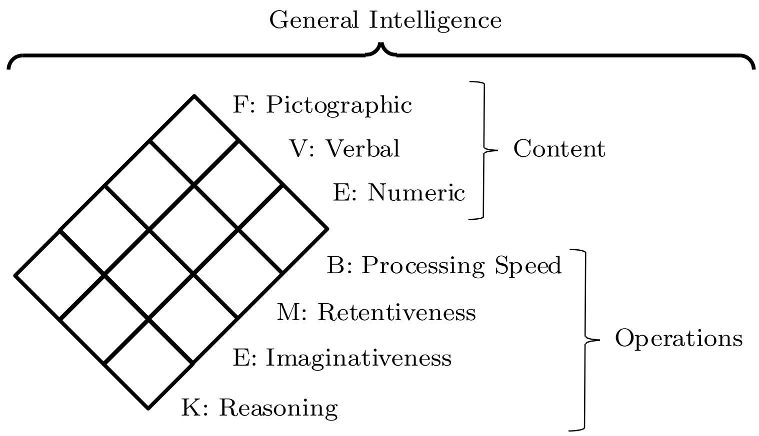


Figure 2.9: Berlin structural intelligence model (Jäger et al., 1997)

This model shows that intelligent mental operations require some form of content to work on, and this content usually consists of things we know. Intelligence and knowledge are thus intertwined: The fact that both abovementioned definitions of intelligence include references to knowledge acquisition underlines the fact that intelligence conditions knowledge acquisition and knowledge application (see also Süß, 1996). This assumption is also backed by the finding that intellectual abilities do not directly influence the results of problem solving, but are preconditions for the acquisition and application of relevant knowledge and/or expertise. Schmidt (1992) reports a path model that can be interpreted as a model of expertise acquisition (Süß, 1996) that backs this assumption. It asserts that general intelligence does not directly influence supervisory ratings of job performance, but influences the acquisition of job knowledge ($r = .46$) which again correlates with supervisory ratings at $r = .34$. According to the American Psychologists Association, "across a wide range of occupations, intelligence test performance accounts for some 29% of the variance in job performance" (Neisser et al., 1996, p. 83).

All in all, intelligence is very general mental capability that, among other things, conditions the acquisition and application of knowledge and thus accounts for a significant explanation of variance in job performance.

2.8 Summary and Research Assumptions

In this chapter, the concepts of knowledge and different knowledge types have been introduced. Knowledge in general has been defined as a set of structural connectivity patters that have proven to be viable for the attainment of certain goals. The terms explicit and tacit/implicit knowledge have been introduced: explicit knowledge is produced through words, tacit, sometimes used synonymously with implicit, through action that cannot be articulated or is at least difficult to articulate.

Since individual knowledge resides in individual long-term memory, different memory models and the neurobiological foundation of memory were introduced. Tulving's content-related memory model distinguishes between the declarative memory system that includes memory for facts and the own biography, and the reflexive memory system that includes priming, conditioning and procedural memory for actions. Engelkamp's multi-modal memory model can explain how memories can be connected to motor programs without conscious access to the memory, a phenomenon that is referred to as embodied knowledge.

Experiments from the field of artificial grammar learning lead to the assumption that implicit memory processes are phylogenetically older than explicit memory processes but have also led to the criticism that different forms of knowledge can be made responsible for a dissociation between a (successful) action and a subject's inability to verbalize the cause and conditions for these actions. Thus, implicit memory has been coined as unconscious access to memory content, including non-verbalizable content, whereas explicit memory refers to conscious access to memory regardless of the type of content. Further cognitive knowledge representation models such as chunks and procedures were introduced in the Section on the ACT-R model. It assumes that declarative chunks are associated with each other through productions.

Thus, the concept of structural knowledge was introduced, a type of knowledge that refers to the internal connectedness of knowledge elements in the cognitive system. Different concepts for memory and knowledge have then been placed into a two-dimensional model with the dimensions "codifiability" (from not codifiable to codifiable) and "consciousness of use" (no use – unconscious use – conscious use). It has been shown that groups of memory concepts that are placed into this system can be mapped to the broader terms tacit, implicit, and explicit knowledge as employed by management science. In this view, Nonaka's notion of the cognitive dimension of tacit knowledge, synonymous with implicit knowledge, is mapped to unconscious access to structural and declarative knowledge. This classification leads to the research assumption that a fraction of individual implicit knowledge – unconscious access to structural knowledge – can be psychometrically assessed through structural knowledge elicitation techniques (this will be further elaborated in the following chapters).

Apart from the memory model, intelligence has been introduced as a mental capability responsible for acquisition and application of knowledge of different forms.

Chapter 3

Elicitation and Representation of Structural Knowledge

The promise of the dimensional framework outlined in Chapter 2 lies in its inherent possibility to empirically evaluate knowledge-based activities that target individual non-explicit knowledge through structural assessment, as an increase in non-explicit knowledge should also lead to an increase in structural knowledge. This possibility justifies further investigations based on the proposed framework. In this chapter, existing approaches for knowledge elicitation are presented and evaluated, which leads to the development of an own approach for structural knowledge elicitation. This approach, the Association Structure Test, combines two existing techniques into one new computer-based structural knowledge assessment tool. The chapter closes with hypothesis on the connection between AST scores and knowledge-based performance.

3.1 Graph Representations of Knowledge Networks

There are a number of empirical studies that try to assess, elicit and measure individual structural knowledge (M. A. Davis et al., 2003; Eckert, 1998a; Goldsmith & Johnson, 1990; Goldsmith, Johnson, & Acton, 1991; Jonassen, 1993; Jonassen et al., 1993; Lee et al., 2002; Schvaneveldt et al., 1985; Schvaneveldt, 1990). In all of these studies, a different approach to knowledge elicitation is employed, but the semantic structure of individual knowledge within a specified domain is always represented as a graph or network. Graphs have been used to formalize network-theoretic models of knowledge organization in the past (Rost, 1980; Sugiyama, 2002). In order to employ graph representations of structural knowledge in the further course of this thesis, the following definitions from graph theory are introduced (based on Bonato, 1990):

Definition 1 *A* graph *or* undirected graph *G is an ordered pair $G := (V, E)$ where V is a finite non-empty set of so-called* vertices *or* nodes *and E is a finite non-empty set of unordered pairs of distinct vertices, called* edges *or* lines*. The* order *of a graph is its number of vertices $|V|$. The* size *of a graph is its number of edges $|E|$. The* degree *of a vertex is the number of other vertices it is connected to by edges.*

Definition 2 *A* directed graph *or* digraph *G is an ordered pair $G := (V, A)$ where V is a finite non-empty set of* vertices *or* nodes *and A is a finite non-empty set of ordered pairs of distinct vertices, called* directed edges*,* arcs*, or* arrows*. An arc $e = (x, y)$ is considered to be directed* from x to y*. The arc (x, y) is called the* inverse *of the arc (y, x). The* indegree *$\delta_G^-(v)$ of a vertex v in graph G is the number of arcs from other nodes to v, the* outdegree *$\delta_G^+(v)$ is the number of arcs from v to other nodes.*

Definition 3 *A* labeled graph *$G = (V, E, L_V, L_E)$ consists of a finite non-empty set V of vertices, a finite non-empty set E of edges, a label function L_V that maps nodes to labels ($L_V : V \rightarrow \{labels\}$) and a label function L_E that maps edges to labels ($L_E : E \rightarrow \{labels\}$).*

Definition 4 *A* path *p in a graph is a sequence of pairwise different vertices $p = v_1 ... v_n$ such that $v_i, v_{i+1} \in E$, i.e. from each of its vertices there is an edge to the next vertex in the sequence. We say p* connects *v_1 with v_n. The* length *of a path p is defined as its number vertices minus one.*

Definition 5 *The* distance *$d_G(u, v)$ between two (not necessary distinct) vertices u and v in a graph G is the length of a shortest path between them. If u and v are identical, their distance is 0. If no path connecting u and v exists, $d_G(u, v) = \infty$.*

Definition 6 *The* eccentricity *$\varepsilon_G(v)$ of a vertex v in a graph G is the maximum distance from v to any other vertex. The* diameter *$diam(G)$ of a graph G is the maximum eccentricity over all vertices in a graph; the* radius *$rad(G)$, the minimum.*

3.2 Knowledge Elicitation Techniques

According to Jonassen, Yacci and Beissner (1993), the assessment of structural knowledge leads to a structural representation of knowledge (e. g., a graph) and consists of three stages which each have certain techniques associated to them: Knowledge elicitation, knowledge representation, and knowledge comparison. In their view, the most effective and established techniques for the first stage, structural knowledge elicitation, are word associations and similarity ratings. In the second stage, knowledge representation, elicited knowledge is represented in a structural way. According to Jonassen et al., tree construction, pathfinder networks, and dimensional representations have proven to be viable measures for the performance of this task.

In the third stage, structural knowledge is assessed; “learners’ knowledge representations are typically compared with the teacher’s, an expert’s, or some content representation” (Jonassen et al., 1993, p. 23). However, methods for eliciting, representing and assessing structural knowledge are presented in the following without a strict difference between elicitation and representation, because usually, certain elicitation techniques are inseperably connecteed wth certain representational formats. For example, the pathfinder approach combines a specific elicitation technique (similarity ratings) with a specific representation (PFNETs) and a specific measure for comparison (C). Accordingly, other researchers have not employed a strict distinction between the above three steps, but have classified methods for structural knowledge assessment in a holistic way. Bonato (1990) introduces interview techniques, card-sorting tasks, MDS scaling procedures, and concept grids as methods for conceptual structure assessment. Hoole (2006) lists loud thinking, questionnaires, sorting and rating tasks, association tasks, and mapping techniques, i.e. concept mapping, as methods for data generation on individual knowledge. The methods which are mentioned by all three authors (Bonato, 1990; Hoole, 2006; Jonassen et al., 1993) are Card-sorting tasks, Interview methods, word association techniques, and mapping or scaling techniques. In the following sections, specific adaptions of these techniques are presented and discussed. Each technique is classified according to Hoole’s (2006) classification of structural assessment techniques. It states that the methods for elicitation of structural knowledge differ with regard to three dimensions: (1) The amount of required conscious processing (card sorts and similarity ratings require more processing than word associations and catigorial judgments), (2) the degrees of freedom in knowledge element elicitation (limited or free) and (3) the amount of hierarchy inherit in the structural representation of knowledge: Word association and category judgments assume a hierarchical representation of knowledge (i.e., a directed graph), card sorts and similarity rating tasks do not assume a hierarchic structure and allow a variety of representations.

Eckert (1998b, p. 66) classifies such procedures for graphical knowledge representations on two dimensions: testform and content. Testform describes the instruction and the media employed (e. g. paper vs. computer) while test content indicates the degrees of freedom with reference to the assignment of concepts and edge labels by a participant (compare table 3.1). This is similar to the category “Cuedness of recall” in Hoole’s framework (compare section 3.2).

Table 3.1: Eckert’s classification of procedures for knowledge representation (Eckert, 1998b, p. 66)

		Relations	
		Free	Given
Concepts	Free	I. Quadrant	II. Quadrant
	Given	III. Quadrant	IV. Quadrant

Combining both Eckert’s and Hoole’s classifications, one can basically ask three questions with regard to elicitation techniques: How are knowledge elements between which a structure is assumed elicited, how is the structure between these elements elicited and how is the complete structure assessed? Both elicitation of knowledge elements and structure can be assessed in terms of the amount of required conscious processing, the available degrees of freedom (i.e. whether cued or free), and in terms of test form. Additionally, the elicitation of structure can be classified in hierarchical and nonhierarchical (see above). For the assessment of the overall structure, it can be determined whether quantifiable coefficients are available for structural knowledge assessment. The combined classification is presented in Table 3.2, along with the techniques that will be presented in this chapter.

3.3 Knowledge Elicitation without Structural Representation

Structural knowledge elicitation techniques aim at the representation of knowledge structure (Jonassen et al., 1993). However, some of the abovementioned techniques do not initially produce a structural representation in the sense

Table 3.2: Classification of structural knowledge elicitation techniques.

		Elicitation of Edges				
Elicitation method	Elicitation of vertices	Labels	Hierarchy	Coefficients	Test form	Conscious processing
Word associations	Free (associations)	–	–	# of associations	Paper or Computer	High
Free term entry	Free (associations)	Binary (Yes/No)	No	# of Clusters	Computer	High
Card sorting	None (given)	Binary (Yes/No)	No	None	Paper	Low
Concept Grids	Cued	Weights	No	None	Computer	High
Concept Maps	Free (associations)	Freely labeled	Yes	None	Computer	High
SLT	Free (associations)	From set	Yes	Many (Bonato, 1990)	Paper	High
MaNET	None (given)	From set	Yes	Many (Eckert, 1998b)	Computer	High
Graph Building	None (given)	Weights	Yes	Euclidean distances	Paper	Low
PFNET	None (given)	Weights	Yes	Similarity C	Computer	Low
AST	Free (associations)	Weights	No	Many (Section 3.9)	Computer	High/Low

that they produce a network but produce a more coarse structural representation, e.g. clusters or groups. Since some of them provide basic or further information to other measures, they are of importance in structural assessment and are thus presented.

3.3.1 Word associations

Word associations belong to the set of recall techniques. "Recall techniques, whether cued or free, allow knowledge structures to emerge without constraining the individuals to a fixed list of concepts.... [T]hese procedures provide information such as the amount of organization as well as the depth of given structure...." (M. A. Davis et al., 2003, p. 203). In a word association task, participants have to produce associations to a given stimulus word representing the knowledge domain; the associated concepts have to be related to that domain. In Hoole's (2006) classification, word association tasks require little conscious processing, they lead to a non-hierarchic representation and the cuedness of recall is either fixed or free, depending on the mode of the task: Strube (1985) differentiates between verbal and written representation and recording of concepts and between cued or free associations. In free associations, all concepts coming to the mind can be associated spontaneously, while cued associations involve limits with regard to certain semantic relations to the stimulus concept. Although the number of association a participant produces for a given stimulus may indicate his or her magnitude of knowledge (Ceglarek, forthcoming), the amount of obtainable structural information from word association tasks is limited as the sequence of words is a one-dimensional form of data.

Jonassen, Yacci and Beissner (1993) see word association tasks primarily as a measure for elicitating concepts that are then subjected to further methods in order to determine a structural information. As a single method, word associations are unsuitable for the creation of knowledge structures, as they do not produce enough information for the creation of a network representation of knowledge; they can only represent knowledge structures as a chain where the concepts are placed in a row after each other in accordance to their occurrence in the flow of associations.

Free term entry

Strube (1984) suggested a mechanism to derive structural information from word associations by analyzing the time sequence in which word associations were uttered by experiment participants. He employed this information to cluster the associated contecpts into groups. A more recent example of such a method was presented by Beatty and colleagues (Beatty, 2000; Beatty & Gerace, 2002; Beatty, Gerace, & Dufresne, 2002). They employed Strube's technique in the following way: "Each response is recorded along with the time spent thinking of and typing it, in an attempt to capture the flow of concepts triggered in each participant's mind" (Beatty & Gerace, 2002, p. 751). Plotting the recorded thinking times between associations reveals typical characteristics (Beatty & Gerace, 2002): Associated terms that are semantically close follow quickly after each other, while a long thinking time between associations will typically indicate a smaller semantic similarity. This idea is based on the theory of spreading activation (J. R. Anderson, 1983), which again assumes a semantic network storage system for memory. The theory of spreading activation states that in a semantic network, the activation of a node (knowledge element or proposition) iteratively propagates out to adjacent nodes. Thus, activation of a concept will most likely activate an adjacent (= semantically close) concept, while a long thinking time is associated with a longer propagation of the spreading activity through the network. In his experiments with physics students, Beatty (2002) reported that his participants would typically provide him with clusters of terms that would come very quickly, interrupted by longer thinking times. Over time, these peaks in thinking times that separate terms that follow each other quickly would increase in length, whereas intra-cluster thinking times also increase. Figure 3.1 displays a typical pattern of thinking times between associations over time.

It is Beatty's explicit aim to elicit "the structure of interconnections between knowledge elements.... the conceptual knowledge structure" (2002, p. 3) through this technique. They hope to achieve this task by assuming "that the duration of pauses between term entries, and the grouping of term entries into clusters separated by longer follow periods, can reveal information about what terms a subject associates closely" (Beatty et al., 2002, p. 5). In Hoole's classification, the free term entry has the same characteristics as uncued word association tasks.

Although this approach is appealing due to its closeness to the participant's actual state of knowledge and because it does not operate on prespecified concepts, but on the participants' own associations, it is unable to reveal a structure in the actual meaning of the word as the resulting data, clusters of associated terms, is one-dimensional in nature (Beatty et al., 2002, p. 5). Other approaches have attempted to represent elicited knowledge as network structures in order to assess qualities of knowledge organization; examples are presented in the next sections.

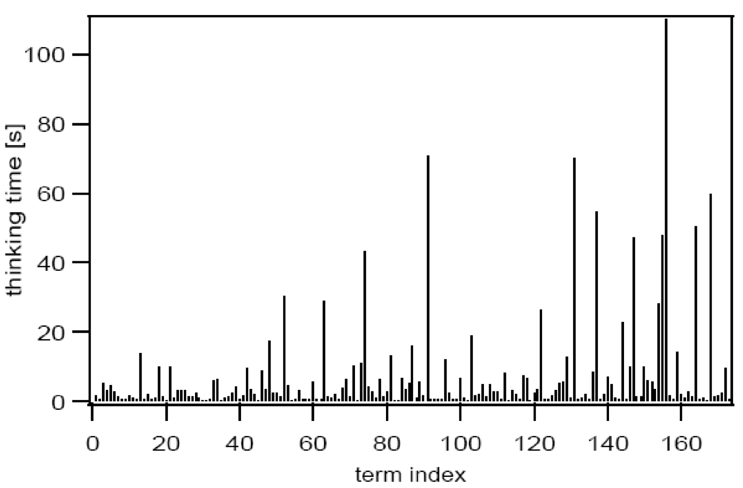


Figure 3.1: Participant01's thinking times between term associations in Beatty's first study (Beatty et al., 2002, p. 7). Reprinted with author's permission. The FTE approach would group the first 12 associated concepts into one cluster. Reprinted with permission.

3.3.2 Card sorting

According to Jonassen, Yacci and Beissener (1993), card sorting techniques "can be used to identify how concepts in a context are organized in a learner's knowledge structure" (p. 45). In card sorting, participants are presented with a set of cards containing words and, optionally, a brief definition. Participants are then asked to sort the cards into stacks or groups according to their similarity and to label these groups (Hirshman & Wallendorf, 1982). The assessment of groups and their associated titles can reveal qualitative features of a participant's knowledge organization on group level. In Hoole's framework of structure elicitation techniques, card sorts require a high degree of conscious processing, they are based on cued recall and do not lead to a hierarchical representation of structural knowledge.

It is also possible to represent card sorting results as a graph by treating concepts as nodes and by placing an edge between two nodes if they were placed in the same group. However, this procedure will not lead to a connected graph encompassing all concepts but to several undirected unweighted graphs – one for each group – that each include the concepts from one group. Thus, a graph representation in this case does not reveal more information than the stacks of cards. This could be changed by altering the instructions in such a way that participants form sub categories with their stack of cards which would lead to a connected and hierarchical graph representations of the concepts.

3.3.3 Summary

The three abovementioned techniques – word associations, free term entry and card sorting – are among the simpler techniques for knowledge elicitation. The first two produce a serial flow of concepts that the participant assumes to be related to a given knowledge domain. Except for card sorting, they do not produce a structure in terms of a graph representation (other than a line).

Card sort creates a graph-representation of knowledge structure but also has its shortcomings, because the method does not account for different association strengths between two concepts (concepts are either related or not). While the first two techniques might well serve as techniques for generating concepts for structural assessment, neither of them appears to be adequate for carrying out the complete task. However, the free term entry method has one advantage that does deserve a special mention: It tries to derive structural information implicitly through reaction times, while card sorts rely on explicit judgments of relatedness. The indirectness of the deduction of a coarse structure through reaction times has undergone little further research since its publication and seems promising for further evaluation.

3.4 Concept Grids

Concept grids or the Repertory Grid Interview is a procedure that combines interview techniques with a structural approach and is used for knowledge elicitation for expert system design (Bonato, 1990; Clases, 2007; Fromm, 2004). It is based on Kelly's (1955) personal construct psychology (PCP), which is used as a method for eliciting personal construct systems.

Construct systems are a set of bipolar dimensions used to classify subjective perception of the environment. According to Kelly (1955), every object entering cognition is classified among a certain set of constructs (e. g. intellectual/not intellectual or critical/accepting when assessing a co-worker). Construct systems are used to make predictions about future events or states. Kelly sees this anticipation and prediction of events as the main driver of the mind in order to better predict actions of other people.

Kelly developed a technique for the elicitation of personal construct systems, the Repertory Grid Interview. A repertory grid is a matrix where rows represent constructs (the opposing poles describing the construct are placed on opposing sides of each row), and columns represent objects, typically other individuals, that are rated on each construct. The cells of the matrix indicate the rating of an object on a particular construct. Ratings can be either nominal (high/low) or ordinal (a number indicating the position of an object on the construct).

In the repertory grid interview, constructs are elicited in an indirect way by firstly asking the interviewee to name a set of important people he or she considers of high importance to him or herself. The interviewee is then presented with a triplet from this set and is asked to name a feature that only two of the three persons possess. The shared feature is used as the construct and the feature that differentiates the third person is used as its contrast. In a final step, all other persons from the set are rated on the construct-contrast dimension by the interviewee.

Kelly's interest was primarily in the clinical field, but the repertory grid interview has been recommended as knowledge elicitation tool by several researchers (compare Gaines & Shaw, 1992). The computer-based AQUINAS expertise transfer system (Boose, 1985; Boose & Bradshaw, 1987) uses the Repertory Grid Interview to elicit expert knowledge; cluster analyses identify construct hierarchies and are thus capable of delivering a knowledge structure (Bonato, 1990, p. 23).

Accordingly, Repertory Grid Interviews have recently been suggested as a method for implicit knowledge elicitation by Clases (2007). He argues that implicit knowledge is "an integral part of symbolically mediated social practice thus seeing the implicit and the explicit as closely connected in human discourse, bridging individual and social aspects of knowledge" (Clases, 2007, p. 286). Since social practice is mediated by construct systems, he continues, repertory grids are capable of tapping into the implicit dimension of knowledge.

In Hoole's framework, the repertory grid interview can be classified in the following way: it requires a high level of conscious processing, involves cued recall and leads to a non-hierarchic structure.

Disadvantages

Construct grids refer to fundamental underlying constructs of an individual. The theory of these constructs as put forward by Kelly (see above) is difficult to relate to the notion of structural knowledge, because it bears no explicit reference to concepts such as declarative, procedural, or structural knowledge. As constructs are thought to guide our assessment of objects in cognition, an unconscious process according to Kelly (1955), constructs in this sense appear to lie deeper in the cognitive system then memory contents, or, in the wording of the dimensional model, are placed further towards un-codifiable on the codifiability axis than structural knowledge in Jonassen's sense (compare Section 2.5). According to the way Kelly (1955) put it, and also in the sense employed by Clases (2007), constructs are more about views of the world or views of objects than knowledge of fact and procedures. Furthermore, construct systems are difficult to elicitate as they require an own interview technique.

After their elicitation, individual construct systems are subjected to a multidimensional scaling technique (for a example, see Clases, 2007). The resulting form is not a graph, but a planar space on which constructs are placed, and vicinity between constructs and objects indicates relatedness. The aim of these representations is to identify meaningful dimensions that can account for the differing distributions. Thus, the aim is not the elicitation of structure itself, but the identification of factors that led to the identified construct structure.

Because of its conceptual distance from the concept of structural knowledge in the abovementioned sense, its elaborate elicitation technique and its representational format, concept systems appear to be only marginally suited for structural knowledge assessment.

3.5 Concept Mapping and Structure Laying Techniques

Concept mapping techniques and structure-laying techniques are closely related to network models of human long-term memory (Bonato, 1990) and are used for knowledge assessment or learning support (ibid). Due to their similar assumptions and elicitation procedures, they are presented in the same section. They both belong to the class of procedures for graphical knowledge network representations that represent knowledge as a graph.

3.5.1 Concept maps

According to Novak and Cañas (2006),

> Concept maps are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts. Words on the line, referred to as linking words or linking phrases, specify the relationship between the two concepts. We define *concept as a perceived regularity in events or objects, or records of events or*

> *objects, designated by a label.* The label for most concepts is a word, although sometimes we use symbols such as + or %, and sometimes more than one word is used (p. 1).

Concept maps are presented in a hierarchical fashion with the most general concept at the top (ibid). However, their representation does not follow the formalism of a directed graph with labeled edges and nodes, because edge labels can further branch edges like in the case of the "are" label between concepts and propositions in Figure 3.2 that illustrates an example of a concept map.

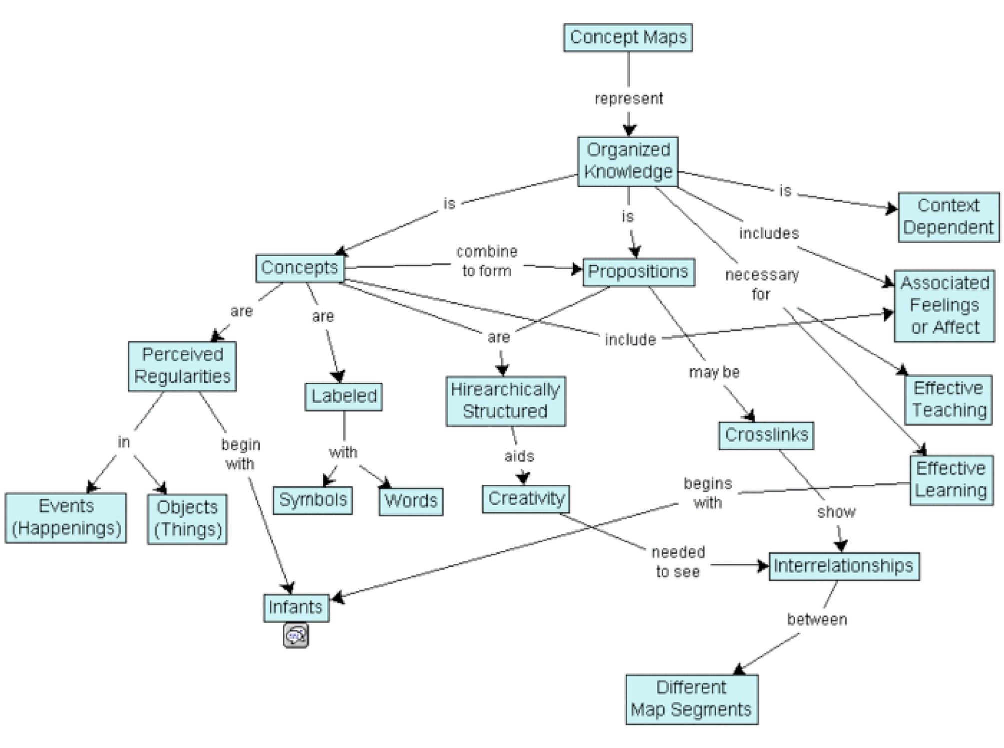


Figure 3.2: Example of a concept map. From Wikipedia, the free encyclopedia (http://en.wikipedia.org/wiki/Image:Conceptmap.gif)

Concept maps were developed by Shavelson in the 1970s (Shavelson, 1972, 1974; Shavelson & Stanton, 1975) and can be elicited in various ways, for example with computer programs or by sticking post-its to a whiteboard (Novak & Cañas, 2006, p. 10). Note that in concept maps, the person constructing the map labels concepts and relationships freely. Thus, according to Hoole's specification, concept maps employ free recall, require a high amount of conscious processing and lead to a hierarchic form of representation. According to Eckert's classification (compare previous section), the users can label both edges and concepts freely.

Concept maps are primarily employed for learning purposes (Novak & Gowin, 1996). Novak and colleagues "believe one of the reasons concept mapping is so powerful for the facilitation of meaningful learning is that it serves as a kind of template or scaffold to help to organize knowledge and to structure it (...)" (Novak & Cañas, 2006, p. 7).

Criticism

Although Novak argues that concept maps resemble structures in human memory (Novak & Cañas, 2006), the way he describes their construction and use (Novak, 1998; Novak & Cañas, 2006; Novak & Gowin, 1996) makes it clear that concept maps are not intended for psychometric elicitation of an individual knowledge structure in terms of a test, but are intended either for the acquisition of a formally agreed and validated knowledge structure or teaching evaluation. For example, a teacher will present a set of concepts to students and ask them to build a concept map from it; the resulting map is then used to assess the student's learning performance (Novak & Cañas, 2006). Another example are several individuals working on a concept map over a longer period of time in order to create a correct representation of a formal knowledge structure for teaching and learning purposes. The possibility to elicit students' subjective knowledge structures with concept map does point towards the possibility of their usage for structural assessment purposes. However, the tool for concept mapping, CMapTools (Cañas et al., 2004) does not include the possibility to analyze concept maps in any quantifiable way (i.e. density, similarity measures, number of edges or nodes etc). The idea of using concept-map-like structures for psychometric knowledge elicitation has been pronounced by other researches; two examples, the Structure-Laying-Technique and its computer-based further developments are introduced in the next sections.

3.5.2 The structure laying technique (SLT)

The 'Heidelberger Strukturlegetechnik' SLT (Scheele & Groeben, 1984, 1988) is a standardized manual concept mapping tool for structural knowledge elicitation which builds on Shavelson's concept mapping techniques (compare previous section). Originally, SLT has been developed for the elicitation of subjective theories. In its original application as described by Scheele and Groeben (1984), a semi-structured interview was employed for the identification of relevant concepts in a certain domain, which are written on cards.

In a second step, the participant has to determine the relationships between concepts. In order to do so, he or she is trained in the application of a set of different types of relations. These include semantic relations such as 'equal', 'sub-category', 'superior category', 'and-relation', 'or-relation', 'manifestation', and 'indicator'. The remaining relation types symbolize empiric relations between variables as found in empiric psychology, e. g. negative and positive causal relations, interaction effects with certain preconditions, second-degree polynomial relationships, correlations etc (for further details, see Scheele & Groeben, 1984). In Hoole's specification, SLT employs free recall, require a high amount of conscious processing and lead to a hierarchic form of representation due to the directedness of available links. According to Eckert's classification (compare Table 3.1), SLT employs freely labeled nodes and given edges.

Bonato (1990, p. 33) assets three main advantages of SLT, which also apply to concept mapping: Firstly, the visualization of concepts and relations allows to easily reconstruct ideas and views of the world. Secondly, the directness of the survey: In opposition to construct grids and word association tasks, relations between concepts are directly elicited and are not inferred based on complex calculations, which only indicate that there is a relationship between two nodes, but not the type of relation. Due to the fact that the relations are chosen from a fixed set, the researcher does not have to interpret a freely labelled connection as is the case in concept mapping (see above). Finally, Bonato argues that the easy manipulation of networks during elicittion pays tribute to the fact that participants usually do not have a complete knowledge network consciously available at the time of the survey, but activate new concepts and relations during the elicitation process.

3.5.3 Graph theoretic description of SLT-elicited knowledge networks

Bonato (1990) introduced the idea to treat SLT-elicited knowledge networks as graphs in the mathematical sense and to use graph-theoretic concepts for their description and for performance assessment. This delivers variables of metric scale that can be used for describing several features of a participant's knowledge network and are thus thought to give a numeric representation of the scope and complexity of a participant's structural knowledge. Based on White's (1985) dimensions describing the cognitive structure of a person's cognitive structure, Bonato suggests a number of graph-theoretic measures describing the structure of an SLT-elicited knowledge network (represented as directed graph) on graph level.

In order to determine the quantity or perimeter of structural knowledge, the number of vertices (also referred to as graph order, compare Section 3.1) and the number of edges (also referred to as graph degree) are used to describe the perimeter of the knowledge graph. Bonato further suggested using the number of unconnected subgraphs, where a number larger 1 represents a less well-integrated knowledge structure, as a measure for knowledge integrity and the density of the graph as a measure of complexity. The density of a graph is calculated by dividing the number of present edges by the number of all possible edges inside a given graph. West (1985) refers to the same measure as Integrated Propositional Knowledge (IPK) in the description of complexity of cognitive structure.

The diameter of a knowledge graph (compare Definition 6 in Section 3.1) is seen as a further valid description for the magnitude of a knowledge graph (Eckert, 1998a). In accordance with Strube (1985), Bonato (1990) suggests a further metric relying on the diameter measure, the Relative Maximal Path Length (RMPL). It is defined as the longest path length inside a knowledge network divided by the maximum possible path length inside a given knowledge network with n nodes. This equals the diameter divided by n - 1. "If RMPL = 1, the knowledge network consists of a single chain. The minimal value of RMPL is $1/(n-1)$. In this case, the diameter is 1 and the knowledge network has a star-shaped structure. Thus, RMPL allows a statement with regard to the compactness of a knowledge network: The smaller the value of RMPL, the more compact the knowledge network" (Bonato, 1990, p. 96, own translation).

The median and mean of distances between all vertices inside a graph are used as a further measure to determine the compactness of a graph. In an empiric study on classroom learning, Bonato elicited students' structural knowledge on genetics and a teacher's reference network with an SLT according to Ballstaedt and Mandl (1985). He then had the teacher assign grades to the student's networks and grouped students' networks into four clusters based on their structural features. A discriminant analysis with the above coefficients and a coefficient based on network similarity between student and teacher network was able to deliver a correct classification percentage of 93.2. Bonato then analyzed whether the clusters differed with regard to the average grade of the cluster and found significant differences in grades between two clusters. His results can be seen as a support for the notion that performance prediction based on graph-theoretic coefficients of structural knowledge graphs is possible. However, methodological criticism can be directed at his way of calculating results. Despite the fact that he had an outside criterion available that captured knowledge in the same domain as his networks intended to (a multiple choice questionnaire), he refrained from correlating the graph-theoretic network scores he obtained with them, which casts doubts on the validity of his measures. Furthermore, he did not correlate the graph-theoretic coefficients with the grade of the networks, depriving himself of another opportunity to establish internal validity.

3.5.4 Computer-based structural assessment of concept maps with MaNET

Partly based on Bonato's findings, Eckert (1998b, 1998a) developed a computer-supported structure laying technique which explicitly assumes that certain graph-theoretic variables can be used for individual performance assessment and prediction. His method, the Mannheimer Netzwerk-Elaborierungstechnik MaNET (Mannheim network elaboration technique) combines computer-based tools for expert-based generation of concepts and edge relations, for reference network generation, for computer-based participant network generation, and for graph-theoretic quantitative analysis of knowledge networks.

With MaNET, Eckert hopes to improve economy and feasibility of execution and analysis of the structure laying technique and to overcome the shortcomings of traditional concept mapping, e.g. the pen-and-paper format. Its main feature is the integrated automatic quantitative analysis. From a participant's perspective, the MaNET test belongs to the class of procedures for graphical knowledge representations where both concepts and edges between concepts cannot be labeled freely, but have to be chosen from a fixed set (Quadrant IV in table 3.1). In Hoole's specification, SLT employs free recall, requires a high amount of conscious processing and leads to a hierarchic form of representation due to the directedness of available links.

Since the set of general semantic edge types in SLT procedures as outlined above proved to be difficult and costly to handle and required extensive amount of training on the side of the participants, Eckert replaced the standard set of semantic relations with a specific set for each knowledge domain. Therefore, prior to testing, a team of experts (domain experts, diagnostics and cognitive psychologists) identifies relevant concepts from a certain domain and appropriate relation types for the domain. This procedure takes place on screen in a special configuration module of the MaNET system. These concepts and edges are then presented to the participant. Thus, in Hoole's specification, MaNET employs no recall as participants are presented with the complete set of concepts and edges, requires a high amount of conscious processing, and leads to a hierarchic form of representation due to the directedness of available links.

MaNET coefficients

In the actual test module, concepts and relations are presented to the participants who then assemble a concept map representing their knowledge from the provided elements on screen. The semantic meaning of a relation is chosen from the available set, and direction and association strength are entered. For each participant, an analysis module automatically computes five graph-theoretic coefficients: The number of edges (the number of links in the graph), the density (the number of edges divided by the number of edges possible in the participant's network;), ruggedness (the number of unconnected subgraphs), the diameter (the longest path in a network) and the node centrality for each node (see below).

According to Eckert (1998b), a higher density hints at high degree of knowledge interconnectedness and a low ruggedness indicates a less rugged, highly integrated knowledge structure. The diameter is treated as an indicator for the quantity of structural knowledge. The centrality of a node is supposed to indicate the centrality of the meaning of that node in a given knowledge structure. Two centrality measures are calculated: The indegree and outdegree (compare Definition 2 in Section 3.1). The degree D of a node u specifies to how many other nodes in the network a direct relationship exists. D is specified as indegree ID (the number of directed edges pointing towards u) and outdegree OD (the number of edges originating from u). For undirected graphs, OD equals ID.

Between all networks that have been elicited with MaNET, similarity can be calculated based on edge similarities between two networks. This comparision can be made between participants' networks or between a participant's network and a MaNET-elicited expert network. The employed comparison analysis requires that both networks consist of the same concepts (nodes) and the same types of edges, which is the case for two MaNET-elicited networks that have been constructed in the same knowledge domains, as relations and concepts are pre-specified in the configuration module (see above). The similarity between two networks, also referred to as correspondence is given in the correspondence index C which is dissimilar to the similarity measure C used in Pathfinder networks (compare next section). MaNET's C is based on comparing all possible node pairs between two networks. It is thus a similarity measure based on path features and not on structural features as Pathfinder's C. MaNET's C is calculated by adding the number of occurrences where two nodes have the same edge between them in both networks and the number of occurrences where the same two nodes have no connection between them in both networks. From this sum, the number of mismatches for both categories is subtracted. The result is divided by the number of possible connections.

This calculation is adjusted for edge weights, resulting in a coefficient referred to as C_w. Eckert also suggests a significance measure u for C_w which is based on the null hypothesis that the sum of correct matches between two networks is equal to the sum of incorrect matches. However, the author states that such a measure will most likely overestimate significance and refrains from using it in his further analyses. Eckert also suggests a method for aggregating a network from several individual networks, so-called modal networks. This is done by including those edges in the modal network that occur above a certain threshold in all the included networks. The threshold is based on the average number of links in the included networks (see Eckert, 1998b, p. 96 for details).

Although Eckert (1998b) conducted an experiment in whih he obtained students' MaNET scores, he does not correlate them with outside performance criteria such as the students' grades although they were available to him. The validity of these measures for performance prediction is thus questionable.

Criticism

Some of the aspects Bonato (1990) mentions as advantages can also be seen as disadvantages of the SLT technique. First of all, learning the 18 different relation types is a demanding and extensive task and a study shows significant variance in label type choice between participants that are experts in the same domain and have to construct a concept map of a small sub-domain of their expertise (Rothe & Warning, 1991). Rothe and Warning showed that their study participants were able to use relations between concepts consistently (by being able to employ analogies correctly), but were unable to name the edges. If an edge is placed in SLT by a participant that appears to be wrong to a domain expert, it is thus impossible to determine whether the participant has wrong knowledge or whether the participant uses a relation correctly, but is unable to label it. This finding is a particularly strong indicator for the implicit aspect of structural knowledge and justifies the assumptions made in Chapter 2.

The fact that participants do not have a complete knowledge network consciously available but construct one as an interactive process during the elicitation procedure casts doubts on the suitability of SLT as a test in a psychometric understanding of the word. An assessment technique is supposed to elicit a construct, and is not supposed to construct and measure a construct in one process.

Despite the above criticism, SLT is seen as a reliable and valid tool for knowledge assessment and performance prediction (Ballsteadt & Mandl, 1985; Bonato, 1990), because a change in learning is reflected by a change in SLT-elicited networks in terms of the number of concepts and a difference in structure of the networks.

Despite the advantages of a computer-based structure laying technique with regard to its economy of the actual test process, a few aspects of the method can be criticized. First of all, the demands for the creation of test items (concepts and relation types) and the reference network are high; Eckert suggests an entire group of specialists for performing these tasks. This demand is likely to counteract the saving in complexity that is achieved due to the modularized computer-based set-up of the method.

Secondly, in the development of his correspondence coefficient C_w, Eckert ignores the wide body of graph-theoretic research on graph similarity algorithms. Goldsmith and Davenport (1990) explicitly suggest a correspondence coefficient based on structural methods for knowledge network comparison and discourage the use of path-based similarity measures. Today's algorithms are able to determine graph similarity based on the structural comparison of subgraphs (Le, Ho, & Phan, 2003), whereas Eckert's measure does not take any structural information beyond dyads of nodes into account. Finally, Eckert's prime measure for performance prediction C_w is not an absolute measure based on norms, but a relative measure with reference to an expert network. This means that for every structural assessment, an expert and his or her reference network must be available and experts themselves cannot be tested at all. Furthermore, the requirement for a specific expert network for comparison makes it impossible to compare MaNET's C_w scores across domains, because each knowledge domain has its own frame of reference (its expert network), and the validity of the expert network is largely unaddressed. Furthermore, the test implicitly assumes that all participants are familiar with the concepts and edge relations presented to them, which cannot necessarily be taken for granted. In the case where a participant does not know what a concept or an edge means he or she will place these elements in a arbitrary way which again casts doubts on the validity of the resulting network. Finally, the central criticism that applies to concept mapping and structure lying techniques can also be applied to MaNET: Presenting a participant with the task of constructing an entire network is likely to initiate a process whose result is unknown to the participant at its onset. This process adds an unnecessary amount of variance to the aim of psychometric knowledge diagnosis. Other methods try to address this issue by dissecting the task of network construction in several less complex and demanding subtasks. These methods are presented in the following sections.

3.6 Graph Building and Multidimensional Scaling

The graph building method is based on Rapoport, Livant and Boyd (1966) and on Fillenbaum and Rapoport (1971). In this approach, participants are presented with an alphabetic list of expert-defined stimulus concepts and are asked to construct a graph out of the stimulus concepts by drawing lines between those they see as similar. Note that the main difference to SLT and concept mapping techniques is two-fold: Edges between concepts are not labeled but only have a connection strength attributed to them, and edge construction only occurs locally between two concepts at a time; the participants do not construct an entire graph but perform similarity ratings between two words. In Hoole's classification, graph building involves no recall except for edge relations, which are cued. The process involves a smaller degree of conscious processing than SLT, and the resulting network representation is hierarchic. The resulting raw data of graph building is a concept $\times$ concept matrix that is the starting point for factor analyses or multidimensional scaling. These procedures aim at discovering structures within the set of stimulus concepts.

"Metric multidimensional scaling (MDS) transforms a distance matrix into a set of coordinates such that the (Eucledian) distances derived from these coordinates approximate as well as possible the original distances" (Abdi, 2007, p. 598). It is often used for visualizing dissimilarities in data, as it can extract the latent structure within empirical similarity judgments (Schvaneveldt et al., 1985, p. 705). In order to obtain such a representation, the similarity data is submitted to a MDS algorithm. Analogous to principal components analysis, MDS detects the locations that minimize distortions to the distance matrix. Its goal is to find a Euclidean distance approximating a given distance (Eigenvector problem). For more information on MDS calculations, refer to Abdi (2007). Figure 3.3 shows the result of a two-dimensional MDS scaling procedure on similarity ratings performed by 16 students between 16 natural concepts (Schvaneveldt et al., 1985 p. 707)[1].

A problem with MDS representations of similarity data is the fact that these are very global in nature and difficult to compare and to describe in a metric way. Schvaneveldt (1985) thus suggested to employ a method that "extracts the latent structure rather than transforming the data" in

[1]This article was published in the International Journal of Man-Machine Studies, Vol 23, Roger W. Schvaneveldt and Francis T. Durso and Timothy E. Goldsmith and Tomothy J. Breen and Nancy J. Cooke and Richard G. Tucker and Joseph C. De Maio, Measuring the structure of expertise, 699–728, Copyright Elsevier (1985).

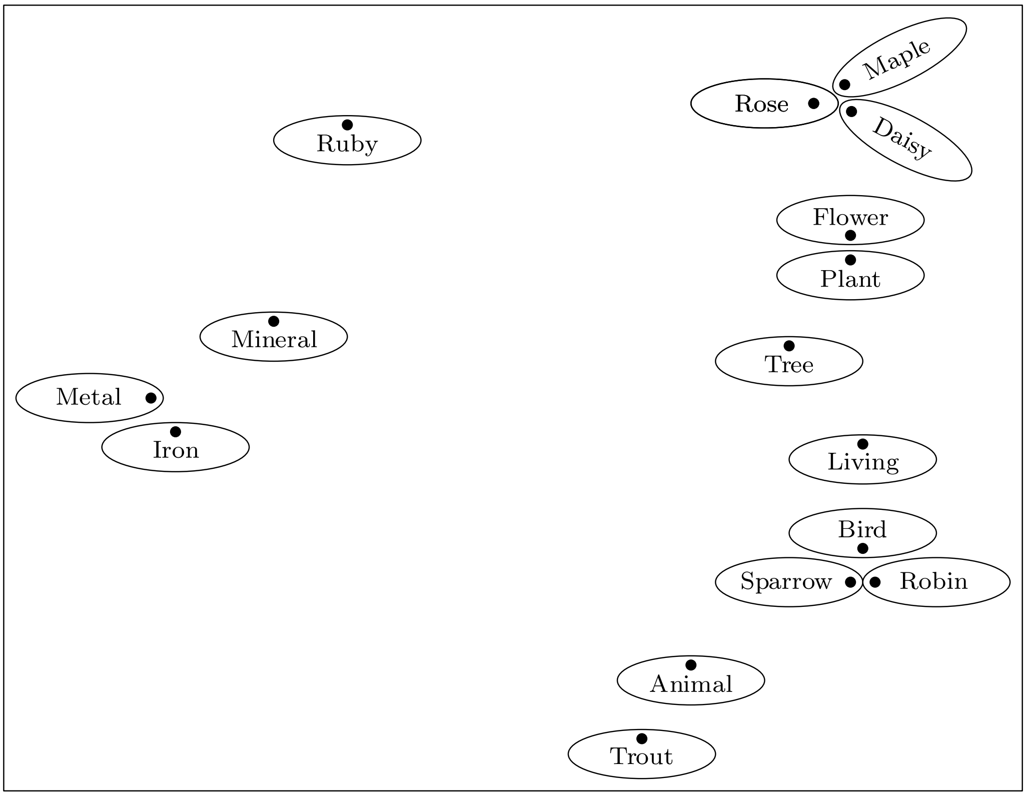


Figure 3.3: Two-dimensional MDS solution for 16 natural concepts, assessed for their similarity by 16 students (adapted with permission from Schvaneveldt et al., 1985, p. 707). The horizontal dimension reflects a living-non-living dimension; the vertical dimension captures a difference between plants and animals for living concepts.

order to "specify the local relations and structure present in a conceptual organization" (Schvaneveldt et al., 1985, p. 709). The method he suggested for doing so is outlined in the following section.

3.7 Structural Assessment with Pathfinder

Pathfinder is a method for arriving at a network representation of empirical proximity data (Schvaneveldt, 1990, p. ix). It was created to transform subjective pairwise ratings of concept-relatedness (similar to graph building, compare previous section) into graph-representations of knowledge networks for psychometric assessment. In order to transform proximity data into network representations, proximities between elements are considered as path lengths in a graph-theoretic point of view (compare Section 3.1). The proximity data supplied to the pathfinder procedure typically stems from "...a participant's estimate of the similarity of each pair of entities in the set" (Dearholt & Schvaneveldt, 1990, p. 2), where set refers to a set of concepts from a certain knowledge domain that has been identified as crucial and representative for that domain by one or more domain experts. The similarity rating between two concepts is interpreted as an edge between the two concepts that are treated as graph nodes. If the participant rates a pair of concepts as dissimilar, there is no edge between the concepts. If the participant can rate the degree of similarity between the concepts, the rating is transformed into a distance label for the edge between the two nodes: A high similarity results in a short distance, a low similarity rating results in a longer edge.

Graphs constructed in this way typically violate the triangle inequality which states that for any subgraph with triangular structure, the measure of a given edge between two nodes must be less than the sum of the other two edges but greater than the difference between the two paths. The shortest path between two nodes inside a graph (spanning several individual edges) is called a geodesic path. A triangle inequality violation is never part of a geodesic path, because there is always a shorter path available. "The omission of the edges which violate some triangle inequality in a network assures the preservation of the (geodetic) distances between all pairs of nodes and provides a simpler structure which possesses precisely those edges which are responsible for the most economical paths" (Dearholt & Schvaneveldt, 1990, p. 2). The Pathfinder method generates a class of networks (PFNETS) from distances between pairs of entities that omit links that violate a triangle inequality for at least a specified number of links (usually two). Distances between entities in a PFNET that are not connected in the original proximity data are calculated on the bases of path lengths between them. PFNETs can either be directed or undirected. For PFNET generation, Pathfinder uses two parameters, q and r. q limits the number of indirect proximities investigated during network generation. q is a value between 2 and $n-1$ where n is the number of nodes in the network. The r parameter defines the metric on which the computation of the distance of paths is based. r is a real number between 1 and infinity, inclusive. A network generated with particular values of q and r is called a PFNET(q, r). A PFNET$(n-1, \infty)$ will be a minimal spanning tree for the edges supplied by the proximity data. For a given graph, its minimum spanning tree is a spanning tree with weight less than or equal to the weight of every subgraph which connects all the vertices in the graph together. If not specified differently, PFNETs generated in the course of this work will have the q, r parameters $n-1$, ∞ by default.

In the practical generation of PFNETs for a certain knowledge domain, domain experts identify a set of concepts that are of high relevance to the knowledge domain in question (similar to MaNET, see above). Two concepts from this set are presented at a time, and participants have to make similarity ratings between them, usually on a five- or seven-point likert scale (Goldsmith & Johnson, 1990). Since these similarity ratings are treated as undirected edges, a participant has to perform $n(n-1)/2$ ratings for a set of n concepts in order to compare each concept to all other concepts in the set. The data generated in this way is then subjected to the pathfinder algorithm for adjustment, which again results in the PFNETs.

A central aspect in Pathfinder-based approaches to structural assessment is the assessment of structural similarity between PFNETs, as performance prediction is based on the similarity between a participant's PFNET's similarity to a reference PFNET elicited from a domain expert. The idea behind this comparison is that the more similar the knowledge structure of a participant to the one of an expert, the more likely this participant will display a similarily high performance than the expert in the particular domain. For the comparison of PFNETs, Goldsmith and Davenport (1990) suggest a range of similarity measures C (Goldsmith & Johnson, 1990). In a subsequent analysis of prediction performance, a specific similarity measure C_1 was identified as possessing the best predictive validity. it "employs a set-theoretic method to compare corresponding neighborhood regions of two networks. The method computes for any two networks a single quantitative index of closeness called C" (Goldsmith & Johnson, 1990, p. 246). "A neighborhood about some node, v, is defined to be the set of nodes that are within distance one from v, excluding v itself" (Goldsmith & Johnson, 1990, p. 77). When comparing two graphs G and G' with the same number and type of nodes such as PFNETs, the neighborhood similarity between $v \in G$ and $v' \in G'$ is calculated by dividing the cardinality of the intersection of the node neighborhoods by the cardinality of the union of the neighborhoods. Simply speaking, the number of nodes that both v and v' link to in G and G' is devided by the total number of nodes that v and v' link to. The mean of all node-level neighborhood similarities between G and G' is the measure of overall graph similarity C. In a study predicting final course points of 40 college students in a college course on psychological research techniques, C_1 similarity measures between students' PFNETS$(n-1, \infty)$ and the instructor's PFNET correlated at $r = .74$ (Goldsmith & Johnson, 1990).

PFNETs $(n-1, \infty)$ applied to subjective similarity ratings of expert-identified concepts have been used in a wide range of studies for the assessment of structral knowledge. In an early application of the Pathfinder method (Schvaneveldt et al., 1985), the structure of PFNETs was able to differentiate between novice and expert pilots. Pathfinder networks have been identified as predective of performance in problem solving (Rowe & Cooke, 1995), they have been used successfully to assess training outcomes (M. A. Davis et al., 2003), and to assess individual and team mental models in computer environments (McDougall, Curry, & Brujin, 2001). Generally, the C coefficient between participants' and experts' PFNETs has been successfully applied for performance prediction in a wide range of studies (Acton, Johnson, & Goldsmith, 1994; Goldsmith et al., 1991; Day, Jr, & Gettman, 2001; Goldsmith & Johnson, 1990; Johnson, Goldsmith, & Teague, 1995; Wyman & Randel, 1998). According to Hoole's (2006) classification, PFNETs require no recall except for edges, require a low level of conscious processing and result in a non-hierarchical representation due to the fact that edges are undirected. Despite their usefulness, two issues of PFNET-based structural assessment can be identified: the need for a reference network and the demand for a fixed set of concepts that represent the knowledge domain.

Because Pathfinder is based on fixed concepts that are presented to all study participants, there is no way to determine whether the participants know anything about the identified concepts. It is simply assumed that the concepts have a meaning for the participants, but this assumption is never tested and has so far remained unchallenged. Thus, it cannot be taken for granted that the similarity ratings between presented concepts have any reference to the presented words because the familiarity of the participant with the words remains unclear and is thus questionable. The reliance on expert-identified concepts presents a trade off in comparison with free recall techniques, as they allow a deeper understanding of a participant's knowledge, as Davis and colleagues point out:

> Recall techniques, whether cued or free, allow knowledge structures to emerge without constraining the individuals to a fixed list of concepts...these procedures provide information such as the amount of organization as well as the depth of given structure, features of structural knowledge that cannot be captured by similarity judgments alone (M. A. Davis et al., 2003, p. 203-204).

Secondly, the C measure requires an expert network for reference. However, situations in which no such network is available appear plausible. Furthermore, the validity of expert networks is widely unaddressed. What qualifies an expert to provide his or her similarity ratings and what happens if networks among several experts in one domain differ? In their study, Davis and colleagues (2003) had two different expert networks available that had been generated by two domain experts. One was based on the consensus of the two experts while the other one was an average-connection structure statistically derived from the two different networks. The PFNETs of 50 participants were compared to both networks individually, and the statistically averaged referent structure provided the best predictive values for C. This finding underlines the validity issue of single expert networks as referent structures. Accordingly, a structural assessment technique that operates on norms instead of expert referent structures appears desirable (Bonato, 1990).

3.8 Discussion of Methods and Requirement for a New Technique

In order to assess structural knowledge in a psychometric way, only those methods that deliver numeric descriptors of a knowledge graph are appropriate for structural knowledge elicitation because such variables can be used for performance prediction and correlation with outside criteria. Thus, four of the presented methods are suitable: SLT, MaNET, Graph Building, and PFNET (compare Table 3.2). However, all of these methods have their weaknesses (compare according sections): SLT is cumbersome to perform due to its exclusive availability in paper-and-pencil form, it is a complex procedure prone to high variance, the set of available relation types is difficult to learn, and it possesses only limited validity. MaNET possesses too many restrictions by presenting the participants with a fixed set of concepts and relations without any means of assuring whether the participants actually know the material they are supposed to use for graph construction. MaNET also relies on expert networks as referent structures with questionable validity and challenge the participant with the task of constructing an entire network in a single process. Graph building produces MDS representations of semantic concept proximity which are difficult to compare and to describe.

Pathfinder Networks reduce the cognitive workload by constructing a graph based on the participants' responses to a series of items (similarity comparisons between expert-identified concepts). It thus reduces complexity and explicitness of the graph construction procedure. The latter feature is of particular importance. In the dimensional model of memory types (compare Figure 2.8), it was stated that implicit access to structural knowledge is part of individual implicit knowledge. Unconsciously available relations between concepts do most probably not carry a readily available explicit label: If a structural knowledge assessment technique is supposed to capture at least a part of unconscious access to structural knowledge, it must feature quick and intuitive similarity comparisons that neither require an explicit label, nor a complex process for their elicitation that involves a high degree of conscious processing. Thus, only the pathfinder methods seems to have a (limited) capability to tap into relations that cannot nessessarily be labelled by a person: Participants are asked to make single quick and intuitive ('gut-feeling') judgements on the strength of relatedness between two presented concepts. The fact that pathfinder networks have been employed in domains where implicit knowledge is likely to influence performance, such as aircraft piloting (Schvaneveldt et al., 1985), underlines this assumption. However, Pathfinder networks do possess two characterisics that limit their use: they rely on expert networks for network assessment and they rely on a set of specified constructs. There is no way of validating whether a participant actually knows the constructs presented to him or her, and the validity of expert structures is widely unadressed, as outlined above.

These issues require a new technique for structural knowledge elicitation that overcomes the abovementioned issues. It should possess the following features: In order to enable a high level of ease-of-use, it should b computer-based. It should involve pathfinder-scaling of structural knowledge networks, as pathfinder-scaled networks have been successfully apllied for performance prediction in a variety of uses (see above). It should be based on concepts that the participant knows and associates with the knowledge domain in question. It should further reduce the amount of required cognitive processing by dissecting the procedure of graph construction into a set of response items as in graph building. It should further exploit multiple sources of information for data generation, i.e. explicit measures supplied by the user, and, if possible, indirect measures such as response time as outlined in Section 3.3.1. Finally, it should deliver graph-theoretic values that quantify the structure of an elicited knowledge graph for performance-prediction purposes. Such values should be obtained without the necessity of expert structures due to their validity issue. That such general coefficients are available and are of value for performance prediction has already been demonstrated by Bonato (1990), compare Section 3.5.3.

In summary, the elicitation of structural knowledge in such a way that edges are elicited that a participant cannot label but assumes present. requires a computer-based test that reduces the elicitation of a knowledge structure into subtasks. The test should also allow implicitness (i.e. unlabeled edges), should employ a mixture of indirect (time) and direct measures, should produce coefficients, and should not require an expert network for their creation. Such a knowledge elicitation technique is presented in the next section.

3.9 The Association Structure Test (AST)

The newly proposed method integrates an association task and pathfinder network similarity ratings into one IT-based test system[2]. Association techniques can measure some aspects of individual's structural knowledge that are potentially missed if only similarity ratings are adopted. An association task assures that the concepts from which a structure is built actually stem from the memory of the specific participant. In this association task, the participant is asked to associate concepts that he or she thinks belong to a specified knowledge domain analogue to the FTE technique (compare Section 3.3.1). These associated concepts are then fed into a pathfinder structural assessment technique based on pairwise concept similarities to which the PFNET $(n-1,\infty)$ algorithm is applied (compare Section 3.7). The concepts that enter the comparison section are those that were freely associated to a stimulus concept by the participant. In this way, the benefits of pathfinder networks are preserved, and the issue of determining whether a participant actually knows a concept presented during pairwise comparisons is solved. This design also offers the advantage of deducting several quantitative measures for knowledge organization (see Section 3.9.3) that can be used for correlation or prediction purposes. By using quick and intuitive judgements of concept relatedness as employed in Pathfinder on concepts provided by the participant, the AST aims at capturing the structural implicitness (compare Section 2.5).

Just saying what comes to mind first "gives leave to treat the direct test like an indirect test" (Dienes & Perner, 1999, p. 738). At least some of these associations, which are no doubt explicitely available because they have been articulated, are seen as a "representation of compounds (typically: compound properties) that leaves the structure of its components implicit" (Dienes & Perner, 1999, p. 740). By asking participants to perform quick intuitive similarity ratings between these associations without labeling the relationship, associations between compounds that have no explicitly available label might be captured. This assumption is based on the fact that pair associations are one of the fundamental levels of learning: Boucher and Dienes (2003) state that artificial grammar learning, one of the key paradigms of implicit knowledge (compare Section 2.3.1), is based on the detection of

[2] I am very grateful to Tobias Lattke, who programmed the system. I am also grateful to Lutz Lippke, Henryk Plötz, and Sebastin Ssmoller for their participation in programming the appliction. Apart from the programming, no other external assistance or support for the design of the AST system was used. The AST has been used in experiments together with other researchers who tested their own hypothesis with the tool (Ceglarek, Meyer, Kermas, & Rothe, 2006; Meyer, Ceglarek, Kermas, Lattke, & Rothe, 2006)

frequency regularities on bigram level. According to the authors, participants exposed to bigrams classified new artificial grammar-based test items very similarly to participants exposed to complete grammatical strings. Thus, the detection of associations between tuples is central to implicit learning, and thus a nonverbal feeling of association between known tuples might serve as a measure of this process.

In summary, the AST combines the free term entry task, term association task, and Pathfinder algorithm techniques, which are the available structural assessment techniques that have proved to be valid and reliable respectively in various research studies (Beatty & Gerace, 2002; M. A. Davis et al., 2003; Jonassen et al., 1993), into one test system that might be able to elicit at least a portion of the structural knowledge that is primarily accessed in an unconscious way.

3.9.1 AST configuration module

The AST test is an internet-based client/server application. It is programmed in the JAVA programming language. The server module contains the administration module that allows configuration of test series and resides on a web server. The server module also distributes tests, collects the data, and performs the conversion from the participant's relatedness matrix into a PFNET $(n-1, \infty)$ based in the algorithms supplied by Schvaneveldt (1990). The configuration module consists of a web-interface to the AST's data base and can be accessed from any Web browser over the Internet. It is currently available at http://abulifa.wiwi.hu-berlin.de:8080/ast/admin. It allows the creation of new AST test series, the generation of new participant codes for an existing test series, and the analysis of collected results.

If a new test series is created, the administrator specifies its parameters (compare Figure 3.4): The name of the test series, up to ten stimulus concepts that require free term associations, the stimulus concept whose associated terms enter pairwise concept comparison, the maximum number of concepts that are selected for pairwise comparisons, the mode of the test (explained in Section 7.1.1), its language (German or English), and its presentation form: In the standard AST (undivided) mode, the pairwise concept comparison take place immediately after term associations. In the split mode, free term entry and pairwise concept comparisons of associated concepts are split into two separate tests that can be performed at different points in time. This latter feature will play a role in some of the validation experiments. The possibility to supply more than one concept for free term entry is designed to allow participants to familiarize with the association task before collecting the vital associations used for pairwise comparisons. Also, it allows for association experiments outside the scope of this thesis.

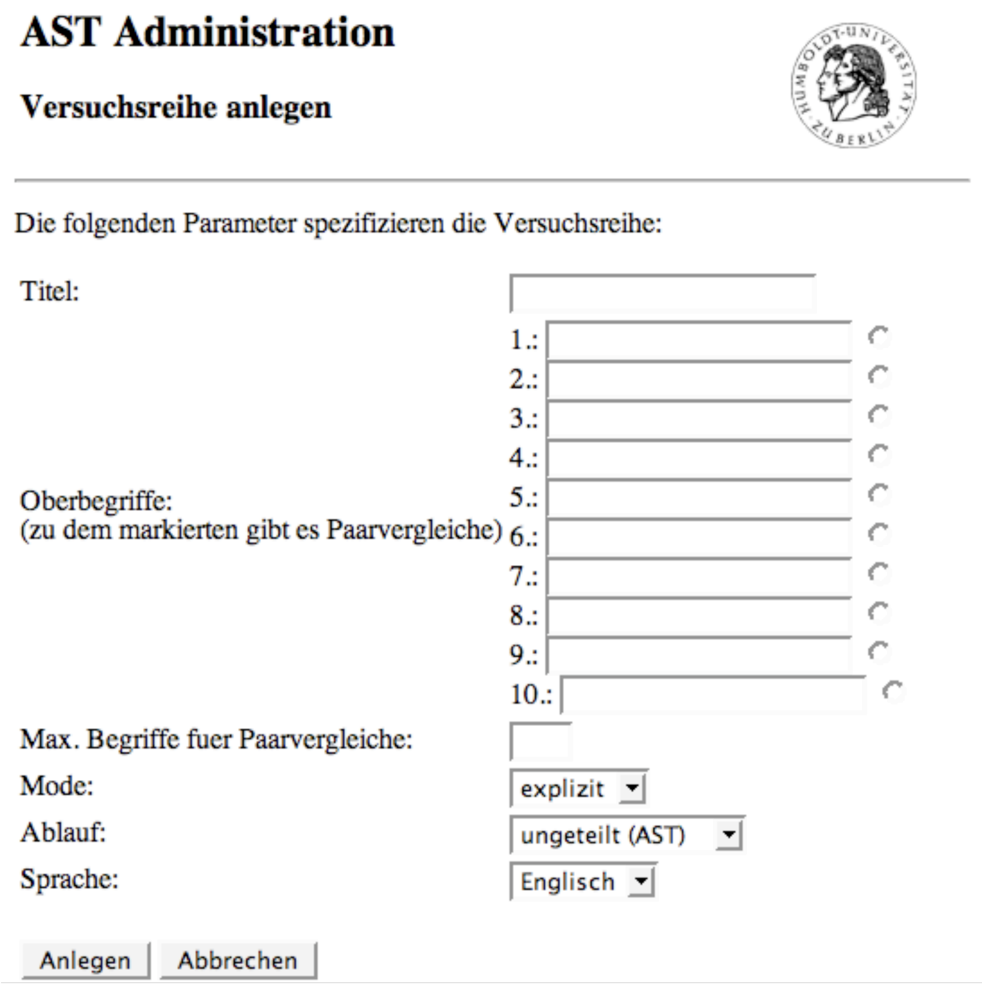


Figure 3.4: Screenshot of the German AST configuration module

The maximum number of concepts that are selected for pairwise comparisons requires the specification of a maximum value because of the following reason: The total number of comparisons ($n(n-1)/2$) that the participants need to make depends on the number (N) of the terms they have typed into the system during the first stage of the test. For example, if the participant entered 20 terms during the first stage of the test, she or he would have to perform 190 ($20(20-1)/2$) comparisons in the second stage of the test. In order to limit the total number of similarity ratings a participant has to perform in the second sub-task, the maximum number of terms for the participants to compare is specified during the configuration. If the number of terms entered within a knowledge domain during the first stage exceeds the specified maximum number, the total number of terms selected into the second stage would be equal to the maximum number. The sample of terms would be chosen from clusters formed in the first stage according to the following principle: The AST classifies the terms entered during the first stage into different clusters according to the measured thinking times. As in the FTE tasks presented in Section 3.3.1, a time series cluster analysis identifies peaks of thinking times between clusters of fast-paced concept associations (compare Figure 3.1). These clusters are thought to represent different sets of individual cognitive structures in the same knowledge domain. If the number of terms the participant entered exceeds the specified number for pairwise similarity comparison, the very first term in each cluster enters the second stage, the rest of the terms is selected from each of the clusters in proportion to the cluster size in order to assure that the selected terms represent the overall aspects of the terms entered in the first stage of the test. In this way, a certain level of preservation of cognitive structure is maintained despite sampling of a limited number of items. Davis, Curtis and Tschetter (2003) determined that there is a tradeoff between reasonableness of the length of a comparison task and its predictive validity. They established that a value of 15 concepts entering pairwise comparisons still delivers an adequate prediction at a reasonable length of the comparison task.

In another section of the administration module, the administrator can generate participant codes at his discretion. These codes are always associated with the test series from which they are created, i.e. when a participant enters his or her code, the AST system is aware of the corresponding test series and serves it. The final section of the administration module is for result analysis. It will be outlined after explaining the test module.

3.9.2 AST test module

The test can be accessed from any JAVA-enabled computer over the Internet using a standard Internet Browser. The test is started by entering the URL of the test (currently, http://abuluifa.wiwi.hu-berlin.de:8080/ast) in the browser. This starts a JAVA applet that displays the test environment and prompts the participant for his or her code. Participant codes are generated in the AST administration module by the administrator (compare previous section). After entering his or her code, the following instruction is presented to the participant:

```
Dear Participant,

The AST test system tests your knowledge on a specific topic
```

area within a domain you are most likely familiar with. The test consists of two parts.

At the beginning of the first part, you will see the name of the topic area that is being tested. Your task is to type in any terms you can think of spontaneously, which you think belong to the presented topic (i. e. that are associated with it). For instance, if the topic area is ‘‘bicycle’’, you may have the terms ‘‘Chain’’, ‘‘Pedal’’, ‘‘Saddle’’, ‘‘Tire’’, ‘‘patching a tire’’, ‘‘holding balance’’, and ‘‘Dynamo’’ in mind (note that each term can consist of several words and may describe an item or an action as well).
You are asked to type these terms into the provided textbox and press the <enter> or <return> key immediately after the last letter of each term, e.g.
Chain <return>
Pedal <return>
Saddle <return>
Tire <return>
Patching a tire <return>
You should type as many terms as you consider being related to the topic area. The pause time between two terms will be recorded. If you do not have any other terms to enter, please click on the ‘‘Done’’ button. Please note that you will receive only one topic area for entering your associated terms, so please do not spare things that come to your mind for later topics. Enter everything that comes to your mind when you see the topic area in large letters on the top of the next screen.

The last two sentences (“Please note...”) are omitted if more than one stimulus concept is specified in the configuration module. After the participant has indicated his or her understanding of the instructions by clicking on a button labeled “understood” below the explanation, the free term entry (FTE) part of the AST is displayed as depicted in Figure 3.5. The participants type one term into the textbox on the left-hand upper side and press enter before typing in another term. The terms they entered would appear in the text area on the right-hand side. Only when the participants believe that they do not have any more terms to associate, they click on the “Done” button on the left-hand lower side.

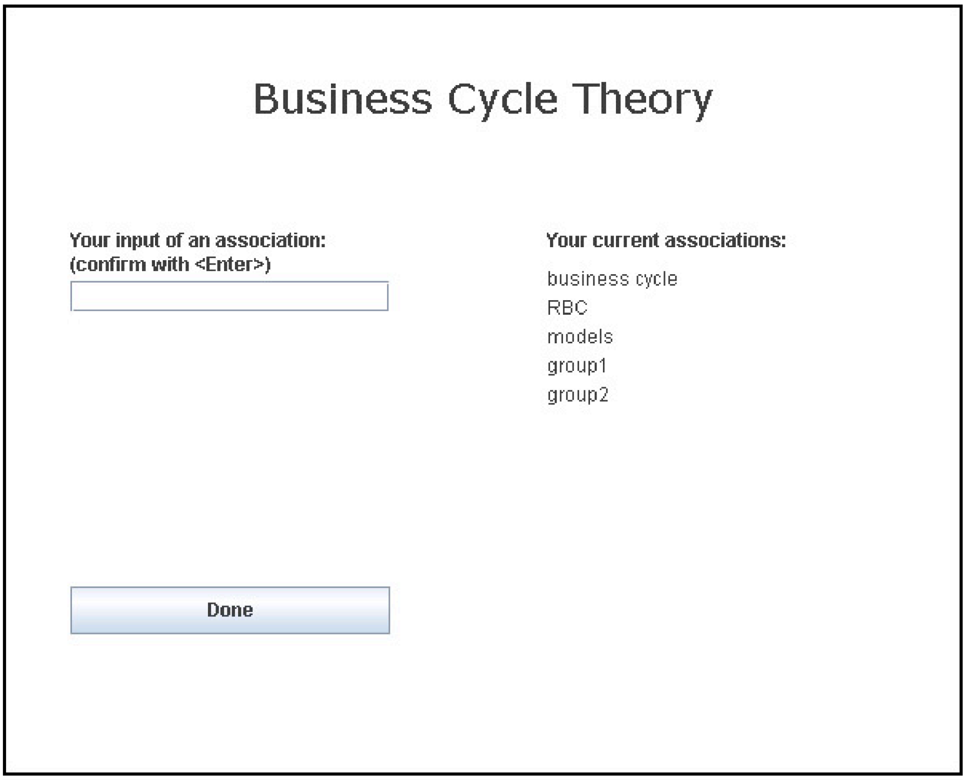


Figure 3.5: AST interface in the free term entry (FTE) task

After the participant clicks the "Done"- button, a second instruction is presented:

```
In the second part of the test, some of the terms that you
have typed in previously will be presented to you as pairs.
You will have to indicate the strength of the relationship
between the two terms with regard to their content. Please
rate the strength of the relationship between the two terms
with regard to their content on a scale from 0 to 4.
0 indicates that there exists no or only a very weak relation-
ship between the two terms. The relationship becomes
stronger with increasing numbers. That is to say, if you
choose 4, it implies that you think that the two terms are
strongly related to each other. Please make spontaneous
and quick judgments without long considerations. You do
not have to specify the nature or type of the relationship, nor
do you need to be fully aware of it. A ‘‘gut-feeling’’ of
```

`relatedness is sufficient for placing a relation between two concepts.`

After confirming their understanding of the instructions, the pairwise comparison task is presented as shown in Figure 3.6. On the top of the screen, a part of the instructions remains visible. Below the instructions, two terms are presented for the participants to determine their interrelatedness by clicking on the appropriate number below the two terms.

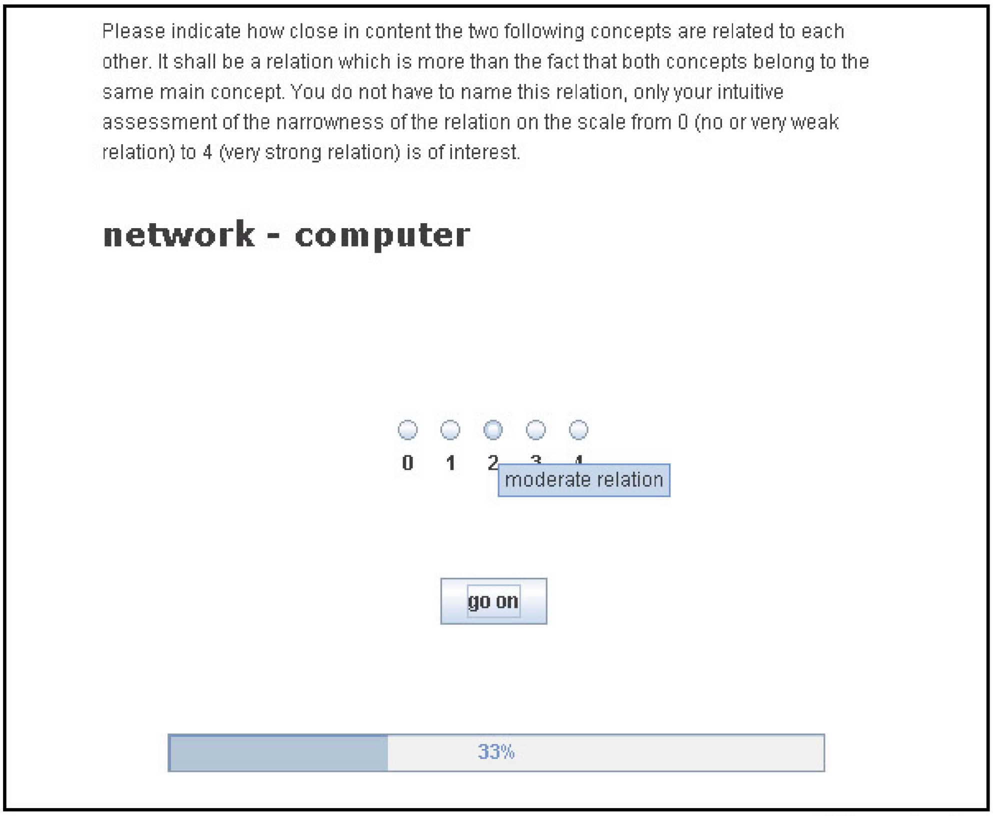


Figure 3.6: AST interface in the pairwise similarity rating task

After a participant has completed an AST test run, the server modules provides several files in the comma-separated format (CSV) that contain the participant's input for further analysis. For each participant, two files are generated for each stimulus concept:

1. A cluster matrix displaying the results of the time series variance analysis of thinking times between concepts. It consists of one line per identified cluster beginning with the number of the cluster that is followed by the associated concepts grouped into that cluster

2. A time matrix containing four lines. The first line holds all associated concepts, the second line displays the number of associated concepts, the third line shows the thinking period prior to the entry of the first letter of each concept in seconds, and the fourth line displays the typing time for each concept.

For the concept whose associated terms entered pairwise comparisons, two additional files are generated:

1. A data matrix containing the raw association strengths as specified by the participants in a term $\times$ term format. Those terms that were not selected for pairwise comparison are marked with a '-' in the appropriate cells.

2. A pathfinder matrix containing the PFNET$(n-1, \infty)$ of the data matrix. Association strengths are recoded into distances with 1 indicating shortest distance and 4 indicating the longest distance.

The AST administration module delivers only two basic values per participant: The number of associated concepts and the number of edges in the graph. For obtaining further graph-theoretic metrics for further analyses purposes, external graph analysis software has to be used. In the context of this study, two software packages that can import AST's pathfinder matrices were employed for this task: UCINET (Borgatti, Everett, & Freeman, 2002) and AGNA (Benta, 2003).

3.9.3 AST coefficients and their relationship

The AST delivers an individual PFNET graph per participant with node labels that have been generated by the participant herself through term associations. Thus, these graphs cannot be compared to each other or to a referent graph as suggested by Schvaneveldt (1990) and Eckert (1998a), because these comparisons require graphs with the same nodes. Thus, global values that describe the graphs are used to quantify AST-generated knowledge networks, based on the suggestions by Bonato (1990, compare section 3.5.3) and Eckert (1998a, compare section 3.5.4). These are:

- Number of associated concepts
- Number of nodes (graph order, equals the number of associated concepts; if a threshold for maximum association is specified and more concepts than the threshold value are associated, this equals the threshold value)

- Number of edges (graph size)
- Graph density
- Graph diameter and
- Relative maximum path length (RMPL)

Furthermore, with reference to the FTE method, AST reports the number of clusters (compare Section 3.3.1). If the coefficients that the AST delivers represent different aspects of structural knowledge, they must correlate with certain outside criteria. Also, different coefficients are assumed to correspond to different knowledge types.

Number of nodes (graph order)

The number of associated concepts is seen as a measure of declarative knowledge. As stated in the model of dimensional knowledge types (compare Figure 2.8), declarative knowledge is knowledge that can be codified; it consists of verbalizable chunks in the terms of the ACT-R model. According to the theory of implicit and explicit knowledge by Dienes and Perner (1999) as presented in Chapter 2, the ability to verbalize a proposition indicates a certain level of explicitness. Tulving's memory model also assumes that declarative knowledge, whether episodic or semantic in nature, is consciously available and thus verbalizable. Thus, the number of concepts that a participant associates verbally to a stimulus indicate the magnitude of available declarative knowledge in the given subject domain.

Number of edges (graph size)

The number of edges is assumed to refer more closely to structural knowledge and to structural implicitness (compare Chapter 2). As they are set by the participant in a quick and intuitive way and do not have to be explicitly labeled, they do not necessarily encompass verbalizable forms of knowledge. In Engelkamp's multimodal memory model (compare Figure 2.4), they refer to connections in the conceptual system. Those can consist of word knowledge, but those can also consist of referent knowledge which is not articulable if not connected to word knowledge. In the ACT-R model (compare Section 2.4.2), edges can refer to productions that link declarative knowledge elements. In Tulving's model (compare Figure 2.3) however, edges linking declarative elements to each other are part of the semantic memory system, which is part of the declarative system. If these edges are accessed in an unconscious way, they can be seen as a part of implicit memory processes (compare p. 13).

The differing characteristics that are attributed to concept-linking edges by the different memory models outline the heterogeneity of the concept: edges are not purely implicit in nature, but combine implicit and explicit aspects, which is also visible by their assignment to both categories in the model of knowledge types (Figure 2.8).

Another reason for the heterogeneity of the edge concept is the fact that edges are always connected to nodes. In fact, an edge is a set of two nodes. Edges thus cannot be seen as independent of nodes, which in turn consist of the associated concepts (which fits into the conceptualization of structural knowledge as something that links certain entities). The simple number of edges is thus seen as an operationalization of what Dienes and Perner describe as explicit knowledge that "can be a representation of compounds (typically: compound properties) that leaves the structure of its components implicit" (1999, p. 740). This leads to the fact that the number of edges depends heavily on the number of associated concepts. Note that despite their interconnection and the resulting positive correlation between edges and nodes, they are not mathematically deductible from each other: Participants specify nodes prior to specifying whether there is a connection between two nodes, which do not necessarily *have* to be connected. The number of edges inside a graph will thus correlate with the number of nodes, but it will not be deductible from the number of nodes.

Density and weighted density

Another measure capturing the number of edges inside a graph is the density (compare Section 3.5.3). It specifies the ratio of present edges in relation to possible edges inside the graph. The density is calculated by

$$\frac{l}{\frac{1}{2}n(n-1)} \tag{3.1}$$

where l is the number of lines in the graph and n is the number of nodes. Note that the density measure is logically connected to the number of nodes and edges in such a way that it will quickly approach values close to 0 if both the number of edges and the number of nodes increase, as Figure 3.7 illustrates. The density measure is thus logically dependent on the number of edges and nodes and will display a *negative correlation* with these two values.

The same applies to the weighted density measure, which takes the edge weights as delivered by the AST into account. Therefore, instead of the (binary) density, the weighted density is employed as proposed by Benta (2003). It is the sum of all edge values devided by the number of all possible edges

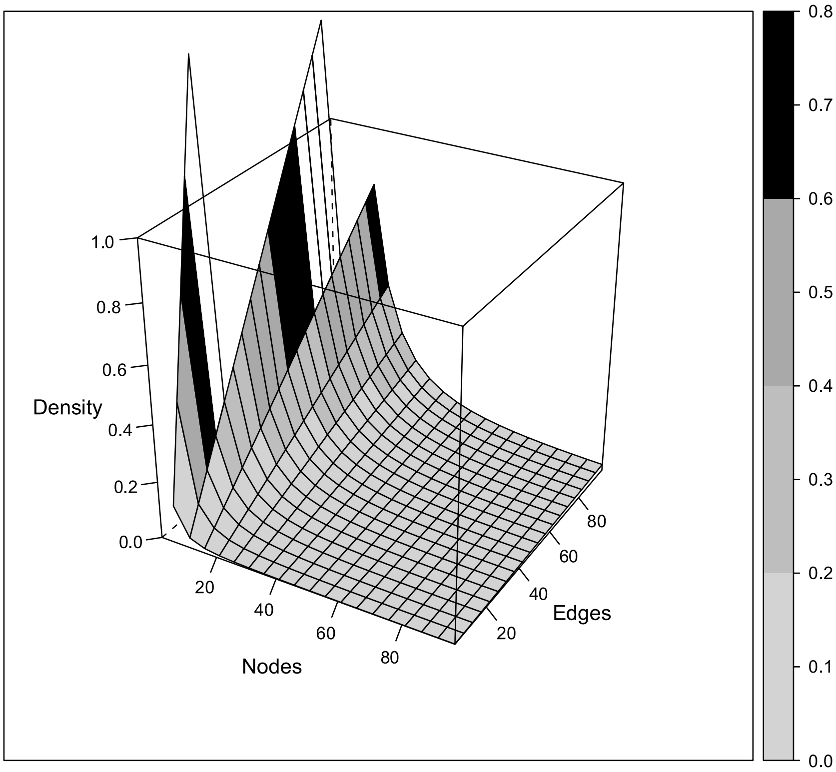


Figure 3.7: Relationship between the graph density measure, the number of nodes, and the number of edges inside a graph. White areas are undefined as the density rises above 1 in those areas. It becomes evident that the density measure approaches 0 for an increasing number of nodes and only visibly increases for a high number of edges at low numbers of nodes; most values will be below 0.2.

in that network. The density of a weighted undirected graph (as supplied by the AST) is calculated by

$$\frac{2}{n(n-1)} \sum_{i=1;j=i+1}^{n} x_{ij} \tag{3.2}$$

where n is the number of nodes in the network and x_{ij} denotes the value of an undirected edge between nodes n_i and n_j. Because the weighted density is the primary indicator of the amount of edges and their strength inside a graph, it is seen as the primary indicator for structural knowledge, and, if applied to those edges that a participant cannot label, for structural implicitness. The weighted density of a knowledge graph is seen as the closest quantification of the heterogeneous edge construct as discussed above. It is mathematically connected to the number of edges and nodes similarly to the regular density measure and will thus also exhibit negative correlations with the number of edges and nodes as shown in Figure 3.7. This will lead to the situation that a high number of associations, which are seen as an indicator of declarative knowledge (see above) will most probably lead to low values of structural implicitness, indicated by the weighted density. As such a negative correlation is not justified by the underlying constructs (there is no argument that justifies the fact that a high amount of declarative knowledge on a subject conditions a low amount of structural implicitness and vice versa), the coefficients delivered by the AST cannot be considered as operationalizations of different knowledge types in the strict sense, but more as an indicator or correlate.

Diameter

Eckert (1998a) treated the diameter (compare Definition 6 in Section 3.1) as a measure for the amount of structural knowledge (compare Section 3.5.4). However, his networks were directed whereas undirected networks as produced by the AST will more likely display short diameter values with less variance. The relationship between the diameter and knowledge types remains open at this point, but it has to be treated as a somewhat meaningful measure for knowledge because it is correlated with outside criteria such as grades in SLT-elicited knowledge graphs (Ceglarek, forthcoming). The diameter is not independent of the number of nodes and the number of edges; larger graphs in terms of number of nodes and edges will increase the probability for larger diameters, as the distances between nodes increase with graph size. The diameter will thus exhibit positive correlations with the number of nodes and the number of edges.

Relative maximum path length (RMPL)

The relative maximum path length denotes the relationship between a graph's diameter and the longest possible path inside that graph (compare Section 3.5.3). Bonato (1990) established it as a measure for compactness of knowledge graphs; an RMPL of 1 occurs if all nodes are aligned on one path while a RMPL of 0 indicates a star-shaped graph with one node in the center and all other nodes connected to it with a path distance of 1. Bonato (1990) argued that smaller RMPL values indicate a more favorable integration of knowledge elements (compare Section 3.5.3). RMPL is calculated by

$$\frac{d}{(n-1)} \tag{3.3}$$

where d is the diameter of a given graph and n is its number of nodes (graph degree). Figure 3.8 visualizes the relationship between the number of nodes in a graph, its diameter and its RMPL value. It becomes clear that the RMPL will exhibit positive correlations with the number of nodes in a graph and its diameter.

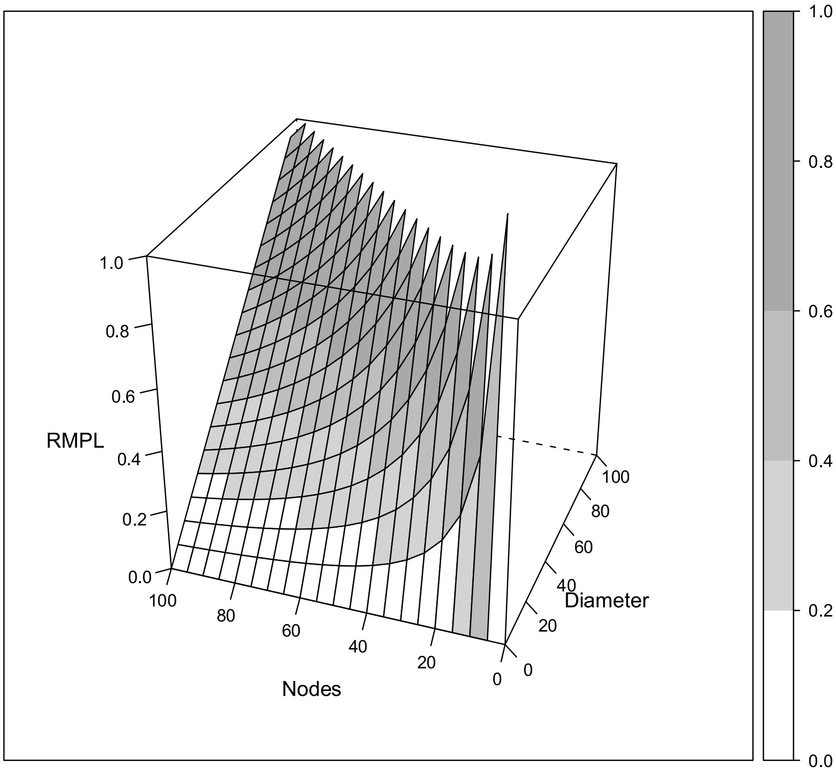


Figure 3.8: Relationship between the graph RMPL measure, the number of nodes, and the graph diameter. It becomes evident that the RMPL measure approaches the favorable 0 for an increasing number of nodes under short diameter values.

Due to the fact that Equation 3.3 involves the diameter, which already includes edges and nodes for calculation in the nominator, and the number of nodes in the denominator, it is by definition more closely linked to the number of associated concepts (explicit knowledge, see above) than to the number of edges (explicit structural knowledge and structural implicitness). It is thus seen as indicator of the organization of explicit knowledge.

Number of clusters

The number of clusters is primarily dependent on the number of verbal associations made by a participant and by their speed. It will thus correlate with the number of associated concepts. The number of clusters is supposed to probe into the conceptual knowledge structure of expert declarative knowledge (Beatty & Gerace, 2002) and thus belongs to the explicit domain.

Conclusion

Not all AST coefficients can be derived from each other mathematically, but all of them will correlate to differing extents due to their dependence on nodes and edges, which are again correlated. For example, both the diameter and the number of clusters in a graph correlate positively with its degree, which also leads to a positive correlation between the diameter and the number of clusters. The correlational structure of all AST-coefficients is presented in Table 3.3. This dependency leads to to the question whether all measures should be combined to a single measure, as it will probably lead to a single-factor structure of AST coefficients.

There are however subtle differences among the concepts. For example, the weighted density contains not only the number of edges with reference to the possible number of edges, but is also the only coefficient that contains the edge strength. All measures will correlate, but a combination of all measures into one coefficient, e. g. through factor analysis, will most probably annihilate such subtleties. The individual measures are thus retained for the time being. The internal consistency of the AST coefficients will have to be determined through correlation and oblique factor analysis.

Table 3.3: Dependency and correlational structure of AST coefficients. $+$ indicates a positive correlation, – a negative correlation, $++$ a positive formal mathematical dependency, and $--$ a negative one.

	Degree	Order	W. Density	Diameter	RMPL	Cluster
Degree		$+$	$--$	$+$	$--$	$+$
Order			$--$	$+$	-	$+$
W. Density				$-$	$+$	$-$
Diameter					$++$	$+$
RMPL						$-$

To conclude this section, Figure 3.9 shows an AST-elicited PFNET of a cleaning apprentice from Experiment 7.1.4.

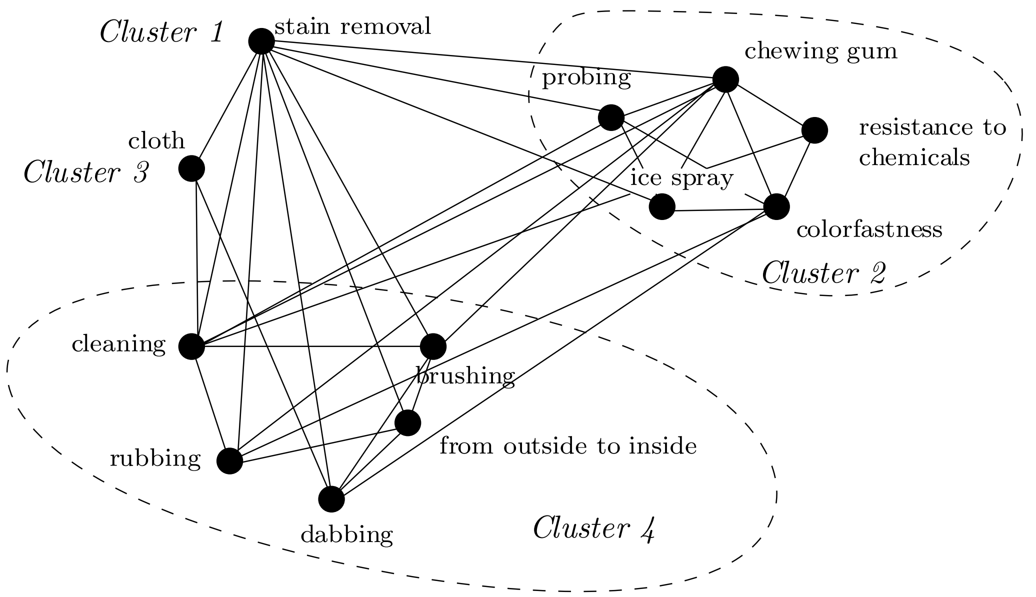


Figure 3.9: AST-elicited knowledge graph. This participant was a cleaning apprentice who completed the AST with the stimulus “carpet cleaning”

3.9.4 Hypotheses

It is assumed that the different AST-elicited coefficients number of associated concepts (graph order), the number of links (graph size), the weighted density, the diameter, and the relative maximum path length have subtly different relations to different types of knowledge, i.e. that they operationalize differing underlying concepts – declarative knowledge and structural knowledge – which are correlated but not the same (see above). At the same time, the mathematical and logical dependencies between the coefficients render the use of a factor analysis over AST coefficients questionable as items are mathematically related. A factor analysis can only serve as a representation of the dependencies summarized in Table 3.3, but an oblique factor rotation might conserve the assumed subtle differences among constructs. It will thus be performed prior to testing the following hypothesis in order to illuminate the AST’s internal consistency.

The elicitation of explicit and implicit knowledge structures that are unconsciously accessed and are part of individual non-explicit knowledge as outlined in Chapter 2 is the prime attempt of the AST system. First of all, the fundamental claim, namely the assumption that the AST elicits connections between knowledge elements that are implicit, needs to be tested. This hypothesis is thus the main AST-related hypothesis.

Hypothesis 3.9.1 *If test participants working on the AST are asked to place unlabeled edges between two concepts in a quick and intuitive way, they place more edges than in a condition where edges have to be labeled.*

If the AST does measure knowledge, it should display sensitivity to learning, the process of knowledge acquisition. This is being put forward in the second AST-related hypothesis:

Hypothesis 3.9.2 *AST coefficients show change in their respective direction towards more knowledge if learning occurs in a field that targets the specific knowledge types.*

Although the AST is not a test in the classical sense as participants create their own items during the test under certain constraints, its construct validity has to be investigated, as it claims to measure a construct (knowledge). Construct validity refers to the degree to which inferences can be made from the operationalizations of a construct – the AST's coefficients – to the theoretical construct on which the operationalization is based (Campbell & Fiske, 1959). Thus, as Campbell and Fiske suggested, the AST is compared to outside criteria for validation. Due to the fact that the constructs are already heterogeneous in nature (compare Section 3.9.3), the following hypotheses are considered to be of less significance than the two above main hypotheses.

Despite its differences to other structural knowledge elicitation techniques (compare Table 3.2), the AST does have similarities to other elicitation techniques requiring the participants to associate concepts to a stimulus, e.g. SLT. Although the SLT captures only edges with explicit labels, such edges are considered to be included in the set of edges that the AST elicits (edges with labels and those edges that the participants cannot label). Thus, the different indices that are both available for the AST and the SLT should display correlations to a certain extent if employed on the same participants with the same stimulus concept (convergent validity).

Hypothesis 3.9.3 *The number of associated concepts, the number of links, the density, the diameter, and the relative maximum path lengths of SLT-elicited knowledge graphs and AST-elicited knowledge graphs correlate.*

Furthermore, if AST-elicited knowledge has proven to be viable for the attainment of certain goals (compare the knowledge definition in Chapter 2), AST scores have to correlate with measures of goal attainment (congruent validity).

Hypothesis 3.9.4 *AST-elicited structural knowledge measures that have been obtained for a certain domain correlate with performance measures that indicate achievement in that particular domain.*

The prime research assumption put forward at the end of the second chapter stated that a part of individual non-explicit knowledge can be conceptualized as unconscious access to structural knowledge (structural implicitness, compare Chapter 2). It has been hypothesized that the AST is able to at least partly capture those unconsciously available concept relations. Therefore, one has to assume that AST scores indicating structural implicitness (i. e. the weighted density) correlate with implicit knowledge.

Hypothesis 3.9.5 *In a task that involves implicit knowledge for successful completion, the AST-elicited weighted density measure correlates with task performance if the AST is invoked with a stimulus concept that captures the nature of the task in question.*

In order to empirically test Hypothesis 3.9.5 in a laboratory setting, a task is required that does require implicit knowledge for completion. Such a task – complex problem solving (CPS) – is presented in the following chapter. It is followed by chapters on the role of knowledge on group level in CPS and its elicitation. These lead to the presentation of the research method in Chapter 7.

Chapter 4

Knowledge in Complex Problem Solving (CPS)

In order to test the hypotheses presented in the previous chapter, a task is required that involves implicit knowledge for successful completion and whose successful completion can be determined with an objective measure in order to demonstrate the usefulness of knowledge assessment through performance prediction: If (implicit) knowledge enables a certain performance, the amount and/or quality of that particular knowledge should serve as a predictor of the quality of the knowledge-related performance. This assumption is employed widely in assessment practice in different contexts. For example, future driving abilities are tested with a theoretical pen-and-pencil test on driving-related knowledge in driving school. Military aircraft pilots' ability to perform certain manoeuvres is assessed by assessing their structural knowledge on flight-related concepts (Schvaneveldt et al., 1985).

In the same way, knowledge assessment with the AST (compare previous chapter) will be used for performance prediction in a task. This should be a realistic task modeling a capability that is useful outside of the psychological laboratory. The task presented in this chapter is complex problem solving (CPS).

4.1 Defining Complex Problems

According to Dörner (1979), a problem generally consists of three aspects: An undesired initial state, a desired target state and some sort of barrier preventing the immediate transformation of the initial state into the target state (see also Kersting, 1999). Problems can be classified into simple and complex problems, depending on the characteristics of the barrier and of the target state. In simple problems, the barrier can be overcome by combining operations known to the problem solver (Kluge, 2004). These problems are often artificial (or interpolative) puzzles like Towers of Hanoi (ibid) and they have a clearly defined target state.

Complex problems resemble realistic everyday problems (Kluge, 2004; Kersting, 1999), as faced by managers and politicians (Endres & Putz-Osterloh, 1994). Complex problems are characterized by the complexity of the situation, by opaqueness, interconnectedness, dynamics, and by polytely (Dörner, Kreuzig, Reither, & Stäudel, 1983). Kluge (2004, p. 6), summarizes the findings (Dörner et al., 1983; J. Funke, 1992; U. Funke, 1993; Kersting, 1999; Strauß & Kleinmann, 1995; Süß, 1996) on these five characteristics in the following way:

- The complexity of the situation refers to the fact that the amount of information to process is beyond individual human processing capabilities, preventing a complete processing of all available information and the arrival at an optimal solution.
- Opaqueness refers to the necessity of an active information search in solving a complex problem, as not all decision-relevant information is directly available.
- Interconnectedness refers to dependencies between the involved variables. The problem solver must discover dependencies between the variables he or she can alter and must discover interdependency structures.
- Dynamics or momentum implies that the situation changes without actions of the problem solver which limits the time available for reflection
- Polytely means that there are multiple, possibly conflicting goals to achieve. This requires "...the careful elaboration of priorities and a balance between contradicting, conflicting goals" (J. Funke, 2001b, p. 72).

According to Funke (2001b), these characteristics can be reduced to two main characteristics of complex problem solving: The connectivity between variables and the dynamic nature of the problem situation. Neither can be achieved with pen-and-pencil techniques, whereas the opaqueness depends largely on how the problem is presented and complexity is mainly a result of the connectivity. "Connectivity characterizes the structural features of the system. The dynamics brings about a procedural aspect in form of a time-dependent characteristic" (Kluge, 2004, p. 6, own translation). "To summarise: in CPS research, tasks are used which consist of two specific, distinctive features, namely connectivity and dynamics. Both attributes need a computer program for their realisation, and cannot be realised by a paper-and-pencil approach" (J. Funke, 2001b, p. 73). Based on the abovementioned considerations, Frensch and Funke (1995, p. 18) provide the following definition of CPS, which is the one employed in this thesis:

> CPS occurs to overcome barriers between a given state and a desired goal state by means of behavioral and/or cognitive, multistep activities. The given state, goal state, and barriers between given state and goal state are complex, change dynamically during problem solving, and are intransparent. The exact properties of the given state, goal state, and barriers are unknown to the solver at the outset. CPS implies the efficient interaction between a solver and the situational requirements of the task, and involves a solver's cognitive, emotional, personal, and social abilities and knowledge.

The competence to solve complex, dynamic and partially intransparent problems can be seen as a key competence for all academic professions (Wittmann, Süß, & Oberauer, 1996).

4.2 Analyzing CPS Performance with Dynamic Scenarios

Computer programs simulating complex problems are referred to as microworlds (Quesada, Kintsch, & Gomez, 2002) or dynamic scenarios (J. Funke, 2001a). These scenarios combine a formally complex task with a semantic embedding (Hussy & Klinck, 1990). Over several intervals, the problem solver is supposed to reach his or her goal and to make changes to the scenario (Kluge, 2004). Such systems simulate, for example, the operation of a forestry (FSYS, Wagener, Hochholdinger, Conrad, & Wittmann, 1997), the work of a peace corps (MORO, Strohschneider, 1991; Endres & Putz-Osterloh, 1994), being the mayor of a mid-sized town (LOHHAUSEN, Dörner et al., 1983), and the operation of a small shirt factory (TAYLORSHOP, Süß & Faulhaber, 1990). The simulations vary by the number of control variables (from two to more than 2000) and by the complexity of the underlying variable structure and algorithms. An overview over a number of microworlds can be found in Funke (1992) and in Kluge (2004).

Psychometric assessment with dynamic scenarios faces two challenges: comparability between participants' performance in the same simulation and comparability between scores in different scenarios (J. Funke, 2001a; Kluge, 2004). With regard to the first issue, the problem in CPS lies in their feature that "once a subject has entered the first response, the subject moves through the microworld on an individual path.... It is virtually impossible that two subjects follow exactly the same pathway through the system" (J. Funke, 1995, p. 248). In other words, the control over the manipulation is difficult because the test situation in a given moment depends on the participant's previous alterations to the system and the test situation between participants working on the same scenario is different after the first alterations, given that they altered the system in two different ways. This feature complicates standardization and control of the test situation significantly (Kluge, 2004, p. 9). However, the fact that stimuli are produced by the participants does not render the entire experiment useless, but it demands a very careful analysis of between-subject effects (J. Funke, 1995).

The fact that different scenarios employ different formal underlying sets of logic complicates comparison across different scenarios (Wittmann et al., 1996). Originally, Dörner and colleagues assumed that dynamic scenarios are a valid operationalization of problem solving capability, a construct they assumed to be mainly independent of standard IQ (Dörner & Kaminski, 1987; Dörner, 1987, 1994; Dörner et al., 1983), comparable to Sternberg's concept of practical intelligence (R. J. Sternberg et al., 2000). However, the fact that correlations between different scenario performance scores differ depending on the employed scenario casts doubt on this construct (Wittmann et al., 1996) and CPS performance is seen as an application of acquired scenario-specific knowledge, mediated by the capacity of information processing (ibid).

According to Funke (2001b), two formal frameworks underlying microworlds satisfy the two main requirements of complex problems (connectivity and dynamics): The linear structural equation (LSE) approach and the theory of finite state automate (FSA). Both are briefly outlined in the following.

4.2.1 Linear structural equations (LSE)

Linear structural models can represent systems that consist of quantitative variables. The variables that can be directly manipulated by the problem solver are represented as exogenous variables, the endogenous variables are influenced by exogenous variables and cannot be directly manipulated. The state of each endogenous variable at a given point in time t can be calculated as a linear combination of exogenous and endogenous variables at $t-1$ that influence the endogenous variable in question. Thus, the state of a system with n endogenous variables at a point in time can be described as a set of n linear equations. Consider the following example (adapted from J. Funke, 2001b, p. 75): A simple dynamic system has two exogenous variables A and B and two endogenous variables Y and Z. The state of Y and Z at time t depends on the weighted states of the variables at time $t-1$ as presented in Figure 4.1.

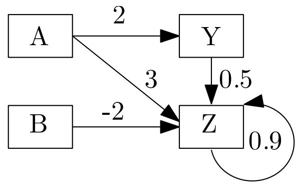


Figure 4.1: Structure of a simple linear system with two exogenous variables A and B and two endogenous variables Y and Z. The arrow weights indicate the state at time t by multiplying the states at time $t-1$ (adapted from J. Funke, 2001a, p. 76).

This system can be represented with two linear equations (one for each endogenous variable) as follows:

$$Y_{t+1} = 2A_t$$

$$Z_{t+1} = 3A_t - 2B_t + 0.5Y_t + 0.9Z_t$$

In linear structural systems, the problem solver's task is "(a) to find out how the exogenous and endogenous variables are related to each other, and (b) to control the variables in the system so that they reach a certain goal value" (J. Funke, 2001a, p. 75).

4.2.2 Finite state automata

Finite state automata can be used to formally describe systems with a more limited scope if it meets three criteria: The sytem can only be in a limited number of states, transitions from one state to another are triggered either by user input or through an automatic process, and user input and system states produce an output signal. FSA are a way of representing dynamic systems whose variables are of nominal scale (Kluge, 2004). Examples for finite state automata are ticket machines, coffee machines, and video tape recorders. FSAs can be represented as state-transition matrices and directed graphs (compare J. Funke, 2001a, p. 77 for further details on FSA).

Taken together, despite their challenges, dynamic systems possess the quality of escaping "both the narrow straits of the laboratory and the deep blue sea of field study, because scenarios allow for a high degree of fidelity with respect to reality and at the same time allow systematic control of influential factors” (J. Funke, 2001a, p. 70).

4.3 Determinants of Individual Complex Problem Solving

According to Funke (1995), three different factors affect individual CPS performance: person, situation, and system variables (p. 250). Person factors are competencies that the subject acquires during interaction with the situation, for example knowledge on how to successfully control a microworld. Person factors can also refer to individual disposition such as computer experience and test intelligence (see for example J. Funke, 1992, 1995, 2004; Kersting, 1991, 1999; Preußler, 1996; Süß, 1996; Süß, Kersting, & Oberauer, 1991). Situation factors refer to a variation of experimental context independent from the used scenario, e. g. to the possibility to experiment with a system prior to task performance or to group versus individual settings. System factors are formal or content related attributes of the scenario, e.g. differences in the semantic embedding or different underlying algorithms. The following sections primarily deal with person factors. Situation factors in the sense of group settings are discussed in Chapter 5, but system factors are not in the focus of this study. An overview of the influence of situation factors on CPS performance can be found in Kluge (2004).

Among person factors, previous studies identified domain-specific declarative knowledge and intelligence, especially reasoning (compare Section 2.7), as influencing complex problem solving performance (J. Funke, 1992, 1995, 2004; Kersting, 1991, 1999; Preußler, 1996; Süß, 1996; Süß et al., 1991). Apart from these established determinants of complex problem solving performance, other individual influences such as other types of knowledge apart from declarative knowledge have been suggested for prediction of CPS performance, for example implicit knowledge (Berry, 1984; Berry & Broadbent, 1995; Broadbent, FitzGerald, & Broadbent, 1986; Buchner, Funke, & Berry, 1995). These individual determinants will be outlined in detail in the following sections.

4.4 Intelligence

The discussion on the role of (test-) intelligence in CPS was sparked by Dörner’s claim that test intelligence and the ability to solve complex problems are statistically unrelated (Dörner, 1979, 1980; Dörner et al., 1983) and has been a rather heated one (Beckmann & Guthke, 1995). Dörner and colleagues found near-zero correlations between their employed general IQ measures and their subjects' performance in the LOHHAUSEN scenario (Dörner et al., 1983). They argued that their employed microworld is a valid operationalization of a real-world complex problem and questioned the validity of IQ-tests, because of their inability to predict CPS performance.

Contrary to Dörner's findings, others (e. g. Strohschneider, 1990, 1991; Wittmann et al., 1996; J. Funke, 1983, 1986, 1992, 2001b; Kersting, 1999; Kluge, 2004) did find significant correlations between measures of IQ and CPS performance. An extended discussion of contradictory findings in CPS with regard to the influence of test intelligence on performance is given by Beckmann and Guthke (1995). They argue that contradictory results are yielded by several moderator variables employed in different studies, such as different interaction effects between solver and tasks, varying goal definitions, differences in semantic embeddedness, transparency, requirements etc. They conclude that both CPS and IQ research have certain deficits and demand a shift away from "unsatisfactory global correlations between intelligence test scores and CPS performance" (p. 186) towards identifications of the role of specific cognitive processes in particular CPS situations. Funke (1995, p. 252) concludes that intelligence measures can predict CPS performance if two criteria are met: (1) Instead of global IQ measures, more specific concepts have to be employed and capacity of information processing as included as "reasoning" in the BIS intelligence model (Operation K, compare Section 2.7) appears to be the most promising one; (2) "CPS quality has to be measured reliably-a condition which is rarely met" (p. 252). Due to the fact that the larger number of studies did find a connection between reasoning and CPS, and based on Funke's argument, it can be proclaimed that one can expect a correlation between the capacity of information processing and CPS performance. The current state of literature as summarized in Kluge (2004) can be interpreted in such a way that with regard to personal factors, CPS performance can be explained as a combination of knowledge and intelligence.

4.5 Knowledge

Funke (2001a) argues that a complex problem-solving task has two main demands, both related to the concept of knowledge: knowledge acquisition and knowledge application. Knowledge acquisition is synonymous to the attempt to identify the structure of the problem, i.e. to "find out details about the connectivity of the variables and their dynamics" (p. 79). The analysis of structure and dynamics are inseparable because system analysis can only take place as a process over time. The process of knowledge acquisition requires identification strategies that establish different qualities of relationships between system variables (i. e. existence, direction, and strength).

Knowledge application refers to a period of system control aiming at the achievement of a certain goal state within the system. In systems based on the linear structural equation model (compare Section 4.2), it requires two subgoals: The transformation of a given state into a target state, and, once achieved, the sustainment of that state.

Funke (2001a) sees knowledge acquisition as a necessary and sufficient condition for knowledge application, but the two processes cannot, despite the possibility of their formal distinction through experiment design, be viewed as independent from each other: in knowledge application, new knowledge is formed and newly acquired knowledge is often applied during the process of its acquisition. Studies by Süß (1999) and Kersting (1991) show significant correlations between knowledge acquisition and CPS performance if subjects are instructed to acquire knowledge. Findings on the role of specific types of knowledge as presented in Chapter 2 are presented in the following subsections, if available.

4.5.1 Declarative knowledge

The role of knowledge in CPS has been empirically analyzed in a myriad of studies. For the German field of research, Kluge (2004) groups the works into schools or groups around certain authors. Without reproducing her extensive classification, the following groups have placed a special emphasis on the role of forms of declarative knowledge in CPS: Kluwe and colleagues, (e. g. Kluwe, 1991, 1996, 1997), Putz-Osterloh and colleagues (e. g. Endres & Putz-Osterloh, 1994; Putz-Osterloh, 1981, 1993; Putz-Osterloh & Lüer, 1981), Funke (e. g. J. Funke, 1992, 1995), and Kersting, Süß and Wittmann (Kersting, 1991, 1999; Süß, 1996, 1999). Some of the findings of these groups will be presented briefly in this section; for a complete review, compare Chapter 2 in Kluge (2004).

All of the abovementioned groups use their own conceptualizations and forms of knowledge, which makes a comparison difficult. In the following, their terminology that they see as a resemblance or subcategory of declarative knowledge as introduced in Chapter 2 will be outlined.

In the context of CPS, Kluge refers to declarative knowledge as domain knowledge (2004, p. 40). Domain knowledge is about the components of the system and includes fact-knowledge on the function and organization of system components and their possible states (ibid). In his 1997 study, Kluwe elicited participants' knowledge on the computer interface of a technical simulation of an asphalt plant and found significant correlations with control performance.

Putz-Osterloh (1993) refers to two types of knowledge required for CPS: “knowledge about ‘how the system works’ ”(p. 332) and “about ‘How to use the system’ ” (p. 332). Due to the fact that she equates knowledge about how the system works with “structure” (p. 332) (i. e. the system’s components, input and output variables, and their interdependencies), it becomes evident that she is referring to conscious access to structural knowledge as introduced in Section 2.5. Her empiric results are thus reported in the section on structural knowledge. Knowledge about how to use the system is explained as “goal-specific methods and goal-specific functions known when accomplishing a specific control task” (Putz-Osterloh, 1993, p. 332) which is categorized as procedural knowledge according to the model in Figure 2.8. Putz-Osterloh (1993) does not specify whether this knowledge is accessed consciously or subconsciously, but her speculations on the importance of non-verbalizable forms of knowledge in CPS (Putz-Osterloh, 1993, p. 334ff) allows the assumptions that she attributes relevance to both modes of retrieval.

Funke (1992) states that the quality of CPS performance is determined by the fit between the subjectively assumed “connectivity structure” (p. 72, own translation) between the system’s variables and its objectively correct connectivity structure. In his research, he elicits this structure with causal diagram analysis: Participants are presented with a diagram containing a box for each system variable. They are then asked to draw lines between those variables where they assume a relationship, and are asked to specify further information, if known: the direction of the relation with an arrowhead, the sign of the relationship with a plus or a minus, and a factor in terms of a number. This operationalization of the subjective mental model is thus a consciously verbalizable form of structural knowledge, and the according empiric results are presented in the section on structural knowledge.

Süß (1996) and Wittmann, Süß, and Oberauer (1996) present a model of knowledge types for CPS-relevant knowledge that distinguishes between two dimensions: what a person knows (fact knowledge and knowledge of actions) and how a person knows: declarative (verbalizable knowledge) vs. procedural (through action), compare Table 4.1.

With reference to the dimensional model of knowledge types (cf. Figure 2.8), these knowledge types translate into declarative knowledge (declarative/fact and declarative/action), conscious access to procedural knowledge (procedural/fact) and to procedural knowledge with arbitrary access mode (procedural/action). Süß, Kersting, and Oberauer (1991; 1993) found significant positive correlations between both postulated forms of declarative

Table 4.1: Taxonomy of knowledge types with examples from the control of a simulated coal power plant (adapted from Wittmann et al., 1996, p. 5)

	Fact knowledge: What is the state?	Action knowledge: What to do?
Declarative knowledge: Being able to answer questions	How is coal supply related to steam pressure?	How can I quickly increase energy output?
Procedual knowledge: Acting successfully	Anicipation of steam pressure curve after a sequence of system alterations	Assembly of an optional sequence of system alterations for the creation of a target curve of energy output

knowledge and microworld performance, and in the Wittman et al. study from 1996, domain-specific declarative knowledge was the strongest performance predictor of the participants' performance in the TAYLORSHOP microworld.

In summary, the above findings allow the following two conclusions: (1) Declarative knowledge has been found to correlate with CPS performance and, more specifically, (2) the quality of consciously available structural knowledge is seen as a predictor of CPS performance and is counted to the domain of declarative knowledge by Putz-Osterloh (1993).

4.5.2 Procedural knowledge

An empiric analysis of the ability to solve a complex problem in the sense of non-verbalizable know-how can be found in Funke (1992). He attempted to operationalize this ability as "differences between target and actual values averaged over intervals and all dependant variables during the control phase" (p. 86, own translation). Due to the skewdness of the observed distribution of this value, he employed a logarithmic scale for his measure, entitled quality of system control (QoS).

The problem with this measure is that it equals the dependant variable CPS performance an thus cannot be treated as a independent variable operationalizing procedural knowledge. The QoS measure further depends on a dynamic scenario with specific target values for each interval in order to calculate deviations for each one. This condition is met in few microworlds, as it conflicts with one of the features of CPS, namely its arbitrary goals (compare Section 4.1).

Other attempts to capture the ability to control a dynamic system have always relied on verbalizable knowledge and thus cannot be counted to the domain of procedural knowledge in the strict sense as outlined in Figure 2.8. An example is the operationalization of scenario-specific declarative action knowledge (compare Table 4.1) by Wittmann, Süß and Oberauer (1996) as rules of thumb. Correct and incorrect rules of thumb for the control of the TAYLORSHOP system were presented as items on a questionnaire and participants had to decide whether the item is either true or false. Although this measure correlated significantly with TAYLORSHOP performance while other measures of declarative fact knowledge did not, this operationalization cannot be seen as anything but a special form of declarative knowledge. In summary, empiric evidence for the role of procedural knowledge in CPS is scarce, although non-verbalizable forms of knowledge are assumed to play a part in CPS, as the following subsection will outline.

4.5.3 Implicit knowledge

In the context of CPS, implicit knowledge can be captured by Kluwe's definenition (2006) as "performance advantages in the accomplishment of cognitive requirements, which are based on an unconscious use of previously perceived and unintentionally stored information" (p. 41, own translation). It is thus closely related to implicit memory processes (compare Chapter 2), which retrieve specific events or experiences "without making the actual content and its meaning conscious" (Markowitsch, 1999, p. 25, own translation).

The concept of implicit knowledge with regard to CPS has been raised and studied by a group at Oxford University, centered around Berry and Broadbent (Berry, 1984; Berry & Broadbent, 1995; Broadbent et al., 1986). In the study from 1984, they employed the SUGAR PRODUCTION FACTORY and the PERSON INTERACTION microworlds. Both tasks are similar with regard to their underlying formal logic: a target state of one variable has to be reached and maintained through the alteration of two input variables (in case of the sugar factory, the number of workers and the behaviour towards the union leader). The goal variable also depended on its state in the previous interval. In the person interaction task, participants had to engage in a computer-simulated interaction and reach a certain level of friendliness in a simulated interaction partner. In both scenarios, performance significantly increased over several repeated trials while practice had no effect on the participants' ability to answer questions on the underlying formal structure. In other words, good problem solvers in terms of scenario performance did not show significantly better results in multiple-choice tests on system knowledge than poor problem solvers. When Berry (1984) provided the participants with decarative knowledge on the system, it led to significantly better results in the knowledge test, but did not lead to an increase of microworld performance in a second experiment. This dissociation between verbalizable, declarative knowledge, and problem solving performance led the to the conclusion that non-salient relationships in the system were learned implicitly and were not accessible in a verbalizable way outside of the context in which the system had to be controlled. Evidence for this dissociation was also found in later studies (Berry, 1991; Broadbent et al., 1986), and Süß (1996, p. 175) reported similar effects with the TAYLORSHOP microworld.

The sugar factory scenario that Berry and Broadbent employed was however not a linear structural model, but a finite state automat (compare Section 4.2.2). Thus, the dissociation between verbalizable knowledge about the automat's states and the ability to control it can be based on the fact that successful operators only have knowledge on desired target states of the automat, while unsuccessful operators have found themselves in a larger number of undesirable states and have thus acquired more knowledge on the system as a whole (Buchner et al., 1995). Other studies (e. g. Marescaux, Luc, & Karnas, 1989; Sanderson, 1989; Stanley, Mathews, Buss, & Kotler-Cope, 1989; Berry & Broadbent, 1995) also challenged the original assumption and led Barry and Broadbent to summarize that "the dissociation between task performance and associated verbalizable knowledge may not be as complete as was at first thought" (Berry & Broadbent, 1995, p. 136). After a series of further experiments with the WHALE GAME scenario (ibid), they still postulate the assumption that repeated practice of scenario control does lead to implicit knowledge on scenario control by stating the following principle: "Keeping the idea that some knowledge is symbolic (verbal) and some is subsymbolic (nonverbal). These broad categories of knowledge are most readily tested by different methods of measurement; questioning or measuring performance. They are also most readily acquired in different ways; verbal instruction and direct task experience being examples" (ibid, p. 146).

Dorfman, Shames, and Kihlstrom (1996) support this argument by stating that intuition and insight also account for CPS performance, which they count to the implicit domain. However, experimental support for this notion is so far laregely limited to the Oxford group's work and has not yet been applied to a group setting.

4.5.4 Structural knowledge

According to Funke (2001b), the quality of a problem solver's identification of a complex system, i.e. the quality of his or her knowledge acquisition process, can be assessed by obtaining the subject's view of the casual relations of the microworld. Thus, conscious access to structural knowledge is seen as a key quality for CPS performance. This causal structure is then compared to the existing one throgh structural assessment. In his study from 1992, Funke operationalized structural knowledge assessment as causal diagram analysis (compare Section 4.5.1). His resulting coefficient, quality of causal knowledge (QoC), significantly correlated with the quality of system control in one study at $r = .54$ (J. Funke, 2001a, p. 96).

In her 1993 study, Putz-Osterloh attempted to provide structural knowledge to a group of problem solvers by providing them with the causal diagram of the system during the knowledge acquisition phase. Surprisingly, the so-equipped group did not show significantly better CPS scores than a control group that had been deprived of the structural information. In a subsequent transfer task involving a modified dynamic scenario, the group that had initially learned from the diagram outperformed the non-diagram group. However, her design makes the fact evident that she employs a different concept of structural knowledge than the one presented in Section 2.5. Instead of viewing structural knowledge as integrated connectedness of individual knowledge organization, structural knowledge "is defined as general knowledge about the variables of a system and their causal relations" (Schoppek, 2002, p. 2) in complex problem solving research and dynamic system analysis (ibid). This form of specific knowledge of system structure, as one could refer to it, thus belongs to the explicit domain, as Preußler (1998) states by referring to the acquisition of this type of system knowledge during learning as "explicit learning" (Preußler, 1998, p. 218, own translation). Preußler tried to elicit this knowledge of system structure with a graph-building technique (compare Section 3.6) and did find positive correlations between this type of knowledge an system performance. Contrary to those results, Putz-Osterloh (1988) did not find such a connection. She employed the structure-laying technique (compare Section 3.5.2) for eliciting participants' knowledge of causal relations between system variables. She provided 20 readily labeled cards with the variable names of the dynamic scenario she employed (MORO) to her subjects and asked them to place the assumed variable relations between the cards. The correct number of placed relations divided by the total number of placed relations was used as an index for structural knowledge and it did not correlate with scenario performance. This operationalization illustrates that Putz-Osterloh employs the term structural knowledge in a different way than the way it was introduced in Chapter 2.

Using the same complex-scenario specific definition of structural knowledge as knowledge on the structure between system variables, Schoppek

(2002) also did not find correlations with CPS performance.

In summary, both Schoppek and Putz-Osterloh attribute an important role to structural knowledge in complex problem solving, define structural knowledge in a very specific and explicit way and fail to find empiric support for their assumption. Thus, empiric findings on the connection between structural knowledge and CPS are somewhat scarce and suffer from different or unstated definitions of structural knowledge. However, the available data leads Funke (2001a) to the conclusion that the assessment of structural knowledge for prediction of CPS performance is not without problems, "but in a number of studies this procedure has been validated" (p. 81). Thus, the role of structural knowledge, defined as the organization of the relationships of concepts in long-term memory providing the integration and organization of declarative knowledge in such a manner to be accessible to procedural knowledge applications (compare Section 2.5) will be revisited in the further scope of this thesis because it is thought to positively influence complex problem solving abilities.

4.6 Summary and Hypotheses

The ability to solve complex problems can be seen as a key competence for all academic professions (Wittmann et al., 1996). Complex problems can be simulated with computer-simulated dynamic scenarios or microworlds. Personal factors influencing individual complex problem solving (CPS) performance are intelligence, especially the capacity of information processing, and knowledge. Apart from declarative knowledge in the sense of the model presented in Figure 2.8, findings indicate that both implicit knowledge and consciously available structural knowledge play a part in CPS. Since a part of implicit knowledge can be conceptualized as unconscious access to structural knowledge, the following hypothesis is put forward:

Hypothesis 4.6.1 *Individual CPS performance is positively correlated with implicit knowledge that is acquired through experience with a dynamic system.*

If H 4.1 holds true:

Hypothesis 4.6.2 *The acquisition of implicit knowledge through experience leads to higher levels of structural implicitness than the acquisition of scenario knowledge without experience.*

If both 4.1 and 4.2 hold true, 4.3 follows:

Hypothesis 4.6.3 *Individual CPS performance correlates positively with scenario-relevant implicit structural knowledge.*

Chapter 5

The Influence of Knowledge on Problem Solving in Groups

Until now, this thesis has mainly referred to individual knowledge, its different forms and shapes, its importance in (individual) complex problem solving tasks, and to the ability to elicit a certain form of knowledge, namely structural knowledge. Thus, person factors of Funke's taxonomy of variables that affect CPS (J. Funke, 1995, compare Chapter 4) have been in the focus of attention so far. This chapter will introduce complex problem solving in small groups as a situation factor of CPS, review findings from literature and present hypotheses on how the distribution of knowledge inside a group affects its ability to solve complex problems. According to Forsyth (2006), "a group is defined as two or more individuals who are connected to one another by social relationships" (p. 3). More specifically, a team is defined as "a distinguishable set of two or more people who interact, dynamically, independently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span of membership" (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000, p. 273).

5.1 Research Strands on Group Performance

As stated in the introduction, assigning tasks to groups is seen as a possibility to overcome individual information processing limitations in complex information environments (J. H. Davis, 1980; Kahnemann et al., 1982). Project-oriented team work is one of the most common forms of collaboration in today's knowledge-intensive businesses (Scholl, 1997). However, research on the effects of group work on problem solving performance has revealed that group processes can lead to an increase in group performance as well as to a decrease in group performance (Williams & O'Reilly, 1998). In order to determine the conditions under which groups show a higher level of performance, different models and mechanisms have been suggested (Knippenberg, De Dreu, & Homan, 2004). One aspect, the effect of knowledge distribution inside groups on their performance, has been the subject of extensive research under different headings in different fields, e.g. in research on team mental models (Mathieu et al., 2000; Mohammed & Dumville, 2001), in research on the effects of group diversity (Sauer et al., 2006; Knippenberg et al., 2004; Knippenberg & Schippers, 2007; Williams & O'Reilly, 1998), and in research on team effectiveness (Scholl, 1996, 2005).

Team mental model research focuses on a group's joint understanding (i.e. of the task) and shared knowledge (Mohammed & Dumville, 2001, p. 89). "Team Mental Models are team members' shared organized understanding and mental representation of knowledge about key elements of the team's relevant environment" (Mohammed & Dumville, 2001, p. 90). The term "refers to an organized understanding of relevant knowledge that is shared by team members" (ibid, p. 89). In Mathieu et al. (2000), the connection between mental models and structural knowledge (compare Chapter 2) becomes evident: "Essentially, mental models are organized knowledge structures that allow individuals to interact wit their environment. Specifically, mental models allow people to ... recognize and remember relationships among components of the environment, and to construct expectations for what is likely to occur next" (Mathieu et al., 2000, p. 274). Accordingly, Rowe and Cooke (Rowe & Cooke, 1995) suggested to use structural knowledge elicitation techniques for mental model assessment in team mental model research, and Mathieu and colleagues (Mathieu et al., 2000) used the Pathfinder technique (compare Section 3.7) to do so.

Team mental model research has dealt primarily with the effects of cognitive *homogeneity* in a group on the group's performance. Research on group diversity focusses on the effects of *differences* among group members on the performance of the group. Thus, the study of team mental models can be seen as a strand of research among the (wider) field of group diversity (Sauer et al., 2006).

Group diversity research deals with differences in group members occurring on different levels (McGrath, Berdahl, & Arrow, 1995). One of these levels is the social categorization level on which group members' subjective categorizations of self and others into subgroups is considered (Knippenberg & Schippers, 2007). The information/decision-making level (Williams & O'Reilly, 1998) considers the effects of the distribution of knowledge, information, and abilities among the group on the performance of the group.

Differences on this level are also referred to as cognitive diversity (Sauer et al., 2006).

Research on general team effectiveness such as Scholl's work (Scholl, 1990, 1992, 1996, 1997, 2004, 2005) has focussed more broadly on the effects of information and knowledge exchange and knowledge *production* in a team on its performance and innovativeness.

However, as already pointed out in the introduction, many studies on group performance do not model an information-rich environment or complex tasks, but employ rather simple and deterministic problems (Endres & Putz-Osterloh, 1994). This is insofar unfortunate, as complex computer-simulated scenarios provide a way for simulating complex problem solving in a laboratory setting (compare previous chapter). The few existing studies that employ complex problems in a group setting, such as Endres and Putz-Osterloh (1994), do not clearly identify whether the amount of knowledge in complex problem solving groups or the amount of knowledge overlap inside a group is the key determinant for CPS performance.

In the following sections, the findings from various strands of research on the relationship between a group's task performance and a group's knowledge will be discussed. It will become evident that some of the studies altered both the way in which information was distributed among group members and the amount of information a single group member had to remember at the same time, which makes certain interpretations regarding the sole role of information distribution in group performance difficult. The discussion will ultimately lead to two opposing views: One promotes cognitive homogeneity as a condition for group success, the other promotes cognitive heterogeneity as a condition for group success. In order to combine these views, Scholl's model of team effectiveness (Scholl, 1996, 2005) is introduced claiming a curvilinear relationship between cognitive heterogeneity and team performance. The section closes with an application of Scholl's model to complex problem solving tasks and according hypotheses.

5.2 Task Types and Group Performance

In various studies, different types of tasks have been employed in order to study the effect of knowledge or information distribution on group or team performance.

According to Steiner (1972, 1976), group tasks can be classified into five groups with reference to how individual inputs are related to the group's product: additive (individual performance is added up to potential group productivity, PGP), compensatory (individual performance is averaged for

PGP), disjunctive (best individual answer is PGP), conjunctive (worst individual performance determines PGP) and discretionary (PGP cannot be determined because the group chooses how to combine individual efforts). A typical example for additive tasks is idea generation in brainstorming (Forsyth, 2006), which is dealt with extensively in brainstorming literature (e. g. Diehl & Stroebe, 1987). Discretionary tasks are tasks in which “group members can perform using their own preferred combination procedures” (Forsyth, 1990, p. 266). The fact that non-discretionary tasks have dominated laboratory research on problem solving is unfortunate (Scholl, 1997) as project assignments, one of the most common forms of current organizational work forms, fall into the discretionary category which may contain any of the other tasks’ characteristics in an undistinguishable way (ibid, p. 385).

5.2.1 Group performance in non-discretionary tasks

In the following, findings from different fields that deal with the effect of cognitive diversity on team performance in different task types are presented. For those task types that rely on a fixed way of combining individual efforts (all above types except discretionary tasks), additive and disjunctive tasks are presented as representing this simple type of tasks. They are followed by results for discretionary tasks before own hypotheses on the influence of cognitive diversity in groups on a discretionary task (coacting complex problem solving) are presented.

Additive tasks

One application of additive tasks in organizational practice is idea generation in groups, i. e. brainstorming. However, the number of ideas generated per person declines with group size in such a way that groups do not outperform a set of individuals with the same size as the group who work on their own (so-called nominal groups). The brainstorming literature has identified group processes that inhibit effective idea generation in groups (compare Gallupe et al., 1992, for an overview). Two of those processes are production blocking and evaluation apprehension. The former “occurs when individuals cannot express their ideas because somebody else is talking” (Gallupe et al., 1992, p. 352), the latter “occurs when individuals withhold their ideas out of concern that others may not approve of them” (ibid). These effects can be countered by employing special techniques, e.g. computer-mediated brainstorming (Gallupe et al., 1992).

Disjunctive tasks

A simple type of disjunctive task employed in small group research are disjunctive puzzles. They typically involve a group that has to arrive at one conclusion or decision to a given problem. An example of these tasks are Laughlin's rule-discovery tasks (Laughlin, 1980, 1988) during which groups have to find a rule that correctly predicts the next card in a series of game cards presented to the group through a process of theory-generation and theory-testing.

Disjunctive learning tasks

Another type of disjunctive tasks are learning tasks. In these tasks, groups are typically instructed to learn non-word letter sequences. The distribution of learning items over group members is varied (i.e. all group members receive the full list of items to remember or some items are exclusively assigned to certain group members, reducing the length of the item list that each subject has to remember). After learning, recall is tested on group level. The group is able to supply a correct answer if at least one group member can provide the correct answer. Such group memory effects, especially optimal item distribution among group members, have led to two empirically confirmed models of group memory: Model A (Lorge & Solomon, 1955) and, building on it, the model of group performance by Zajonc and Smoke (1959).

Model A specifies the probability that at least one member of a group of the size n will come up with a correctly remembered item. The model assumes that the probability that a group member will provide a correct solution is equal over all group members and states: $p_{Group} = 1 - (1 - p_{Individual})n$, where p_{Group} and $p_{Individual}$ are fixed parameters for a specific group and a specific problem (Hartmann, 1996, p. 21). It becomes clear that this model assumes that the superiority of the group in comparison to the individual increases with problem difficulty and group size. This "truth wins" - model was able to predict group performance in many simple problem solving tasks with high accuracy (Hartmann, 1996, p. 22).

Model A predicts that group performance increases if all group members know as much as possible. Since human information processing capacity is limited, Zajonc and Smoke combined Model A with a model of individual memory performance in order to "a priori predict the optimal way of distributing information in a group so as to maximize later group recall of information" (Tindale & Sheffey, 2002, p. 6). Zajonc and Smoke found that the optimal level of group performance is independent of group size and is reached if each group member receives approximately 16% more items than he or she can remember. Thus, this model promotes partial overlap of knowledge items among group members in information rich environments for “truth wins” tasks and has been empirically confirmed by Tindale and Sheffey (2002). In a similar study, Ohtsubo (2005) tested whether redundant information or partially redundant information distribution among dyads and triads is most effective for group problem solving performance. He employed a logic puzzle requiring the correct combination of seven clues to solve as the dependant variable. In dyads and triads, he either distributed all seven clues to all group members (the shared condition), or he distributed only one item to all group members and all other items exclusively to certain group members (unshared condition). Thus, in the shared condition, each individual had to learn seven items in one minute while in the unshared dyad, each individual had to learn four items, and in the unshared triad only three items within the one-minute learning period. There was no significant performance difference between the unshared and the shared dyad, but the unshared triad significantly outperformed all other groups.

Hidden-Profile tasks

Another example for a disjunctive task is the one employed by Stasser and Titus (1985, 1987) in their ‘hidden profile’ experiments. In their study, a group was asked to choose a candidate for a vacant job position. All information about all candidates was given to the group, but only some bits of information (e.g. certain candidate skills) were distributed to all group members while the remaining information was distributed exclusively to some group members in a non-overlapping way. If the group members exchanged all unshared information, the optimal candidate would be identified while the shared information alone would lead to the selection of an unsuitable candidate. Note that this experimental setting differs insofar from the previously reported studies as the candidates were required to learn from each other, i.e. share their information for making a decision. Stasser and Titus (1985; 1987) found that there exists a bias that mainly shared information is discussed and unshared information is withheld from group discussion. This widely replicated phenomenon is referred to as the ‘common knowledge effect’ (Tindale & Sheffey, 2002). According to these studies, a larger overlap of knowledge inside a team leads to better performance, because more knowledge would be shared and made available to all members.

Summary

In summary, the studies on disjunctive tasks promote a certain level of knowledge overlap inside a group for optimal group performance. For disjunctive learning tasks, optimal overlap is reported to be at 16% of items, while for hidden-profile tasks, a larger overlap conditions higher performance. However, in Ohtsubo's study, the number of clues given to each group remained constant (seven), but the amount of information that each individual had to learn differed over the conditions. In one of the shared conditions, each group member had to memorize seven clues, in another four, and in another condition three clues. Learning time was one minute regardless of the condition. The fact that the shared groups performed poorer thus appears to be somewhat artificial as one minute is merely enough to read all seven clues, let alone remembering them. In both Ohotsubo's and Tindale and Sheffey's study, both magnitude of knowledge overlap between group members and the amount of items each group member had to memorize were altered *at the same time*. As item overlap and items per subject are thus confounded, it is difficult to judge whether the effects stem from a characteristic of the group (overlap) or from group members' individual learning difficulties in the condition where individual item load was high. It would thus be desirable to keep individual characteristics at a constant level in order to determine effects on group level.

As noted earlier, the abovementioned studies employed memory recall of meaningless words or solving puzzles as tasks. The transferability of such results to organizational practice is questionable (Endres & Putz-Osterloh, 1994). Complex problem solving scenarios or microworlds represent a class of tasks that is of high practical importance as it shares characteristics with problems that politicians or managers face: complexity, dynamics, intransparency and interconnectedness (ibid, see also Section 4). These tasks belong to the category of discretionary tasks (Scholl, 1997); the influence of cognitive diversity on group performance in complex problem solving tasks is elaborated in the following section.

5.2.2 Group performance in discretionary CPS

A small number of studies on the effect of knowledge distribution on group performance have targeted complex problem solving (Endres & Putz-Osterloh, 1994; Rulke & Galskiewicz, 2000). Endres and Putz-Osterloh concluded that two classes of variables influence group performance in such tasks: individual features of group members and features of the group interaction process.

According to Endres and Putz-Osterloh (1994), on the group level, cohesion, communication structure and patterns of interaction determine group performance. They concluded from two experimental studies during which the participants worked on the TAYLORSHOP microworld (Süß & Faulhaber, 1990) that individual capabilities are a strong predictor for group performance in complex problem solving. They also found that group interaction processes differ depending on group structure, but these differences in interaction style did not lead to different outcomes. The authors hint at the possibility that the process of collaborative work on the microworld was so error-prone that these errors superimpose the effect of interaction style on performance. It should be noted that the authors did not systematically vary the amount of domain-specific knowledge in the dyads they studied, but assessed individual capabilities on the basis of an ex-post assessment of general economic knowledge.

In a different study, Rulke and Galskiewicz (2000) analyzed the effect of prior knowledge distribution on group performance in a complex management simulation game that lasted for several weeks. The game, judged by the author's description, fulfils the requirements of complexity in Dörner's sense (compare Section 4.1). Rulke and Galskiewicz found that in terms of knowledge distribution, generalists groups outperformed groups consisting of specialists because there was more knowledge overlap in generalists groups than in specialist groups, where non-overlapping knowledge resided in few specialists. These findings indicate that information shared among many group members is more likely to be retrieved in complex problem solving than unshared information. This indicates that the 'common knowledge effect' (compare previous section) is also present in CPS tasks. Rulke and Galskiewicz (2000) further argue that non-overlapping knowledge leads to difficulties in sharing and discussing and prevents optimal decision-making. However, in their understanding, generalists have "a mastery of knowledge across various domains" whereas in specialists groups knowledge is "concentrated in individuals" (p. 613). Generalists groups had more members with broader knowledge than specialist groups. Since individual knowledge of group members is a strong predictor for group performance in complex problem solving (compare Endres & Putz-Osterloh, 1994, above), and in this experiment, generalists groups included more individuals with broader knowledge than specialists groups, the finding that generalist groups outperformed specialist groups may solely be due to the individual characteristics of group members: Generalists groups are more likely to include individuals that have an especially rich knowledge of the problem domain. Furthermore, Rulke and Galskiewicz did not experimentally control for the amount of knowledge that each group member possessed, but used an ex-post classification of participants' self-assessment that casts doubt on their categorization as specialists and generalists. Thus, the amount of knowledge inside a group may have influenced the group's performance.

Another strand of research in cognitive diversity is team mental models, which identifies shared team mental models, i.e. cognitive homogeneity, as improving team performance, while another strand of team diversity research identifies "increasing team diversity as a means to improve overall team performance" (Sauer et al., 2006, p. 935). Studies from the field of shared mental models (compare Sauer et al., 2006, for an overview) "generally suggest that a strongly shared mental model is advantageous for team performance" (ibid, p. 937). An example for the assumption that cognitive homogeneity leads to performance increase is a study by Mathieu and colleagues (Mathieu et al., 2000): They hypothesized that cognitive homogeneity among two-person simulated fighter pilots would increase their performance. Sauer (2006) ascertains that the role of a shared team mental model depends on the type of task in Steiner's classification and admits at the same time that it is unclear how team performance is affected by the different types of tasks. According to Mohammed and Ringseis (2001), complex problem solving group tasks – that belong to the category of discretionary tasks – benefit from heterogeneity. Note that this assumption is contrary to Mathieu's assumption, whose simulated scenario also qualifies as a complex problem.

In order to test whether team heterogeneity or homogeneity are most profitable for team performance, Sauer (2006) analyzed cognitive diversity in teams operating a computer-based multiple task environment. Two types of cognitive diversity were examined and altered in dyads: System understanding and team specialization. System understanding was altered on an individual level; it was either trained as a deep and system-oriented understanding (SOT) or as a more superficial procedure-oriented knowledge (POT). In this way, there were SOT dyads, POT dyads, and mixed dyads (MIX). Teams were also altered with regard to their training about dealing with errors. A total of twelve errors were trained. In one condition, both team members received training on all twelve errors, while in the specialist condition; each team member received a more intensive training on six (non-overlapping) errors. MIX Teams achieved higher performance scores and generalist teams detected more errors than specialized teams.

5.3 Intermediate Summary

To sum up findings so far: the results from disjunctive tasks favored a partial knowledge overlap among group members for optimal group performance. The abovementioned findings from the field of discretionary tasks indicate that cognitive homogeneity (i.e. knowledge overlap) inside the group is proportional to group performance, but the amount of knowledge available on group level may also be of importance. Since some of the above-mentioned studies did not control for the actually acquired problem-relevant knowledge and also indicated that individual knowledge of group members is a strong predictor of complex problem solving group performance, insights with reference to effects of knowledge overlap in groups on complex problem solving performance are vague. In cognitive diversity research, one strand of research claims that cognitive homogeneity leads to increased team performance, while another strand of research claims that heterogeneity delivers better team results.

Based on these limitations and contradictions of previous studies on group knowledge and complex problem solving, it is desirable to experimentally control the specific domain-relevant individual knowledge in complex problem solving and its distribution among a group. It would further be desirable to keep the amount of problem-relevant knowledge per individual constant and to solely alter the amount of knowledge overlap and the amount of knowledge inside the group in order to rule out the influence of individual performance on task performance and to be able to explain differences in group performance solely with group characteristics. A systematic variation of the amount of knowledge at group level could reveal its influence and clarify Rulke and Galaskiewicz's (2000) findings.

5.4 Scholl's Model of Team Effectiveness

In his model of team effectiveness, Scholl (1996, 2005) formulated hypotheses with regard to the effect of knowledge acquisition and distribution in groups on group complex problem solving performance. He argued that positive effects of teamwork are based on the increase of relevant knowledge in the team, and that the magnitude of this effect is proportional to the complexity of the problem, i.e. the more complex the problem, the higher the impact of knowledge increase in teamwork. In other words, knowledge increase among group members during task performance leads to task performance increase.

Scholl (1996) further argues that knowledge increase through discussion depends on the cognitive overlap among group members: If it is very large, there is little that people can learn from each other, and if it is very small as in highly heterogeneous or diverse groups, people can hardly understand each other due to different mindsets. This idea resembles Polanyi's concept of "interpretative frameworks" (Polanyi, 1958, p. 74): Transmitted information carries the interpretative framework of the sender, i.e. his or her prior knowledge, beliefs, and context (compare also Rauchhaupt's knowledge definition in Section 2.1.1). In order to transform the transmitted information into knowledge, it has to be interpreted through the receiver's framework. If the interpretative frameworks are very different, knowledge acquisition cannot take place. If the conveyed information is equal to information that the receiver already possesses, no knowledge acquisition can take place either, because there is nothing new to learn. Thus, the relationship between knowledge overlap and knowledge increase must be of a second-degree polynomial kind. This proposed relation between knowledge heterogeneity and performance is also based on the argument that cognitive heterogeneity and group potential have an asymptotical rising relationship which is superimposed by a exponentially growing relationship between knowledge heterogeneity and processes inhibiting actual group performance (Scholl, 1996, p. 137), compare Figure 5.1.

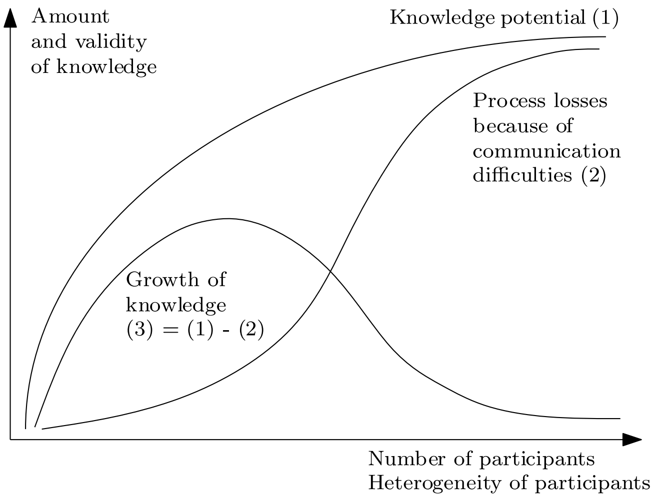


Figure 5.1: Scholl's model of team effectiveness assuming a polynomial relationship between cognitive heterogeneity of participants, amount of knowledge, and knowledge increase (adapted with permission from Scholl, 1996, p. 137). Copyright 1996 From Effective teamwork - a theoretical model and a test in the field by Wolfgang Scholl (in Understanding Group Behavior, Vol. 2 by Erich Witte and James H. Davis (Eds.)). Reproduced by permission of Routledge, Inc., a division of Informa plc.

Note that Scholl's model integrates opposing views that either claim a superiority of cognitive homogeneity or a superiority of heterogeneity by assuming an n-shaped relation between group cognitive heterogeneity and group performance.

5.5 Hypotheses

The assumptions of Scholl's model of team effectiveness in the context of complex problem solving are rephrased into the following hypotheses:

Hypothesis 5.5.1 *In complex problem solving groups, the extent of knowledge overlap among group members has a curvilinear n-shaped impact on knowledge increase among group members.*

If knowledge increase during the performance acts on complex problem solving performance, complex problem solving performance and knowledge overlap should display the same relationship. Based on Hypothesis 5.5.1, it is stated:

Hypothesis 5.5.2 *In complex problem solving groups, knowledge increase among group members increases group CPS performance.*

If a curvilinear n-shaped relationship between knowledge overlap and knowledge increase is visible and knowledge increase correlates with group CPS performance, the following should also hold true:

Hypothesis 5.5.3 *CPS performance on group level displays an n-shaped relation with knowledge overlap inside the group.*

Chapter 6

Prediction of Collaborative Performance with an IT-System

The previous chapter concluded that knowledge distribution among group members influences a group's performance in complex problem solving. This chapter will introduce an IT-system that allows to determine the distribution of knowledge among a group, allowing predictions on the group's performance. Therefore, the development of an algorithm for calculating the similarity of self-assessed knowledge between individuals is outlined. This algorithm is implemented into an IT-system, the skillMap[1], which is introduced first. After system introduction, regulations for the calculation of the similarity measure with regard to the system are explained. The chapter ends with predictions regarding the connection between the proposed similarity measure and group performance.

6.1 IT-based Organizational Knowledge Management

Until now, it has been referred to individual knowledge and knowledge in small groups. From an organizational point of view, creation and use of both individual knowledge and knowledge at higher ontological dimensions such as project groups or teams needs to be supported (Nonaka & Takeuchi, 1995).

[1]The skillMap was originally conceived by Sarah Spiekermann and has been developed jointly since the original idea. Sarah Spiekermann also participated in the editing of Sections 6.1.2, 6.1.3, and 6.1.6 which have been previously published together (Meyer, Spiekermann, & Hertlein, 2005; Meyer & Spiekermann, 2006a, 2006b).

Individual knowledge is a precondition for organizational knowledge that results from the publication of technical and/or individual knowledge and of its consolidation in organizational communication structures (Klimecki & Thomae, 2000). This consolidation of individual knowledge in organizational structures (e.g. in methods, models, documentation, and culture) is also referred to as the organizational knowledge base (Rehäuser & Krcmar, 1996, p. 15). According to Damerow and Lefèvre (1998), such external representations have the same psychological functions as internal, individual representations and are based on the same mental capabilities. Individual knowledge enlarges organizational knowledge (Amelingmeyer, 2004, p. 122) and individual learning is a central element in organizational learning (Argyris & Schön, 1996). Thus, organizational knowledge is based on individual knowledge.

Knowledge management (KM) is defined as "the process of continuously creating new knowledge, disseminating it widely through the organization, and embodying it quickly in new products/services, technologies and systems" (Takeuchi et al., 2004, p. ix). KM has become an important management concept in the change towards knowledge-based economies (compare Chapter 1), that see knowledge as "the one sure source of lasting competitive advantage" (Takeuchi et al., 2004, p. 29). The aim of the European Union to become the world's leading knowledge-based economy (European Council, 2000) underlines the political importance of this view. KM approaches are thus considered to form the basis for innovation and economic success (Davenport & Prusak, 1998; Dick & Wehner, 2002; Nonaka & Takeuchi, 1995; Scholl, 2004; Takeuchi et al., 2004). Especially in the works of Nonaka, Takeuchi, and colleagues, the concept of non-explicit knowledge is considered to play an important role in organizational knowledge creation and innovation (Nonaka & Konno, 1998; Nonaka & Reinmoeller, 2000; Nonaka & Takeuchi, 1995; Nonaka, Toyama, & Konno, 2000; Takeuchi et al., 2004). Nonaka's popular SECI model (Nonaka & Takeuchi, 1995), which sees the cyclic processes of knowledge conversion from tacit to explicit and back over different levels (individual, group, organization) at the heart of innovation, has received widespread attention (Snowden, 2002).

According to Nonaka and Takeuchi (1995), organizational knowledge creation takes place through a continuous process of knowledge transformation: From tacit to explicit knowledge and back, from the individual level to group to organizational level. The SECI-model postulates the following processes: Individual tacit knowledge is shared between two individuals through socialization. An indiviual can, through metaphors and analogies, make his or her (newly) acquired knowledge explicit through externalization. This explicit knowledge can then be combined with other existing knowledge, e.g. on the team level, to new explicit knowledge. This new explicit knowledge can then be transformed into tacit knowledge through internalization and can transcend into tacit knowledge on group level (compare Figure 6.1).

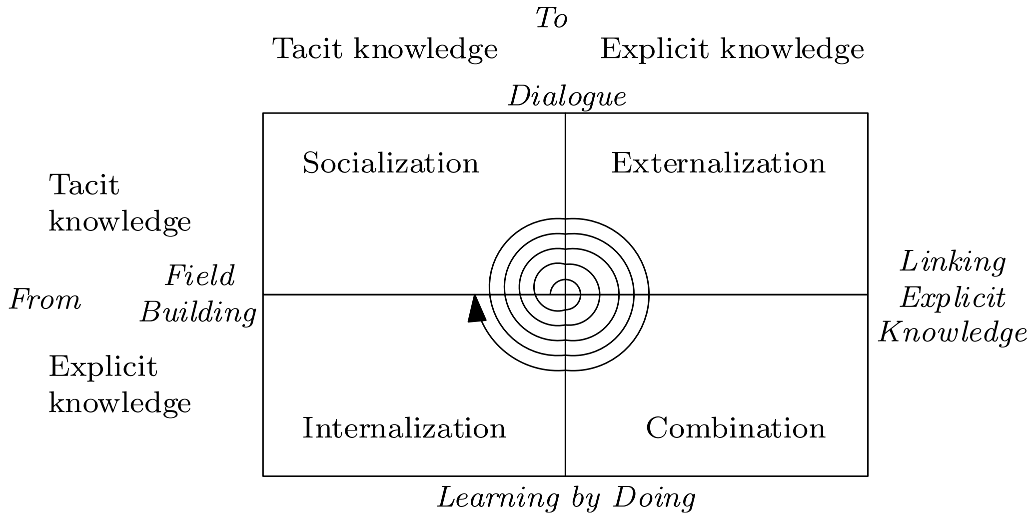


Figure 6.1: The SECI-Model of organizational knowledge creation (based on Nonaka & Takeuchi, 1995, p. 62 & p. 71)

6.1.1 Organizational Knowledge as Structural Connectivity Pattern

The knowledge potentially available in an organization goes beyond individual knowledge and is contained in the organizational knowledge base: "The organizational knowledge base is comprised of several different individual and collective knowledge repositories that can be accessed for task completion. It spans the data and information repositories, on which individual and organizational knowledge is built " (Probst, Raub, & Romhardt, 1999, p. 46, own translation). Knowledge Management can also be defined as the organization and design of the organizational knowledge base (Probst et al., 1999, p. 47).

According to Trier's knowledge management entity model (2005), the organizational knowledge base is mainly made up of the following knowledge entities: Individuals (employees), processes, documents, and topics. These entities are linked together into a network through communication and technology (i.e. references in information systems). Trier's model further states that concrete knowledge management measures target the establishment of new links between those entities or the modification of existing ones. For example, the support and/or establishment of communities of practice, informal groups of experts that exchange knowledge on a specific topic that are seen as key drivers of innovation processes (Wenger, 2003; Scholl, 2004), can be conceptualized as establishing links between certain individuals and topics. Based on this model, the organizational knowledge base can be understood as a structural connectivity pattern (compare definition in Chapter 2), and knowledge management targets the alteration of this structure on an organizational level.

6.1.2 IT-based organizational knowledge management

Knowledge management processes in organizations are usually supported by IT-based knowledge management systems, a heterogeneous group of systems (Gronau, 2002), whose objective is to "support creation, transfer, and application of knowledge in organizations" (Alavi & Leidner, 2001, p. 107). They can be divided into two main categories (Bush & Tiwana, 2005): Knowledge repositories focusing on the externalization and storage of explicit knowledge (see also Schütt, 2003) and knowledge networks, focussing on the establishment of individual contact for knowledge exchange through personal interaction (Alavi & Leidner, 2001). The strategy behind knowledge repositories has also been referred to as the codification approach (Hansen, Nohria, & Tierney, 1999), because it implies codification of individual explicit knowledge into documents and data bases which can be accessed from anywhere inside the organization. Note that this view employs the terms knowledge and information synonymously (compare Rauchhaupt's knowledge definition on p. 6 for an explanation for this synonymous use). The strategy behind knowledge networks is also referred to as personalization approach (Hansen et al., 1999), because it enables personal contact.

Codification-only approaches have often turned out as failures because they do not take the personal context of individuals into account (Schütt, 2003). In their concept of ba, Nonaka and Konno (1998) state that the exchange of both tacit and explicit knowledge between individuals is optimal in a shared physical context. Ba is "... a context which harbors meaning..., a shared space that serves as a foundation for knowledge creation" (Nonaka & Konno, 1998, p, 40). Nonaka and Konno (ibid) define four different bas for different stages of knowledge creation that correspond to the four cycles of knowledge creation of the SECI-Model (see above): Originating ba for face-to-face interaction during socialization, interacting ba for peer-to-peer interaction during externalization, cyber ba in group-to-group interaction during combination, and exercising ba for on-the-site interaction during internalization. Despite the fact that a physically shared space for interaction is optimal, individuals cannot always get together in order to exchange knowledge in virtual or complex organizations. In this case, the shared context between individuals is reduced to shared “interpretative frameworks” (Polanyi, 1958, p. 74) that originate in shared knowledge. This shared context can be described as shared activities, social relations among organizational members, fields of work and expertise (“topics”) and related documents: as links in the structural connectivity pattern that comprises organizational knowledge (see above).

6.1.3 Graph representations of organizational knowledge

In the same way as graphs can be used to represent individual structural connectivity patterns (compare Section 3.1), they can be used to visualize the structural connectivity patterns that comprise the organizational knowledge base (see above). The representation, display and modification of such an organizational knowledge graph thus poses an operationalization of cyber ba.

Graph-based knowledge management systems model the interrelations between entities that comprise the organizational knowledge base and attempt drawing conclusions from those. Graph-based systems contain vertices (or nodes) representing knowledge entities and edges (or arcs) between them (compare Section 3.1). By visualizing semantic relations between entities, graph-based knowledge representations allow capturing complex knowledge domains. Examples include creativity support systems like D-ABDUCTOR (Sugiyama & Misue, 1996), knowledge networks that help to structure data according to semantic relationships such as theBrain© or social software such as Orkut©, LinkedIn©, or Xing©.

Social networks contained in these systems are created by the means of social network analysis (SNA). This can either be done by manually specifying relations to others or by data-mining techniques on logged point-to-point communication data, such as e-mail traffic, web-based schedules and IM-protocols. In a company context, social networks are considered as an adequate representation of knowledge creation (Jansen, 2004), because they reveal how “work really gets done in organizations” (Cross & Parker, 2004, title page) and how information search and retrieval is carried out throughout firms (Hansen et al., 1999). Thus, knowledge bearers and knowledge dissemination in firms can be identified on the basis of social networks, increasing knowledge transparency and facilitating knowledge exchange (Cross & Parker, 2004).

In the following sections, three knowledge management systems that visualize knowledge networks as a graph and offer opportunities for its modification for KM activities are introduced: KnowWho, Commetrix, and skillMap.

Afterwards, a method for assessing knowledge heterogeneity between individuals in the skillMap system is developed which can be used for performance prediction for knowledge-intensive work.

6.1.4 KnowWho

Fujitsu's KnowWho-System (Igata, Tsuda, Katayama, & Kozakura, 2003; Tsuda, 2003) visualizes the semantic structure of products and innovations at Fujitsu as a graph: Products are connected to the product they have evolved from (i.e., fingerprint scanners are connected to flatbed scanners because they are based on flatbed scanning technology). Selecting a product switches to a social network view with the expert on the selected product as the central node (compare Figure 6.2). With reference to Trier's knowledge entities (see above), KnowWho visualizes the connections between employees, between employees and topics (products in this case) and documents (associated with products). The system's visualization engine is based on the static visualization engine from the D-ABDUCTOR system (Sugiyama & Misue, 1996). The data for KnowWho's social network is automatically retrieved from Fujitsu's repositories. Meeting schedules and joint publications of technical documents provide the basis for network construction. The semantic network of product-interrelations is built via data-mining on technical documents. Yet, however advanced the data-mining behind KnowWho, its visualization has certain limitations. Generally, graph based systems face the specific challenge to integrate a highly comprehensive visualization. Sugiyama (2002, p. 11) gives a detailed overview of graph drawing rules and frameworks. Graphs typically have to obey a set of 20 static structural rules (e.g. placement of high degree vertices in the centre and identical drawing of isomorphic sub graphs) and have to be editable and changeable dynamically in such a way that "the human mental map of a graph" (Sugiyama, 2002, p. 11) and the continuity of cognition is preserved. The issue with Fujitsu's System is that it does not preserve the continuity of cognition because clicking a node in either network switches to a static new layout of the graph, and the transition between the two states of the network are not presented in a continuous way.

6.1.5 Commetrix

Commetrix (Trier, 2005) is another example of a graph-based system that seeks to support organizational knowledge processes through graph visualization. It is primarily focussed on identifying and supporting intra-organizational expert groups (communities of practice). "It aims at eliciting the structure of the expert group by analyzing the communication networks stored in data

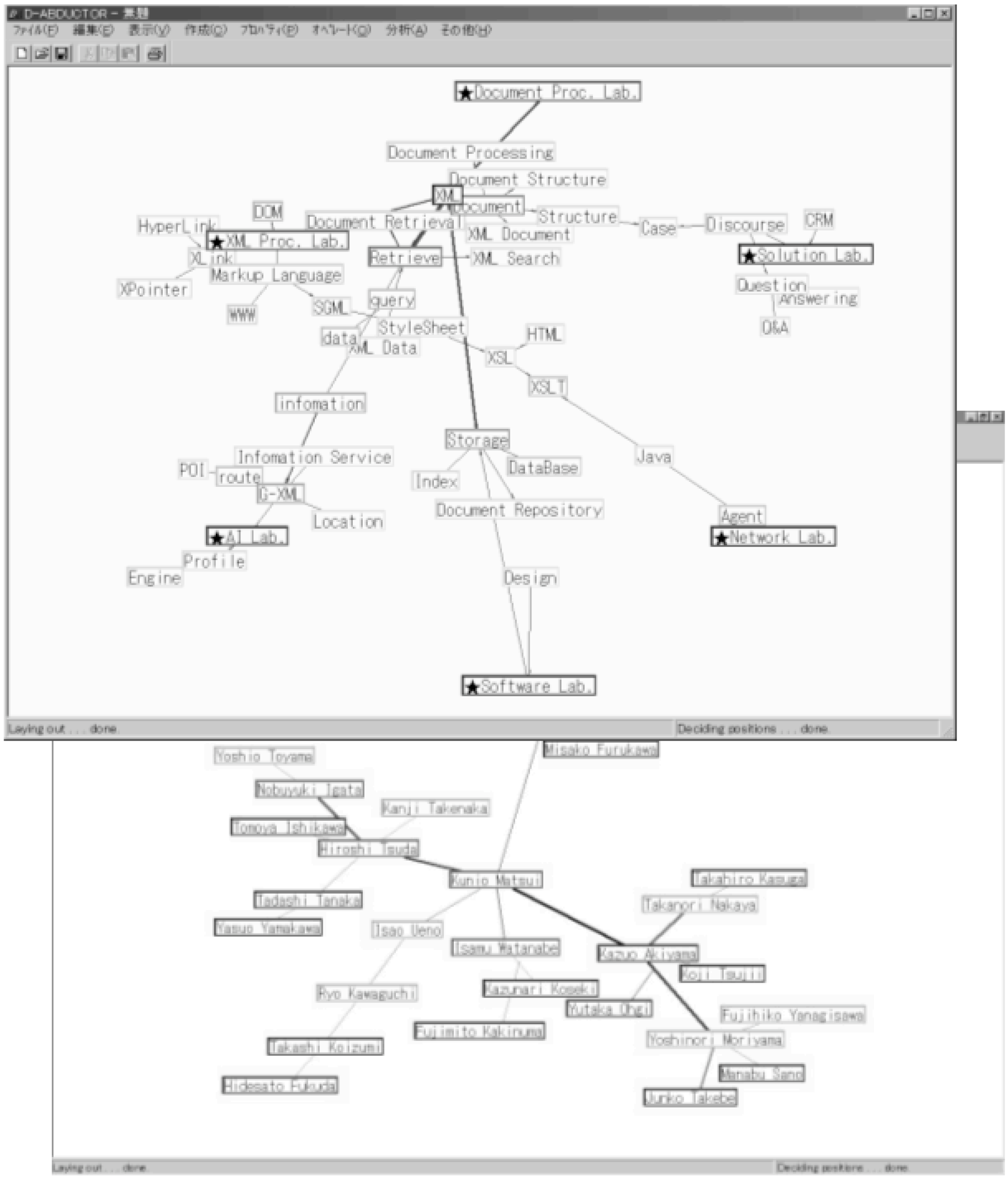


Figure 6.2: Fujitsu's KnowWho Human Knowledge Navigator (Igata et al., 2003). The top window displays the semantic structure of related products, services, and technologies. Clicking a node brings up a second window (below) that displays a social network with the expert on the selected technology in the center.

archives.... This set of data can be visualized as an actual network of experts" (Trier, 2005, p. 979). The system displays individuals (employees) as nodes and e-mail communication between individuals as edges between nodes. The more frequently two individuals communicate, the thicker the edge between them. The edges are labeled with the most frequent topics of communication between two individuals that are obtained through data-mining techniques.

Commetrix includes only employees as graph nodes. Edges between employees are labelled with topics, but contrary to KnowWho, these topics are not semantically related as nodes of a graph. The advantage of organizing topics semantically lies in the ability to discover related topics and in making statements on the degree of their relatedness, which can be modelled through path distance. In Commetrix's visualization, it is not possible to visualize topics that are semantically closely related to the topic of a knowledge exchange between two individuals. However, these closely related topics may be the subject of frequent communication of other employees that are linked to the individuals in the visualization. Thus, the actual size of the community of practice may be underestimated. Another advantage of visualizing topics as semantically connected nodes instead of labels of communication edges lies in the possibility of discovering undesired parallel developments, i.e. two people working on the same topic without being aware of each other (that is, without having an edge between them that symbolizes communication). Such a constellation can be visualized as a topic node with edges to two individuals that don't have a communication edge between them. Commetrix may be very useful for identifying communities of practices on a strict topic within organizations, but it may not be as suitable for discovering such parallel developments.

6.1.6 skillMap

In this section, the skillMap-System (Meyer et al., 2005; Meyer & Spiekermann, 2006a, 2006b) and its different knowledge visualizations are described. Its description is followed by the introduction of an algorithm for knowledge similarity calculation.

System design

The system consists of two interlinked components: A social network of individuals and a graph (semi-lattice) of the individual fields of expertise or knowledge. Figure 6.3 illustrates this structure.

With reference to Trier's knowledge management entity model (see above), the skillMap includes employees and topics as nodes and is thus able to vi-

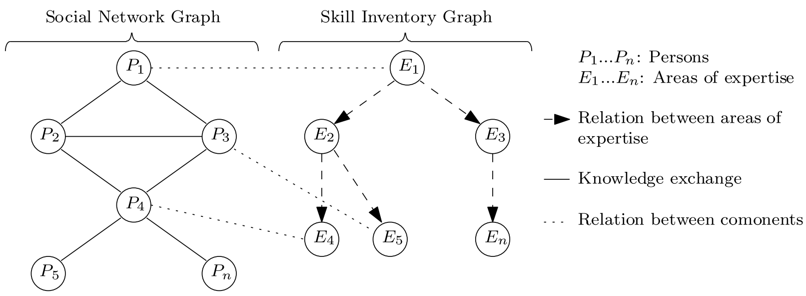


Figure 6.3: Schematic representation of the skillMap system (Meyer et al., 2005)

sualize the semantic relation between both. The skill inventory graph can be edited by all users of the system. As users edit and amend the graph freely for their individual fields of expertise, the skill inventory graph converges towards a stable status of structure consent. The result is a self-assessment of the entire organizational expertise. The inclusion of knowledge entities as different node types makes the system capable of visualizing shared organizational context (see Figure 6.4). The skillMap visualizes this context with the help of its two-component design as follows: Person-nodes from the social network graph are connected to the expertise nodes in the skill inventory graph according to individual expertise with undirected edges. The users themselves specify these connections manually. In this way, a person vertex from the social network graph is linked to all the fields of expertise that he or she has in the skill inventory and to all documents that he or she authored. The newly combined graph is referred to as skillMap. Placing an employee node in the centre of the visualization (compare Figure 6.4) thus visualizes the employees' organizational context.

Assigning people to competencies is not a new concept and is already in practice with yellow pages systems (Ruggles, 1999). However, yellow pages systems only store a link between two objects (person and skill(s)) without further connecting the objects at the ends of the connecting line. By merging the social graph and the skill inventory graph into one skillMap, several uses of the system become evident:

1. The system can visualize who-knows-who and can report a line of connections from any given person inside the organization to another.

2. The system can visualize who-knows-what by displaying the fields of work and expertise for any member of the organization.

3. By visualizing persons that are connected to the same expertise node but have no social edge between them, the user can identify members

of the organization that work in the same field without knowing each other (e.g. P1 and P4 in Fgure 6.3, compare Figure 6.7 for an example).

Especially the last characteristic is a benefit new to KM systems because it can reduce undesired parallel developments and thus poses a possibility to increase organizational performance.

Visualization and technical background

The skillMap is implemented in Java as a distributed client-server web based application and relies on JavaWebStart for launch. Continuity of cognition in the skillMap is achieved by smoothly animating all changes of graph layout. Therefore, preFuse (Heer, 2004; Heer, Card, & Landay, 2005), a toolkit for interactive information visualization, was chosen. Its capability of dealing with a vast number of nodes as well as its impressive graphic processing performance were crucial for the decision for preFuse, which is used for creating skillMap's graph structure and for its visualization and navigation. PreFuse has been designed in order to "augment human cognition by leveraging human visual capabilities to make sense of abstract information, providing means by which humans with constant perceptual abilities can grapple with increasing hordes of data" (Heer et al., 2005, p. 1) and it is hoped that the continuity of cognition will be achieved. For the skillMap and social network views, the user can choose between a static Fruchterman and Rheingold layout and a dynamic fore-directed layout (compare Sugiyama, 2002, for a description of both), both supplied by preFuse. Especially the constantly hovering spring embedded algorithm has received positive feedback from beta testers because of its playful appearance.

A platform independent standard for describing the structure of the graph was chosen in order to allow simple data exchange with other applications. XGMML (Punin & Krishnamoorthy, 2001), an XML-derivate used in many graph-based applications, was chosen as standardized data exchange format.

The underlying graph structure is stored in a SQL-database from which the XGMML representation of the graph is generated by the server and transmitted to the client. Figures 6.4 to 6.7 give examples of the skillMap's different visualizations.

6.2 Calculating Knowledge Similarity with the skillMap

In order to determine knowledge-based similarities between individuals represented in the skillMap, a skillMap is conceptualized as a graph in the fol-

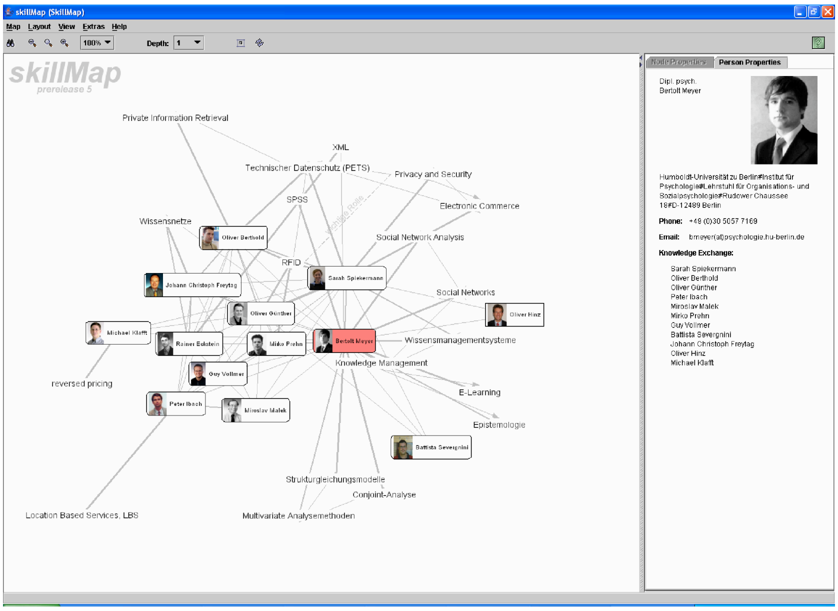


Figure 6.4: skillMap system interface after login. The first-degree social network of the user, his or her fields of work and expertise and the fields of work and expertise of his or her colleagues are visible.

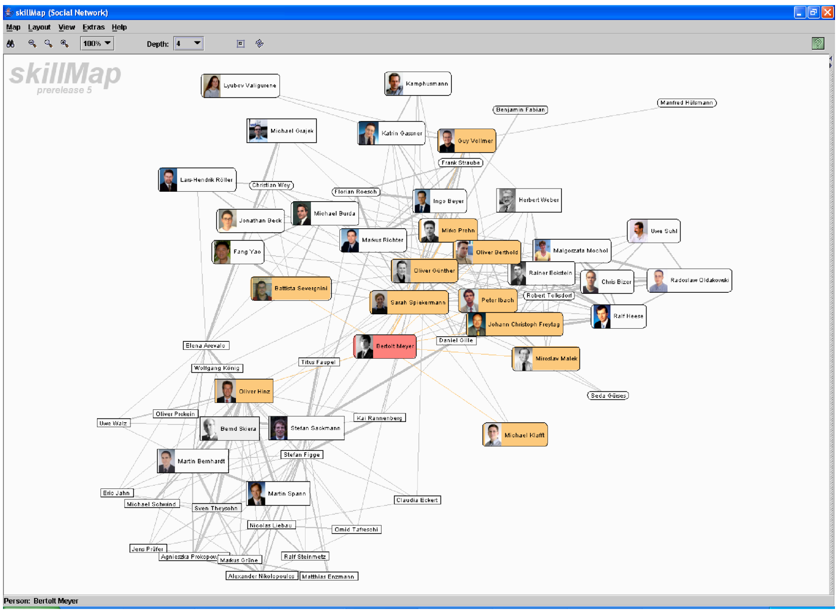


Figure 6.5: skillMap visualization of of the social network graph. The own node is displayed in red, all other directly connected individuals are in yellow

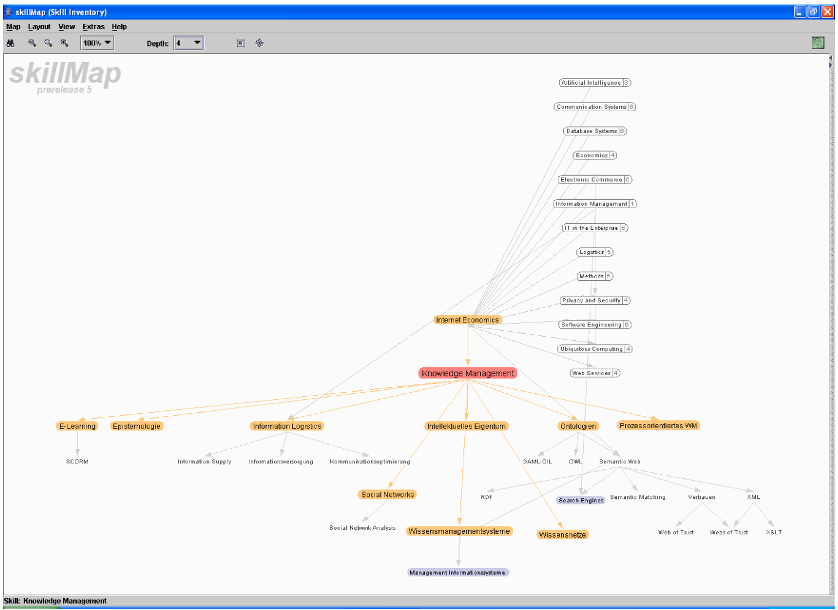


Figure 6.6: skillMap visualization of the skill inventory graph with the ‘knowledge management’ node in the centre.

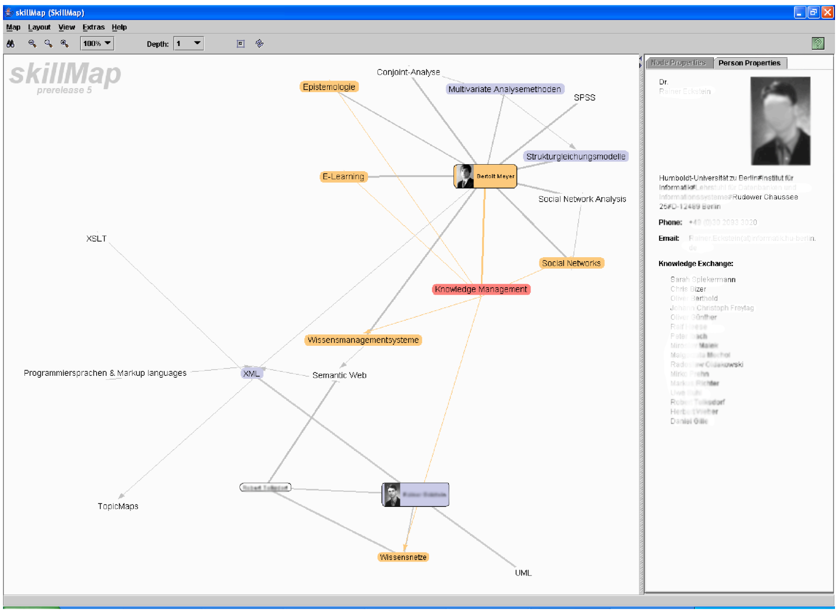


Figure 6.7: skillMap visualization of a possible unrecognized parallel development: Two individuals work with the same technology (XML) and have expertise in similar fields but do not know each other.

lowing section and a way of calculating the similarity between two individuals represented in the skillMap system is introduced[2].

6.2.1 The skillMap as a graph with two node types

A skillMap is defined as a labeled directed graph G (compare definitions 1 to 6 on p. 33) with two different types of vertices, referred to as *persons* p and *skills* s: $G = (V_p \cup V_s, A, L^{(exp)}, L^{(depth)})$ where

- $A_p = A \cap (V_p \times V_p)$
- $A_s = A \cap (V_s \times V_s)$
- $A_{ps} = A \cap (V_p \times V_s)$
- $V_p \cap V_s = \emptyset$
- $(p, s) \in E$ for $p \in V_p, s \in V_s \Rightarrow (s, p) \in E$
- $L^{(exp)}$:= label for edges $\in A_{ps}$, called *expertise*.
- $L^{(depth)}$:= label for edges $\in A_s$, called *depth*.

6.2.2 Similarity calculation

In order to identify possible parallel work between all persons stored in the skillMap, the skill similarity calculation is introduced. The aim of this calculation is to quantify the similarity of individuals based on the skills they share. This skill similarity (denoted *sksim*) ranges from 0 to 1, where 1 denotes that both persons are incident to the same skills without any further incident paths from them. The similarity between two persons is 0 if there are no paths between these persons with only skills on them. For a given skillMap G, the calculation produces a similarity index for each pair of person nodes (v, u). It will thus be possible to obtain a similarity matrix for all individuals represented in the graph which can then be used for analysis and layout, for example for an MDS scaling procedure (compare Section 3.6). In this way, the skillMap can be laid out according to the similarity of persons. Furthermore, this measure can be used for identifying persons similar to a specific person by listing similarity indexes of all persons in reference to the

[2]I am very grateful to Sarah Spiekermann, Manuel Hertlein, and Tobias Lattke for their valuable comments and feedback on the similarity calculations. I am especially grateful to Hans-Florian Geerdes for his help with the mathematical formalisms.

target person in ascending order. An average similarity measure over a group of individuals could be used for the prediction of group performance.

Such a similarity calculation has been developed in the reviewed version of this thesis. As it will be included in future commercial versions of the skillMap system, notions of confidentiality prevent its publication in the public version of this thesis. The suggested similarity measure is refered to as *sksim* and has the following properties:

Properties of sksim

- $sksim(u, v) = sksim(v, u)$
- $sksim \in [0, 1]$
- $sksim(u, v) = 1$ if $\sigma(v) = \sigma(u)$
- $sksim = 0$ if u and v are not connected in G'

The suggested similarity measure is implemented in a Java command-line application[3] that parses a skillMap XML-file that holds a particular skillMap graph and returns a matrix containing pairwise similarity ratings for all pairs of person vertices included in the system. In order to test the similarity calculation, a test graph is implemented in the skillMap system (compare Figure 6.8). The output of the application is in line with the suggested calculation:

```
similarity(P4, P3) = 0.2888889
similarity(P1, P4) = 0.22222222
similarity(P3, P2) = 0.22222222
similarity(P3, P1) = 0.22222222
```

6.3 Summary and Hypotheses

The sksim similarity measure calculates the similarity of knowledge-self assessment between individuals represented in the skillMap knowledge management system. If knowledge overlap in a small group has an influence on its performance in complex problem solving as argued in Chapter 5, the similarity measure obtained with the method above should display some connection with CPS performance. Due to the fact that the similarity measure is

[3] I am very grateful to Tobias Lattke for implementing and programming the similarity calculations.

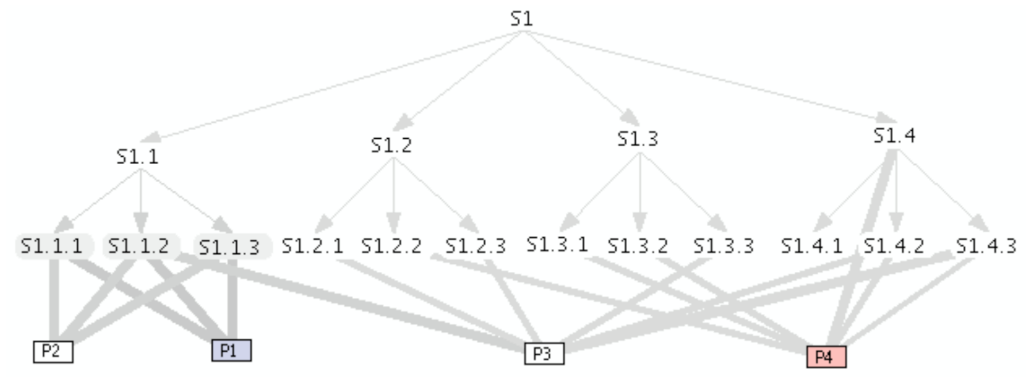


Figure 6.8: skillMap test graph for similarity calculations

supposed to increase with intra-group cognitive homogeneity, and a medium cognitive heterogeneity is supposed to lead to highest problem solving performance, a linear relation between the similarity measure and group performance cannot be assumed: Low and high similarities are supposed to coincide with low group performance, medium similarities are supposed to coincide with high performance scores.

Hypothesis 6.3.1 *For small groups in which each group member assesses his or her knowledge on a dynamic system in the skillMap knowledge management system, the proposed similarity measure for intra-group knowledge overlap displays a curvilinear n-shaped relation with CPS performance.*

Chapter 7

Method

This chapter has two main sections. In the first one, preliminary experiments are described that determine whether the AST is suitable for the elicitation of structural knowledge. The second section will outline a laboratory experiment that aims at testing hypotheses from Chapter 4 with regard to implicit and structural knowledge and complex problem solving performance and from Chapters 5 and 6 with regard to group knowledge distribution and group performance.

7.1 AST Validation

In this section, three pilot studies aiming at AST validation are presented. They are based on Hypotheses 3.9.1 to 3.9.5. For hypotheses 3.9.1 through 3.9.4, an experiment will be conducted. Hypothesis 3.9.5 is tested as part of a larger experiment that also aims at testing the hypotheses from Chapter 4. It is presented in the second section of this chapter.

7.1.1 Labeled and unlabeled AST testing

The first pilot study was conducted in order to test the assumption that the AST actually taps on the implicit knowledge domain. Hypothesis 3.9.1 assumes that AST-elicited knowledge graphs will contain more edges if edges are placed in a quick, intuitive, and unlabeled way by experiment participants. In order to test this assumption, 26 participants of a graduate seminar on organizational psychology at the department of psychology at Humboldt University Berlin completed the AST on the seminar's subject as stimulus concept ('Wissensmanagement', German term for 'knowledge management') at the end of the semester. As the seminar involved many practical tasks and role-play situations, participants were expected to have acquired some structural implicitness. The AST was presented in two different versions and participants were assigned randomly. In one condition, participants worked on the AST as described in Section 3.9 (standard condition). In the other condition (explicit condition, compare Section 3.9.1), the AST differed insofar as in the second part on pairwise concept comparisons, participants were only allowed to indicate a relationship between concepts if they were able to explicitly label the kind of connection. This was enforced by displaying a text box in which the nature of the connection had to be entered if a value greater than ‘0’ had been chosen. The standard condition should thus contain edges for which subjects would have been able to provide an explicit label as well as those which could not be labeled while the explicit condition is supposed to capture only the former type of edges. Thus, the number of edges elicited in the explicit condition should be smaller than the number of edges elicited in the standard condition (compare Hypothesis 3.9.1) while the overall number of associated concepts should be equal between the two modes since the section on term association was identical in both conditions.

7.1.2 AST sensitivity to learning

The second pilot study was conducted in order to determine the sensitivity of the AST to learning. Hypothesis 3.9.2 assumes that all six coefficients display a shift towards their respective more favorable states if learning occurs. In order to test this assumption, the participants of a graduate seminar at the Institute of Psychology at Humboldt-University Berlin completed the AST with the seminar’s topic (Wissensmanagement, German for knowledge management) as stimulus concept in the first session and in the last session of the seminar. The time span between the first and the second test was two months. Note that this was a different seminar than the one mentioned in the previous subsection. In this seminar, the focus was mainly on the acquisition of academic explicit knowledge. The AST’s parameter were set to PFNET$(n-1,\infty)$ and a maximum number of 25 concepts entered pairwise comparisons. If one assumes that at least some learning took place, at least those AST-parameters that are thought to elicit knowledge at different levels of explicitness (number of associated concepts, number of edges, number of clusters, diameter, and RMPL) should display an increase from the first to the second measurement.

7.1.3 Convergent validity

In order to establish construct validity (compare Section 3.9.4), the convergent validity (the degree to which concepts that are related theoretically are interrelated empirically, compare Campbell & Fiske, 1959) was surveyed by comparing AST scores to SLT scores. The Heidelberger Strukturlegetechnik SLT (Scheele & Groeben, 1984, 1988) is a standardized manual concept mapping tool for structural knowledge elicitation which builds on Shavelson's concept mapping techniques (Shavelson, 1972; Shavelson & Stanton, 1975). When applying this technique, participants are asked to visualize a knowledge structure by labelling cards and connecting them with labelled edges (compare Section 3.5.2 for further details on the SLT technique). In order to analyze the AST's validity, AST parameters are compared to SLT parameters. Since the SLT technique requires subjects to label the placed edges, both versions of the AST are compared with the SLT procedure: The standard AST that asks for quick and intuitive judgements for placement of unlabeled edges, and the more explicit version introduced in Section 3.9.1 that requires participants to label relations of strength above 0 they place during the pairwise comparison task.

Positive correlations between both AST versions and the SLT are expected as all three of them are intended to elicit structural knowledge. Higher correlations between the explicit version of the AST and SLT as compared to the correlation between the standard version of the AST and SLT are expected, because the standard AST is supposed to tap into implicit use of structural knowledge in a stronger extent than the explicit version.

In order to obtain a comparison between AST and SLT scores, both tests are employed on the same subset of associations. Therefore, the AST system is modified in such a way that the term association task is split from the pairwise word comparison task. After term association, the AST system stops and saves the associations in a database for further use. At a second point in time, pairwise similarity ratings can be performed, but the associated concepts can also be read from the database for other analysis purposes.

Undergraduate psychology students at Humboldt University and University of Potsdam who were a week short of their exam on biological psychology were asked to associate concepts to ten stimulus concepts from the field of biological psychology. The stimulus concepts had been selected as relevant by the professor for biological psychology at Potsdam University. The students then associated to these ten concepts in the association module of the AST and were dismissed. In a second session, they were then asked to perform an SLT on the concepts they had previously associated to the term 'nervous system', and in a third session they were asked to do pairwise comparisons in the AST comparison module for the same associations. The order of SLT and AST tasks in sessions two and three was randomized. Half of the subjects worked on the AST in the standard mode as presented in Section 3.9, and the other half worked on the explicit version as presented in Section 3.9.1. In this way, the number of associated concepts was equal for both tests, and the measures obtainable for both tests (number of edges, density, diameter, and RMPL) are correlated.

7.1.4 Congruent validity and discriminant validity

As the AST is supposed to predict performance in real-life settings, another experiment outside the university setting is conducted in order to determine whether AST-elicited structural knowledge measures that have been obtained for a certain domain correlate with performance measures that indicate achievement in that particular domain as stated in Hypothesis 3.9.4.

At a vocational school for cleaners and plumbers in Berlin[1], students were asked to complete a regular AST on a stimulus concept identified as crucial to their current state of learning by the teacher. For the cleaners, this stimulus was 'Detaschur' (a German professional term for the cleaning of textile floor surfaces), and for the carpenters, it was 'Dach' (roof). The students worked on the AST a few days after the completion of a written exam that was focused on the stimulus concepts. These exams included writing tasks but also practical tasks such as calculating a floor area or dimensioning a roof. AST-elicited coefficients were then correlated with the total grade for all students from both subject areas. Further sample details are presented in the results section.

This experiment also offered the possibility to evaluate the discriminant validity (the degree to which concepts that are theoretically independent are so empirically, compare Campbell & Fiske, 1959) of the AST by comparing it to a German vocabulary test. Vocabulary tests correlate highly with the general intelligence (g; Neisser et al., 1996) and are thus suited for obtaining a quick estimate of g. In a study comparing different German vocabulary tests with the general IQ as obtained by the German Hamburg Wechsler Intelligence Test for Adults (HAWIE-R), Satzger, Fessmann, and Engel (2002) determined that the German vocabulary test (Wortschatztest) WST (K.-H. Schmidt & Metzler, 1992) delivers the highest correlations with the HAWIE-R IQ. In order to determine the interdependency of the AST's coefficients with intelligence measures, the students, prior to completing the

[1] I am very grateful to Siegfried Richter who suggested to test his students and who supported the experiment with his motivation and determination.

AST, filled out a WST, which consists of 42 items, each consisting of five words. Of these five words, one actually exists and the others do not. The subject is asked to tick the existing word; difficulty increases from item to item. There is no time limit. WST scores are then correlated with AST scores and grades. Due to the fact that intelligence and knowledge are partly overlapping constructs (compare Section 2.7), a certain degree of overlap has to be expected.

7.2 Laboratory Experiment

The following experiment aims at testing Hypothesis 3.9.5 that assumes a correlation between the AST-elicited weighted density measure and task performance in a performance task involving implicit knowledge. The experiment further aims at testing the hypotheses from Chapter 4 with regard to implicit and structural knowledge and complex problem solving performance and from Chapters 5 and 6 with regard to group knowledge distribution and group performance. It is supposed to test whether the experience with a complex scenario leads to more structural implicitness on scenario control, and whether the distribution of knowledge in a group influences the group's performance in a complex problem solving task. The experiment is a 3 (shared knowledge elements – partially shared knowledge elements – unshared knowledge elements) $\times$ 2 (practical experience during learning – no experience during learning) $\times$ 2 (practical experience during discussion – no experience during discussion) factorial design.

7.2.1 Sample

The study took place from April to August 2006 at the section of Organizational and Social Psychology at Humboldt University Berlin[2]. It was advertised as an assessment center simulation as it involved elements that are typical for assement centers such as an IQ test, a group discussion, and a complex management computer simulation. Participants were offered a detailed individual feedback on their performance and the top ten teams were offered a reward of 10-15 EUR per member. 150 persons, mostly students from different fields and universities in Berlin, participated, forming

[2]I am very grateful to the following students for their work as experimenters and for their support in processing the data: Janina Blume, Barbara Haller, Ludmila Herdt, Christina Karsten, Conny Lochelfeld, Goran Milacevic, Anja Rettig, Emanuel Schmider, Marion Schwenke, Nicole Steckloina, Sho Tsuji, Xenia Weißbecher-Klaus, and Thomas Wildermuth.

75 dyads. Individuals were distributed evenly over experimental conditions. A screening of the data revealed that the AST-data for six participants was unusable: Due to technical errors, the data of three participants was incompletely recorded, and three participants failed to comply with the instructions by entering random words. Three subjects showed extreme deviations in taylorshop performance (more than five standard distributions off the population mean). As measurement errors cannot be ruled out as a cause, these subjects are excluded from the sample. Four subjects failed to fully complete all questionnaires and the taylorshop data of two subjects was not recorded due to technical errors. This reduces the total effective sample size to 134 individuals. The distribution of participants over experimental conditions is presented in Table 7.1.

Table 7.1: Distribution of participants on individual level after case exclusion over experimental conditions

		Knowledge overlap in the dyad			
Learning	Discussion	Little	Partial	Full	Total
From texts	No experience	14	12	12	38
only	Experience	12	12	8	32
From texts &	No experience	10	10	8	28
experience	Experience	10	14	12	36
Total		46	48	40	134

The remaining effective sample consists of 73 female and 61 male participants. Average age of participants was 26.3 years, ($\sigma = 5.98$). The average individual score for reasoning (BIS-K) was 100.15 ($\sigma = 10.8$).

As the hypotheses on group level do not include assumptions with regard to AST scores, those subjects with missing or invalid AST scores are included in the aggregation of participants to dyads, leading to a slightly different sample of 70 dyads (140 individuals, 73 female and 63 male) The distribution of dyads over experimental conditions is presented in Table 7.2.

Of the resulting 70 dyads, 23 were same-gender female dyads, 20 same-gender male dyads, and 27 mixed-gender dyads. The average IQ for reasoning is the same as above.

7.2.2 Complex problem solving task

Complex problem solving performance is measured with Süß's A version of the TAYLORSHOP (SCHNEIDERWERKSTATT) microworld (Süß, 1996). The two main reasons for this choice were the availability of a validated

Table 7.2: Distribution of dyads after case exclusion over experimental conditions

		Knowledge overlap in the dyad			
Learning	Discussion	Little	Partial	Full	Total
From texts	No experience	6	6	7	19
only	Experience	7	6	4	17
From texts &	No experience	5	5	6	18
experience	Experience	5	7	6	18
Total		23	24	23	70

questionnaire on declarative knowledge on TAYLORSHOP scenario control, the WIS (Kersting & Süß, 1995) and its short form (in Klocke, 2004) and the finding that TAYLORSHOP scenario performance correlates with performances in other complex problem computer simulations (Endres & Putz-Osterloh, 1994; Wittmann et al., 1996). Furthermore, the employed version of the TAYLORSHOP scenario received the highest evaluation with regard to its suitability for psychometric assessment in Kluge's overview of available scenarios (Kluge, 2004, p. 71).

In the TAYLORSHOP scenario, participants play the owner of a small shirt factory and have to achieve as much profit as they can during a simulated period of 12 months. In each month, twelve variables can be altered by the participant: purchase of raw material for shirt production, market price for shirts, marketing budget, number of shops, number of traveling salesmen, number of small shirt-producing machines (producing 50 shirts per month), number of large shirt-producing machines (producing 100 shirts per month), the number of workers for each type of machine, machine maintenance budget, wages, and social welfare. The changes made to these variables influence the values of the observable variables account balance, overall profit, demand, shirt production, shirts on stock, machine damage, number of shirts produced, production loss, and workers' motivation. The original German wording of the control variables and observation variables can be taken from the image of the scenario's interface, depicted in Figure 7.1.

The changes of the monitor variables is based on a complex interconnection scheme including further invisible variables (Süß, 1996, p. 102). It is depicted in Figure A.1 in Appendix A.1. The system is dynamic (some changes of variable states occur regardless of the participants' actions) and the scheme of variable interrelationships is highly complex. The target of the game is to maximize the overall profit after the 12-month period. Thus, the variable overall profit serves as dependent variable that operationalizes task

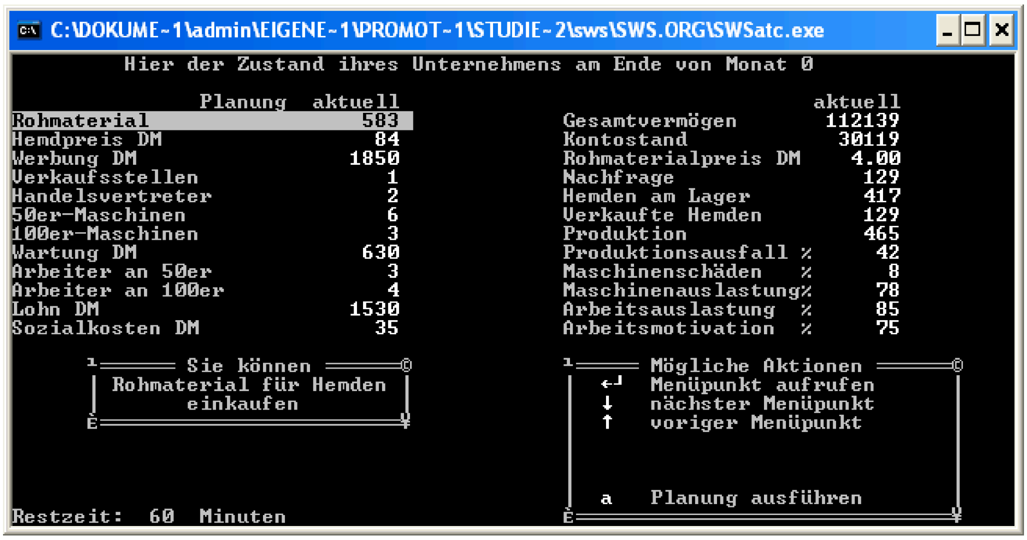


Figure 7.1: Original German interface of the TAYLORSHOP dynamic scenario (Süß, 1996) as used in the experiment. The control variables are presented in the left column, the variables that monitor the state of the factory on the right.

performance or effectivity.

7.2.3 Procedure

In order to operationalize Hypotheses 5.5.1 and 5.5.2 assuming a connection between group knowledge increase and performance, phases of knowledge exchange among group members and phases of task execution need to be separated. Since groups go through several phases during problem solving processes (Bales & Strodtbeck, 1951), this separation is not interfering with natural patterns of group processes: During orientation, information is gathered; during the evaluation phase, discussion (and thus knowledge sharing) and assessment of strategies take place whereas the actual performance of the task is carried out in the control stage. Thus, evaluation and control phases of task performance should be separated in order to test the hypotheses. This separation is based on the assumption that group processes are both beneficial to learning and to task performance, as shown in several studies (Hausmann, 2005, p. 2).

Participants were assigned to dyads in order to minimize social effects that occur in larger groups and due to the fact that dialogue is considered one of the most powerful forms of learning and knowledge sharing in problem solving (Hausmann, 2005).

At the beginning of the experiment, each participant individually worked on the reasoning scale of the short form of the Berlin Structural Intelligence Test (BIS-K, Jäger et al., 1997), as reasoning is known to correlate with TAYLORSHOP scenario performance (Süß, 1996; Wittmann et al., 1996).

Manipulation of knowledge overlap

After the BIS-Test, the learning sequence began. During learning, participants individually acquired knowledge on TAYLORSHOP scenario control from instructional texts developed by Klocke (2004). In our study, we assigned sections of his instructional text to labels G1, G2, A, B, C, D, E, F, G, H, I, and J, where G1 and G2 represent general, introductory information on the system and elements A...J encompass specific knowledge elements on how to perform successfully (compare Table 7.3). The originally employed German texts can be found in Section A.1 of the Appendix.

Each group member received items G1 and G2 and five Items from the set A...J, embedded into a running text for learning[3]. For the latter, the intra-group overlap between group members was experimentally varied: In the shared condition, both group members received the same five elements. In the partially shared condition, two of the five elements were assigned to both group members, and three elements were exclusive to each member. In the unshared condition, each group member received five knowledge elements that the other members did not receive. Note that cognitive load was about the same for the individual group members over all conditions (two general elements G1 and G2 plus five specific elements), but the number of knowledge elements present inside the dyad differed: Apart from the general elements G1 and G2, there were five elements in the shared condition, eight in the partially shared condition, and ten elements in the unshared condition. Learning took place individually and learning time was ten minutes. Participants were allowed to take notes at their discretion during learning.

Due to the connectionist structure of the existing instructional text edited by Klocke (2004), it was not possible to present each knowledge element independent of each other, as some text passages refer to each other. The following combinations of knowledge elements could be formed without rendering the text illegible: ABCIJ, DEFGH, BCDEF, and EFGHI. Dyads in the unshared condition received texts containing elements ABCIJ and DEFGH, dyads in the partially shared condition received texts containing BCDEF and DEFGH, and dyads in the shared condition received varying sets of two identical texts. The question whether these combinations are individually of

[3] An example text with knowledge elements A, B, C, I, and J in German can be found in Section A.3 of the Appendix.

Table 7.3: Knowledge elements for TAYLORSHOP scenario control that were embedded in instructional texts based on Klocke (2004).

Element	Title	Content
G1	Stable system	Small changes lead to small consequences; do not act too cautiously. The bank offers generous credit and a negative value on the bank account is not a problem. If in financial trouble, do not sell assets such as machinery and shops.
G2	Equilibrium	The demand should meet production, the number of machines should be equal to the number of workers, one should continuously increase sales, and expand business.
A	Raw material	The price of raw material on the market is independent of all other variables and is subject to market fluctuations. There should always be enough raw material in stock to meet production capacity.
B	Stocking	Stocking creates costs. Do not produce too much but enough to satisfy market demand.
C	Production and demand	The number of shirts in stock plus raw material should equal demand. If the demand is unequal to production, it is better to increase one variable instead of decreasing it.
D	Investment strategy	In order to pay off investments such as new machinery and new shops, they should be made at an early point in time.
E	Machine efficiency	Displays (in percent) usage of machinery capacity. If the value falls below 100%, machines may be damaged, may have too few operators, workers may not be motivated or too little raw material may be present.
F	Worker efficiency	Displays (in percent) usage of work capacity. If the value falls below 100%, workers may be unmotivated (depending on wages and welfare) or raw material may be short.
G	Investing in machinery	One should only buy machines that produce 100 shirts per month as these produce the same costs in maintenance as machines producing 50 shirts a month.
H	Damage to machinery	Spending on machinery maintenance prevents damages and should never be 0. If damages rise above 10%, maintenance should be increased. More machines require more maintenance and damage to machinery is independent of workers' motivation.
I	Demand	The marketing budget, the number of shops, and the number of traveling salesmen increase demand. The shirt price has a stronger influence on demand than marketing.
J	Expenses	Marketing budget, number of shops, the number of traveling salesmen, stocking and expenses per worker (wages and welfare) increase costs and reduce profit.

equal worth for performance is addressed in the result section. The text was two pages in length and included a picture of the TAYLORSHOP interface as well as the initial state of the scenario. This enabled the participants to discuss possible strategies for the alteration of the unfavorable initial state of the factory in the discussion part that followed at a later stage of the experiment.

Manipulation of scenario experience

In order to induce a dissociation between performance and verbalizable knowledge (which would indicate the presence of implicit knowledge, compare Section 4.5.3), half of the participants had a TAYLORSHOP simulation available during the learning period and were allowed to alter only those variables that were covered in their instructional texts. In this way, they acquired hands-on scenario experience through learning-by-doing and were expected to hold more implicit knowledge on scenario control compared to those participants who only learned from texts. After learning, participants performed a self-assessment: For each title of the ten specific knowledge elements, they were asked to assess their own level of expertise on a four-item scale. This self-assessment was then coded into a skillMap that contained a simple skill inventory of the ten knowledge elements. The structure of the graph is visible in the upper area in Figure 7.2.

Participants were also asked to rate their computer experience and the degree of their prior economic knowledge on a five-point Likert scale, as these features influenced problem solving capabilities in other studies (Süß, 1996). After self assessment, participants completed the short version of the WIS questionnaire on TAYLORSHOP scenario declarative knowledge (Klocke, 2004). They were not allowed to use their notes during the test. Participants were then asked to teach each other as much of their acquired knowledge as possible (knowledge exchange). Participants were told that their performance would be assessed on the basis of the mean of individual performances after the experiment, and were thus motivated to actually exchange their knowledge. Participants were allowed to bring their notes to the discussion and were allowed to make further notes. Time for discussion was fifteen minutes and the discussion was not structured in any further way. The instructions emphasized the aspect of knowledge exchange, which aimed at keeping normative influences to a minimum and to maximize informational processes. Such a setting can be seen as a viable operationalization of situations occurring in organizational practice: Two individuals freely discuss a complex problem in a fifteen-minutes meeting and take notes and proceed with individual work afterwards. For half of the dyads, a computer

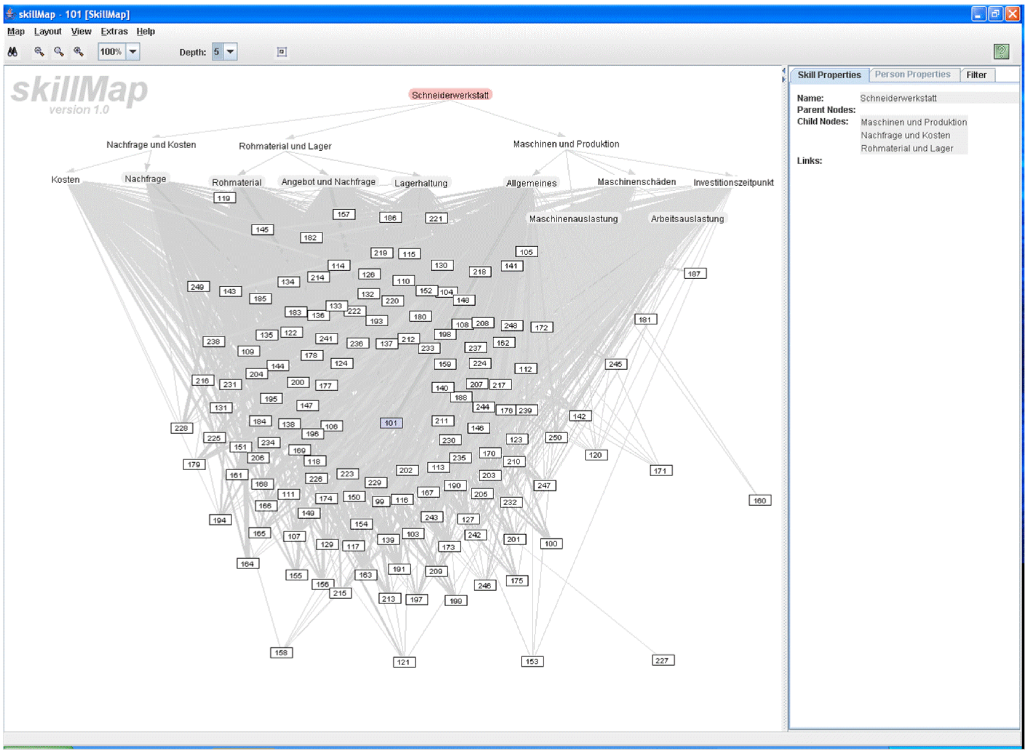


Figure 7.2: skillMap with the knowledge graph containing the labels of the ten knowledge elements provided. The boxes represent participants, edges between boxes and TAYLORSHOP labels represent the participants' self assessment.

running the TAYLORSHOP simulation was present during discussion for experimenting and trial and error. After discussion, the participants filled in the short version of the WIS questionnaire for a second time (again, the use of notes was not permitted). This second test allows the quantification of individual declarative knowledge increase caused by the discussion. These two post-discussion WIS results are averaged in order to indicate knowledge increase on group level. Afterwards, participants worked on the AST (compare Section 3.9) with the stimulus word "Taylorshop" and a maximum of 15 concepts for pairwise similarity ratings. Participants were not allowed to use their notes during AST-based structural knowledge assessment. Finally, participants worked on the TAYLORSHOP microworld individually but were allowed to use their notes. In this way, group members had access to knowledge from the group discussion process, but were able to perform the task on their own. Group performance was then calculated by averaging the two individual scores obtained for TAYLORSHOP profit. In this way, group processes from the knowledge exchange phase enter the task execution stage, but task execution itself does not interfere with interaction processes and possible process losses due to other interaction phenomena, as probably occurred in the study by Endres and Putz-Osterloh (1994). This approach is also referred to as the collective approach: "the collective approach targets the knowledge of individual team members and then aggregates this information" (Cooke, Salas, & Stout, 2000, p. 164). After the completion of the simulation, participants were thanked and debriefed. The entire experiment lasted approximately 2.5 hours. The resulting experimental design is a 3 (shared knowledge elements – partially shared knowledge elements – unshared knowledge elements) × 2 (practical experience during learning – no experience during learning) × 2 (practical experience during discussion – no experience during discussion) factorial design (compare Table 7.1). They layout of the laboratory setting is depicted in Figure 7.3.

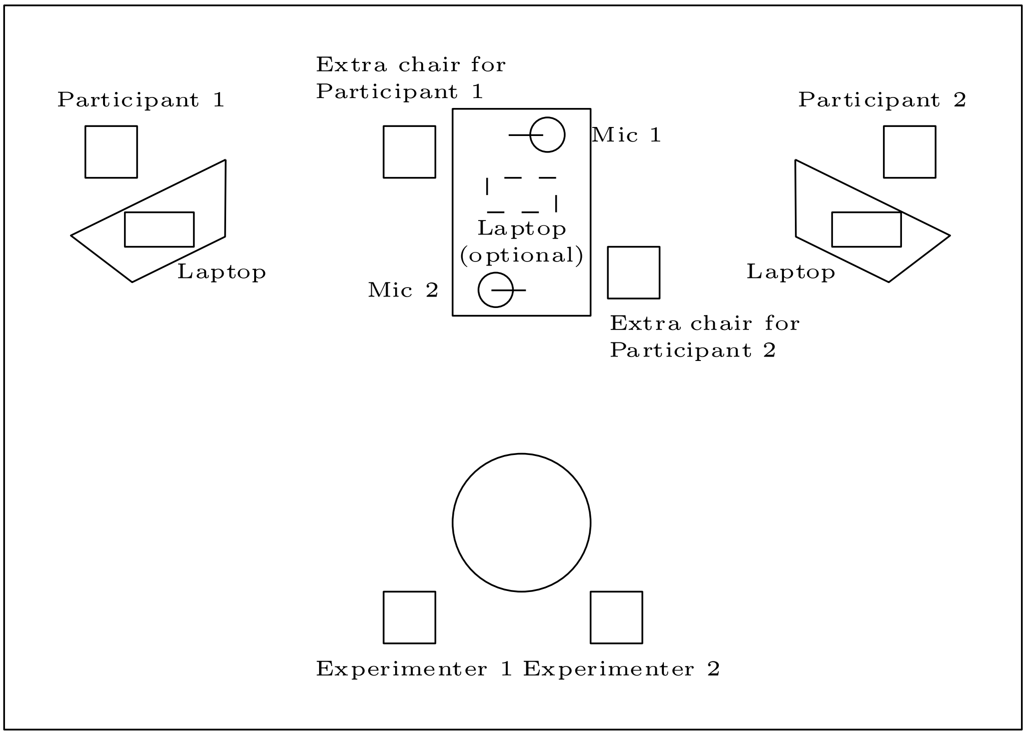


Figure 7.3: Spatial layout of the experimental setting

Chapter 8

Results

The result section is divided in three parts. In the first one, results from the AST validation experiments are reported. It is followed by the second section on results of the laboratory experiment on the individual level, where findings with reference to the connection between structural knowledge, implicit knowledge and complex system control are reported. Results of the laboratory experiment on the group level are reported in the third section, which is mainly on the connection between knowledge overlap inside the dyad and the dyad's complex problem solving (CPS) performance.

8.1 AST Validation

In order to test the hypotheses on AST validity that have been put forward in the third section, four separate experiments were conducted that each tested one of the Hypotheses 3.9.1 through 3.9.4. They are presented in the following. Results relating to Hypothesis 3.9.5 are presented in Sections 8.2.4, 8.2.5, and 8.2.6.

8.1.1 AST internal consistency

As outlined in Section 3.9.3, the six AST coefficients number of associated concepts, number of links, weighted density, diameter, and the relative maximum path length are either mathematically deductible from each other or will correlate due to their underlying logical structure (compare Table 3.3). This subsection will confirm this pattern by reporting empirical correlations among the above-mentioned coefficients. This will allow an assessment of the AST's internal consistency: As all coefficients will correlate, the question is whether they must all be counted to one underlying factor or whether the subtle differences in actual correlations justify the existence of more than one underlying factor.

As this analysis requires a dataset of the six AST coefficients of an appropriate size, no separate experiment is required. Instead, individual AST scores from all available sources will be assembled into one large AST sample for analysis. Factor analysis procedures require a large sample size: Guadagnoli and Velicier (1988) found that for stable parameter estimates, the absolute sample size is more important than functions of sample size. They suggested a minimum sample size of 100 to 200 cases. In order to obtain such a sample, the data is aggregated from Experiments 7.1.1 through 7.1.4 and from Experiment 7.2 (compare next sections). Only participants that worked on the AST under the standard condition (quick and intuitive placement of edges without the necessity to label them) are included. In anticipation of the sample characteristics of the following sections, the aggregated sample contains 193 cases (107 male, 86 female, average age = 24.59, $\sigma = 3.29$). The correlations of AST-coefficients are presented in Table 8.1.

Table 8.1: Spearman rho rank order correlations between AST coefficients ($N = 193$). * indicates significance on 5%-level, ** on 1%-level, and *** on 0.1%-level.

	1	2	3	4	5	6
1. Concepts	–	$.64^{***}$	$.77^{***}$	$-.72^{***}$	$.54^{***}$	$-.90^{***}$
2. Edges		–	$.49^{***}$	$-.24^{***}$	$.39^{***}$	$-.63^{***}$
3. Clusters			–	$-.54^{***}$	$.42^{***}$	$-.70^{***}$
4. Weighted density				–	$-.51^{***}$	$.58^{***}$
5. Diameter					–	$-.17^{***}$
6. RMPL						–

All but one coefficients correlate according to the predictions of Table 3.3. The diameter and the RMPL exhibit a small correlation in the opposite direction of the relationship illustrated in Figure 3.8. This is most probably due to the small range of actually occurring diameter values: in the given sample, the diameter ranges between 1 and 5 (the latter occurred only twice) and very small diameters coincide with small RMPL valus as visible in Figure 3.8. Overall, the empiric data mirror the dependencies established in Section 3.9.3.

After correlation analysis, the six AST coefficients are subjected to a factor analysis with oblimin rotation ($\delta = 0°$). The Eigenvalue > 1 criterion results in a two-factor solution, which accounts for 61.4% of the overall variance. The first factor accounts for 42.4%, the second for 19%. The pattern matrix of the rotated solution is presented in Table 8.2.

Table 8.2: Pattern matrix of the rotated factor solution for the aggregated sample of 193 AST participants (oblimin rotation)

	Factor 1	Factor 2
# of nodes	.90	
# of edges	.72	.28
Clusters	.76	
Weighted density		.98
Diameter	.32	-.28
RMPL	-.73	

Note. Loadings $< |.20|$ are omitted.

Three things become evident from the rotated factor solution. Firstly, the four AST coefficients number of nodes (graph degree), number of edges (graph size), number of clusters, and RMPL load on one factor, presumably because of their close tie to the number of nodes (compare Section 3.9.3). Note that for RMPL, smaller values are favorable, which explains the negative sign. The fact that the number of nodes displays the highest loading on the first factor underlines the fact that these four coefficients are predominantly bound to the number of nodes.

Secondly, the subtle differences among AST coefficients are mirrored in the observation that the weighted density does not load in the first factor, despite its logical ties to the number of edges and nodes (compare Figure 3.7). This might be due to the fact that the weighted density measure is the only AST coefficient that includes the edge weights in its computation (compare Section 3.9.3). Since the weighted density is the only coefficient loading substantially on the second factor and is thought to be the closest indicator of structural implicitness (compare Section 3.9.3), the second factor could be labeled as such. The close logical link between the weighted density measure and the number of nodes is mirrored by the inter-factor correlation of $r = .41$ ($p < 0.001$).

Thirdly, the diameter does not load substantially on the two factors. This indicates that it it measures something else than the weighted density and those coefficients exhibiting a tight connection to the number of nodes (Factor 1). This is insofar interesting as the diameter is significantly correlated with all other coefficients (compare Table 8.1) and has been attributed with carrying structural information by others (compare Bonato, 1990, in Section 3.9.3).

Taken together, the analysis of the internal consistency shows that all AST coefficients depend heavily on the number of nodes and that their correlations mirror their logical dependency. However, this dependency does not justify a combination of all coefficients into one measure, as factor analysis indicates that two coefficients, the weighted density and the diameter, do not load on the same factor than the other four coefficients. The coefficients are therefore kept separate for further analysis.

8.1.2 Labeled and unlabeled AST testing

The fundamental claim of the AST is its assumed capability to capture structural implicitness, i.e. to capture links between explicit concepts that lie in the implicit. Hypothesis 3.9.1 thus assumes that those participants who are asked to make quick and intuitive pairwise concept comparisons will place more edges than those participants who are asked to place only edges that they can explicitly label. This would also lead to denser networks. As word associations are made before any concept relations are placed in the AST system, the number of associated concepts (the graph degree) is not supposed to exhibit any differences between the two conditions for concept placement (standard vs. explicit). The same applies to the number of clusters, which depend heavily on the number of associations, but have no connection with the placement of edges.

In order to test Hypothesis 3.9.1, 30 students at the Institute of Psychology at Humboldt-University Berlin (19 female, 11 male, average age = 23.59, $SD = 2.23$) participated in the first experiment (compare Section 7.1.1). Kolmogorov-Smirnov tests for normal distribution reveal normal distributions for the number of associated concepts ($K-SZ = 1.145, p = .14$), the number of edges ($K-SZ = 1.130, p = .15$), the number of clusters ($K-SZ = 1.309, p = .065$), the weighted density of the graph ($K-SZ = 0.883, p = .42$), and relative maximum path length scores (RMPL, measure for graph compactness, smaller values indicate a more compact organization of the AST graph, $K-SZ = 0.953, p = .32$). The diameter is not normally distributed ($K-SZ = 1.708, p = .006$). The hypothesis is thus tested with independent-sample t-tests for the normally distributed variables and with the Mann-Whitney Test for the diameter variable. The results are presented in Table 8.3.

The networks elicited under the standard (quick, intuitive, and unlabeled) condition have a significantly higher number of edges than those elicited under the explicit condition. On average, they have more than twice as many edges and are also significantly more compact as visible in the significant difference in RMPL-scores. Neither the number of clusters, nor the number

Table 8.3: Comparison of means with independent sample t-tests and Mann-Whitney's U-Test comparing graph-theoretic coefficients for the standard AST mode with the explicit mode (see text).

	Test mode							
	Standard ($n = 13$)		Explicit ($n = 13$)					
	$\bar{x}$	SD	$\bar{x}$	SD	df	T / Z	$p(t/z)$	d_u[a]
# of concepts	13.40	6.0	10.87	5.67	24	1.189	.244	.43
# of edges	53.47	45.35	25.53	15.73	24	2.254	.032	.82
Cluster	5.27	2.789	4.80	3.10	24	0.434	.668	.15
Diameter [b]	2	n/a[b]	2	n/a[b]	n/a[b]	-0.046	.967	n/a
Weighted Density	1.13	0.45	0.98	0.61	24	0.755	.457	.28
RMPL	0.24	0.07	0.32	0.11	24	-2.538	.017	.87

[a] Effect size Cohen's d with bias correction for small samples.

[b] Due to the non-normal distribution of the diameter, Mann-Whitney's U-Test was used for the comparison of means. Instead of the mean, the median is reported. Measures of dispersion are not available (n/a) for the median.

of associated concepts significantly differs between the groups. Hypothesis 3.9.1 is thus supported with regard to the number of edges and the RMPL. Networks elicited in the standard condition are also denser than networks elicited in the explicit condition, but this difference does not reach statistical significance. Networks elicited in the standard mode do not display different diameter values than networks elicited in the explicit mode.

Taken together, these results support the fundamental assumption that the AST is able to capture at least a fraction of structural implicitness by prompting test participants for quick and intuitive similarity ratings on previously associated concepts. If participants are asked to place only those connections they can explain, they place considerably fewer connections. Thus, edges elicited under the standard intuitive connection encompass the labels that a participant cannot label (structural implicitness) and the labels that can be labeled.

8.1.3 AST sensitivity to learning

Hypothesis 3.9.2 assumes an increase or a change towards the more favorable direction of AST-coefficients on a given knowledge domain if learning occurs in that particular domain. In order to test this assumption, participants of a seminar were tested on the seminars' topic (knowledge management) before and after the completion of the course.

Of the 32 seminar participants that participated in the experiment, 19 participated in both measurements: Eight females and eleven males (average age= $24.01, SD = 3.24$). Of the six measurement variables that were elicited at both points in time, only two variables deviate from the assumption of normal distribution in the pretest: The relative maximum path length (RMPL) and the diameter (compare Table B.1 in Appendix B). All variables in the post test display a standard normal distribution (compare Table B.2 in Appendix B). The comparison between pre- and posttest values for the two variables RMPL and diameter is thus performed with the non-parametric Wilcoxon signed ranks test, the other variables are compared with paired-sample t-tests (compare Table 8.4)

Table 8.4: Paired sample t-test comparing graph-theoretic coefficients for pre-learning and post-learning AST-elicited structural knowledge parameters. * indicates significance on 5%-level (two-tailed), ** on 1%-level (two-tailed).

	Test							
	Pretest ($n = 19$)		Posttest ($n = 19$)					
	$\bar{x}$	SD	$\bar{x}$	SD	df	T / Z	$p(t/z)$	d_u [a]
# of concepts	9.67	4.39	15.90	7.59	20	-4.148	$< .001^{**}$	.98
# of edges	27	23.42	54.89	35.44	17	-3.675	$.002^{**}$	.82
Cluster	4.19	1.83	6.57	2.98	20	-4.200	$< .001^{**}$	.94
Diameter [b]	2.43	n/a	2.86	1.06	n/a	-1.369	$.085^{+}$	.41
Weighted Density	1.14	.71	1.01	.57	15	.980	.171	-.11
RMPL [b]	.25	n/a	.21	n/a	n/a	-2.475	$.013^{*}$	-.65

[a] Effect size Cohen's d with bias correction for small samples.
[b] Due to the non-normal distribution of this variable, the Wilcoxon signed ranks test was used for the comparison of means. Instead of the mean, the median is reported.

Of the six measurement variables, four display a significant increase from the first measurement to the second. This indicates that learning is reflected by changes in AST these four coefficients over time. The diameter's increase reaches marginal significance, while the density slightly decreases from the first to the second measurement. Thus, Hypothesis 3.9.2 is supported by four of the six measurement variables. The diameter has been proposed as a measure for structural knowledge by Eckert (1998a). However, the diameter encompasses something different. In order to obtain a large diameter, two conditions have to be met. Firstly, a certain number of nodes has to be present for the diameter to span a longer distance. Secondly, the nodes have to be somewhat scarcely connected, bacuase a highly dense, star-shaped network will most likely deliver a shorter density. Thus, a large density requires the activation of a chain of word associations with start and end points that lie far away from each other in their respective network. A large diameter requires a lot of verbal associations that span a wide range of subject areas. It thus can be seen as a qualitative measure of the span of declarative word knowledge and not as a measure of structural knowledge. The marginally significant increase of the diameter during the course of a seminar would also support this notion.

The weighted density has been suggested to indicate structure implicitness (compare Section 3.9.3). Therefore, the lack of a significant change of the measure from pretest to posttest could be caused by a lack of structural knowledge acquisition during the course of the seminar. The slight visible decrease of the weighted density measure from the first to the second measurement is also due to the internal logic of the AST (compare Section 3.9.3): As the weighted density correlates negatively with the number of concepts and the number of edges, a significant increase of those two variables has to be accompanied by a decrease of the weighted density measure.

8.1.4 Convergent validity

Hypothesis 3.9.3 assumes that the number of associated concepts, the number of links, the density, the diameter, and the relative maximum path lengths of SLT- elicited knowledge graphs and AST-elicited knowledge graphs correlate.

Only eighteen students participated in the experiment (compare Section 7.1.3), fourteen female and four males. Severe technical problems with the data base system led to a usable subsample of only eleven cases (four males, seven females, seven in the explicit condition, four in the standard condition), which renders the explanatory power of this particular study highly questionable. Although some argue that a correlation with a little as five cases might be of use (Kareev, 2000), this is heavily debated (Juslin & Olsson, 2005) and from a stistical point of view, the smallest sample size that has a possibility of delivering an interpretable correlation coefficient is eleven, employing Kendall's tau (Bonett & Wright, 2000). The sizes of the subsamples do not meet this criterion. They are thus combined into one sample and are correlated with SLT scores on aggregate level.

Average age of the participants was 26.21 ($SD = 7.567$) years. All measurement variables except for the diameter exhibit a standard normal distribution in the Kolmogorov-Smirnov Test (compare Table B.3 of the Appendix). Due to the small sample size, Kendall's tau rank order correlations between aggregated AST scores and SLT scores are calculated with R (R Development Core Team, 2007) and are presented in Table 8.5.

Table 8.5: Kendall's tau correlations between AST-elicited coefficients and SLT-elicited coefficients of knowledge networks on the stimulus "nervous system". * indicates significance on 1%-level.

	# of edges	W. Density	Diameter	RMPL
$r_{AST_{(std+explicit)},SLT}$ ($N = 11$)	.07	.53	.63*	.26

On the aggregated level, a significant correlation between the AST-elicited diameter and the SLT-elicited diameter is visible. The weighted density measures and RMPL scores also display a positive correlation, but both do not reach levels of statistical significance, which is probably caused by the extremely small sample size. SLT-elicited edges exhibit a close-to-0 correlation with AST-elicited edges. That is probably due to the fact that the aggregated AST-sample contains edges that the participants would not label, while the SLT sample contains only labeled edges. All in all, hypothesis 3.9.3 is refuted by these results; SLT and AST only correlate on only one of the four coefficients. This may however be caused by the extremely small sample size.

8.1.5 AST congruent validity and discriminant validity

In order to test whether AST scores correlate with outside criteria of knowledge based performance, AST-scores are compared to exam grades. Hypothesis 3.9.4 postulates a positive correlation between the two. The discriminant validity of the AST is tested by comparing AST scores to the results of an IQ-test (compare Section 7.1.4).

A total of 52 students from three classes (two cleaners, one plumbing) participated in the experiment (49 male, three female). Average age was 20.31 ($SD = 3.48$). With regard to their education prior before training, six students came from a special school, 26 have the German Hauptschulabschluss, (an equivalent of a CSE certificate that qualifies below O-Levels), 15 hold a German secondary modern school certificate (Realschulabschluss), and two have Abitur, the diploma from German secondary school qualifying for university admission. Average WST-IQ score of the sample is 87,08 ($SD = 8,86$). The average total grade of the reference examination was

4,39 ($SD = 1.08$) on the standard German school grade scale ranging from 1 (excellent) to 6 (unsatisfactory).

According to the Kolmogorov-Smirnov test for normal distribution, five measurement variables fail to display a standard distribution: Grade ($K - SZ = 1.423, p = .035$), number of clusters ($K - SZ = 1.549, p = .016$), number of edges ($K - SZ = 1.378, p = .045$), PFNET diameter ($K - SZ = 1.657, p = .008$), and RMPL ($K - SZ = 1.420, p = .034$). Only the number of associated concepts ($K - SZ = 1.209, p = .109$), the weighted density of the PFNET ($K - SZ = 0.908, p = .382$), and the average WST IQ ($K - SZ = 0.871, p = .434$) display a normal distribution. Thus, spearman rank correlations are employed for correlating the AST-elicited measures with IQ and grade scores (compare Table 8.6).

Table 8.6: Spearman rho rank order correlations between AST scores, exam grades and WST-elicited IQ scores ($N = 52$). $^+$ indicates marginal significance on 10% level (one-tailed), * indicates significance on 5%-level (one-tailed), ** on 1%-level (one-tailed).

	WST	Concepts	Edges	Clusters	Weighted Density	Diameter	RMPL[b]
Grade[a]	.41**	.22^{+}	.31*	.25^{+}	-.36*	.18	-.41**
WST-IQ	1	.34*	.13	.37**	-.07	.13	-.15

[a] The employed standard German grades range from 1 – 6; 1 is best, 6 is worst. In order to simplify the interpretability of the correlations, this grade scale is reversed, i.e. larger values are desirable and positive correlations with the grade indicate higher achievement.

[b] Note that smaller RMPL values are desirable as they indicate a more compact organization of knowledge.

Of the six AST-coefficients, two are correlated with the exam grade in the expected direction at marginally significant levels: the number of concepts and the number of clusters. Two coefficients exhibit a significant correlation with exam grade in the expected direction: The number of edges and the RMPL. These four coefficients thus support Hypothesis 3.9.4. The diameter also correlates positively with exam grade, but fails to reach statistical significance. The weighted density displays a significantly negative correlation with exam grade. Although this is contrary to the hypotheses, it is in line with the logical structure of the AST (compare Sections 3.9.3 and 8.1.1): the weighted density correlates negatively with the graph order (number of concepts) and the graph size (number of edges). As those two coefficients correlate positively with exam grades, the weighted density *has* to exhibit a negative correlation with exam grades.

Taken as a whole, these results are more supportive of Hypothesis 3.9.4 than opposing as it is supported by four of the six coefficients.

Apart from this observation, it becomes evident from Table 8.6 that in this case, the construct measured by the AST and the construct measured by the WST overlap to a significant extent. In order to evaluate whether the AST measures something else than the WST, a partial correlation of the AST-scores with grade is performed, controlling for WST. As partial correlations are based on Pearson's correlation coefficient, the given data do not allow this analysis in the strict sense. It is however performed in order to establish a trend; the results will be interpreted with the appropriate caution. As visible in Table 8.7, no significant correlations between AST scores and grade remain if the influence of the WST IQ variable is kept constant, although the weighted density measure and RMPL score might achieve significant values if a larger sample size was employed. These results indicate a low discriminant validity of the AST if compared to the construct of verbal intelligence as measured by the WST. Due to the AST's heavy dependency on verbal capabilities for term associations, this is not surprising and the question remains whether operationalizations of IQ with lesser dependence on verbal skills will also correlate to these extends. This question will be addressed in Table 8.13 in Section 8.2.6.

Table 8.7: Partial correlation between AST-elicited variableas and exam grade controlled for WST-IQ ($N = 52$)

	Concepts	Edges	Clusters	Weighted Density	Diameter	RMPL
r	-.00	-.04	-.14	.24	.12	.23
p (one-tailed)	.497	.437	.264	.141	.294	.147

8.2 Laboratory Experiment Results on Individual Level

Prior to hypotheses testing, the data is screened for normal distributions. Furthermore, it is tested whether the manipulation of knowledge overlap was also subjectively experienced by the participants. Afterwards, the hypotheses are tested by means of analysis of variance (ANOVA), correlations, and error-bar charts. The section concludes with further findings not originally included in the hypotheses.

8.2.1 Operationalization of scenario performance

As a first step, a variable operationalizing TAYLORSHOP CPS performance needs to be selected. Since the participants were told to achieve as much profit as possible, the variable profit has the highest face validity and the use of a different operatinalization of profit should only be used if severe reasons prevent the usage of profit. However, past researches encountered difficulties with this variable. Funke (J. Funke, 1983) found skewed distributions of this variable and argued that the TAYLORSHOP scenario was too difficult for the majority of his participants. He also found only small correlations between intelligence scores and profit and thus suggested a new operationalization of performance, the trend measure. The trend measure denotes the number of simulated months during which a participant is able to achieve a profit in comparison to the previous month. Süß (1996, p. 144) found that in the original version of the TAYLORSHOP as employed by Funke, the number of sold shirts negatively correlated with overall profit. He simplified the system and obtained satisfactory difficulty measures for both the profit and trend measure. However, he also suggested a different measure, the sum of the Z-transformed scores of shirt sales per month and profit margin per shirt.

A one-sample Kolmogorov-Smirnov Test reveals that the assumption of bivariate normal distribution does not hold for the four measures profit ($K - SZ = 1.412, p = .037$), trend ($K - SZ = 1.534, p = .018$), average profit per shirt ($K - SZ = 3.335, p < .001$), and average shirt sales per month ($K - SZ = 2.190, p < .001$). Thus, the combined measure of average shirt sales and average profit per shirt cannot be used as suggested by Süß, because the conditions for the z-transformation are not met. An examination of the remaining two performance indicators reveals the following: The profit measure resembles a normal distribution more closely than the trend does (compare Figures 8.1 and 8.2), but was the more difficult goal to achieve: The average profit is -36780.30 DM and the modal interval (43 participants) represents a loss between 0 and 50000 DM. In the trend measure, 107 participants managed to achieve an increase in profit compared to the last month in at least one month. However, the distribution of the trend variable is more skewed than the profit variable. In summary, an overall profit was more difficult to obtain for the participants (43 of 134 achieved a profit larger than 0) than a trend larger than 0, but the data quality of the profit measure is better than the trend measure due to its closer resemblance of a normal distribution. A non-parametric Spearman-rho correlation of the profit and trend measure with the established determinants of TAYLORSHOP performance, declarative knowledge (WIS$_{post}$, BIS-K, computer experience (CE) and prior economic knowledge (EK)) displays significant positive correlations indicating validity of both measures (compare Table 8.8). Since the two measures correlate at .90, they seem to capture the same ability and are thus both kept as operationalization of TAYLORSHOP performance. As the trend measure has proven to be a more robust against participants' control errors (J. Funke, 1983) and due to the finding that the trend and the performance have displayed different correlations with outside criteria (ibid), the two measures are kept separate.

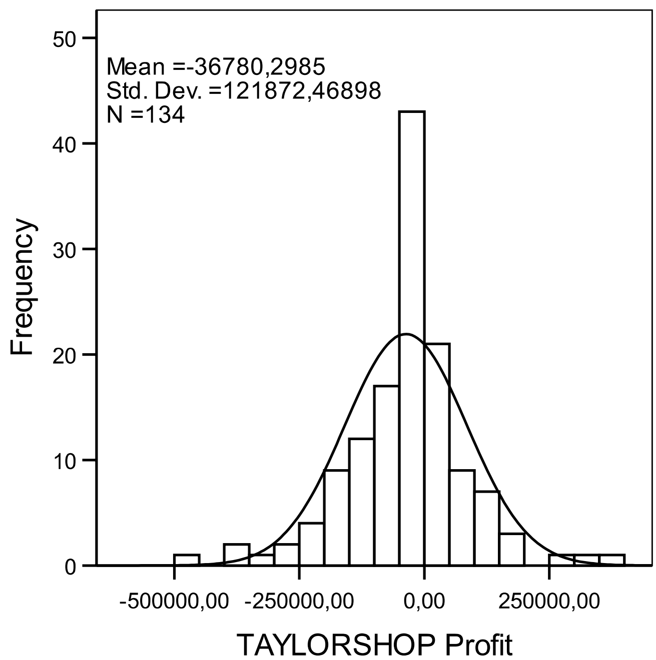


Figure 8.1: Histogram with normal curve of the variable profit on individual level

8.2.2 Comparability of individually learned knowledge elements

In order to meet the demand that all participants have a quantitatively comparable amount of scenario knowledge after learning the five TAYLORSHOP-related elements and to make sure that no combination of knowledge elements was superior to others, it is tested whether all four combinations of knowledge elements (ABCIJ, DEFGH, BCDEF, and EFGHI, compare Table 7.3)

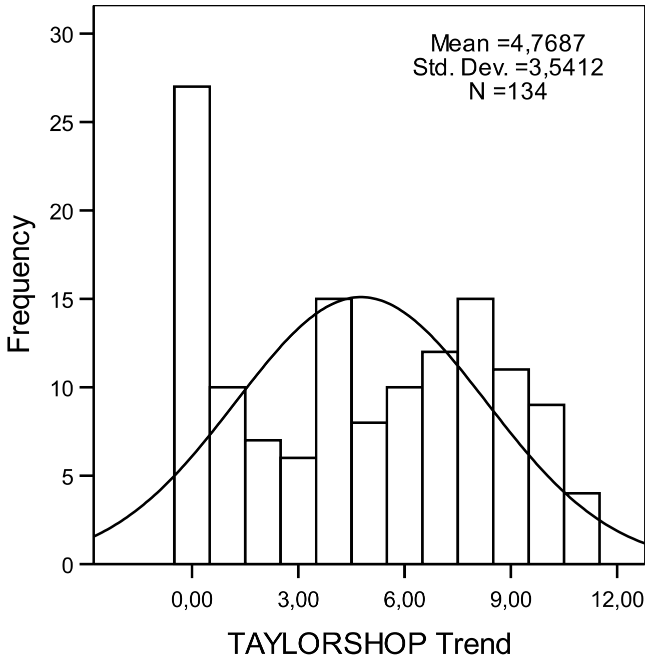


Figure 8.2: Histogram with normal curve of the variable trend on individual level

Table 8.8: Spearman rho correlations between profit, trend, and established predictors of TAYLORSHOP performance on individual level ($N = 134$). $^+$ indicates marginal significance on 10% level (one-tailed), * indicates significance on 5%-level (one-tailed), ** on 1%-level (one-tailed).

	1	2	3	4	5	6
1. Profit	–	$.91^{**}$	.00	$.21^{**}$	$.32^{**}$	$.40^{**}$
2. Trend		–	.10	$.19^{*}$	$.34^{**}$	$.44^{**}$
3. Computer experience			–	$.36^{**}$	$.13^{+}$	$.16^{*}$
4. Economic knowledge				–	$.13^{+}$	.04
5. BIS-K					–	$.50^{**}$
6. WIS_{post}						–

led to comparable pre-discussion WIS-Test results (WIS_{pre}) in the declarative knowledge test. A one-factorial ANOVA of WIS_{pre} over the four combinations of knowledge elements reveals no significant effect of the learned knowledge element on individual pre-discussion declarative knowledge score ($F(3, 133) = 0.339, p = 0.80$). Thus, on individual level, all participants enter the experiment with sets of declarative scenario knowledge that lead to a comparable amount of declarative knowledge across learning variations. The question whether different combinations of knowledge elements lead to superior implicit scenario knowledge on individual level cannot be tested, as no pre-discussion measure of implicit knowledge exists. But it is not very plausible to assume differences in implicit scenario knowledge because no participant had any prior experience with the scenario.

8.2.3 Analysis of intra-group dependencies of variables

Hypotheses 3.9.1 through 3.9.5 make assumptions on the individual level, but the data has not been surveyed on a purely individual level. Experiment participants are embedded in dyads (compare methodology in Section 7.2), which has to be considered in the analyses. In a dyadic setting, an observed correlation "can reflect relations between the variables at an individual level, at a dyadic level, or both.... The observed 'overall' correlation for individuals interacting in dyads can be seen as a mixture of both the individual-level and the dyad-level relation" (Griffin & Gonzales, 1995, p. 430). It thus has to be determined whether the observed correlations on the individual level partly stem from the relations on the dyadic level. Therefore, it is analyzed whether the variables are independent, i.e. whether the membership to a certain dyad influences the expression of measurement variables.

There are several methods available for this analysis, depending on whether the two members of a dyad are interchangeable or not (Kashy & Kenny, 2000). In the design of the laboratory experiment, different knowledge elements were assigned to the members of the dyad. Group members are also distinguishable in terms of gender and age. However, such a distinction should only be made if it affects responses (Kenny, Kashy, & Cook, 2006). As Section 8.2.7 will establish, male participants usually score higher than female participants. Thus, in order to be on the safe side, the distinguishable case is assumed. The measure for nonindependance for distinguishable dyads is the Pearson product-moment correlation between the two members of a dyad. Kenny, Kashy and Cook (2006) strongly advise to control this correlation for independent variables, because "failing to control for these variables may lead us to mistakenly conclude that there is nonindependence when, in fact, there is none" (Kenny et al., 2006, p. 30). For the dependent variables TAYLORSHOP profit and trend, the intragroup Pearson correlation between the dyads's members is thus controlled for reasoning (BIS), scenario-specific knowledge (WIS_{pre}, WIS_{post}, and WIS_{Δ}), prior economic knowledge, AST weighted density, the gender composition of the dyad (compare Section 8.3.8), and a variable representing knowledge overlap inside the dyad. The resulting intragroup correlation is $r = .32$ ($p = .020$) for profit, and $r = .13$ ($p = .345$) for the trend measure. The intragroup Pearson correlations for the independent variables are presented in Table 8.9.

Table 8.9: Assessment of dyadic nonindependence of independent measurement variables: Pearson product-moment correlation. Correlations are adjusted by partialling out known predictors where appropriate. * indicates significance on 5% level.

Variable		r
	Economic knowledge	.09
	Computer experience	.07
	BIS-K	.14
	WIS_{pre}	.16
	WIS_{post}	.01[a]
	WIS_{Δ}	-.12
AST Coefficients	Associated concepts	.37*
	Edges (graph size)	.08
	Cluster	.09
	Diameter	-.06
	Weighted density	.03
	RMPL	.14[b]

[a] controlled for BIS-K, WIS_{pre} and WIS_{post}.
[b] controlled for associated concepts.
* indicates significance on 5% level (two-tailed).

The discovered nonindependence of the dependent variable profit and the independent variable associated concepts is not surprising, as these were most probably influenced by the dyad's interaction during the discussion. Kenny, Kashy, and Cook (2006) advise an analysis on group level through group mean calculation if nonindependence among dependent variables is identified. This will be conducted in Section 8.3. In order to maintain an analysis on individual level, the individual correlations must be treated carefully, as they are influenced by two components: By a portion of variance shared between the dyadic partners and an "individual component representing the portion of the variable that is unshared or unique between dyadic partners" (Griffin & Gonzales, 1995, p. 433).

This can be achieved with a covariance theorem decomposition of bivariate two-level correlation introduced by Robinson (1950). This procedure has also been referred to as Within-And-Between-Analysis (WABA) by Danseraeu, Alutto and Yammarino (1984), who discussed this approach extensively.

WABA delivers a group-size weighted correlation of group means and, more importantly for this case, a group-mean-centered within-group correlation between two variables. This within-group correlation is adjusted for group effects and represents the correlation at individual level. For further discussion of the WABA approach cf. Bliese (Bliese & Halverson, 1998; Bliese, 2000). Table 8.10 reports the group-mean-centered within-group correlations between the dependent variables and independent variables profit and trend and the uncorrected individual correlations for comparison. WABA correlations were obtained with the multilevel package (Bliese, 2006) of the R software (R Development Core Team, 2007).

Table 8.10: WABA group-mean-centered within-group correlations and non-adjusted individual correlations of dependent and independent variables. The uncorrected correlations are Spearman rank-order correlations. $^{+}$ denotes marginal significance on 10%-level, * denotes significance on 5%-level, ** denotes significance on 1%-level, and *** on 0.1%-level.

		WABA		Uncorrected	
	Variable	Profit	Trend	Profit	Trend
	Economic knowledge	$.19^{*}$	$.17^{+}$	$.21^{*}$	$.19^{*}$
	Computer experience	.09	.06	.10	.10
	BIS-K	$.28^{*}$	$.28^{*}$	31^{***}	$.34^{***}$
	WIS_{pre}	$.36^{**}$	$.33^{**}$	$.25^{**}$	$.30^{***}$
	WIS_{post}	$.45^{**}$	$.40^{**}$	$.40^{***}$	$.44^{***}$
	WIS_{Δ}	-.16	-.15	-.11	-.12
	Associated concepts	-.11	-.04	-.08	-.08
	Edges (graph size)	-.09	-.07	.02	.02
AST	Cluster	-.10	-.02	-.08	-.08
Coefficients	Diameter	-.10	-.07	.02	.01
	Weighted density	$.18^{*}$	.08	$.18^{*}$	$.16^{+}$
	RMPL	.06	-.08	.05	.06

A comparison with the uncorrected values reveals one substantial difference among the significant correlations: The pre-discussion knowledge score WIS_{pre} correlates .11 weaker with the TAYLORSHOP profit measure if uncontrolled for group effects. All other significant correlations differ no more than $|.03|$ between corrected and uncorrected (except for WIS_{post}, where the difference is .05), and with two exceptions, the uncontrolled correlation is the smaller correlation.

Another visible difference between corrected and non-corrected correlations is the lack of a significant mean-centered corrlation between the weighted density measure and the trend measure. This is insofar negligible as the weighted density's effect on the trend measure will not be analyzed further, because the dissociation between verbalizable knowledge and performance will only become visible for the profit measure in Section 8.2.4. In summary, the use of the uncorrected values on individual level cam be justified for further analysis in this section due to their minor deviations from the corrected values, especially since an analysis on average dyadic level will be performed in Section 8.3.

8.2.4 Implicit knowledge and TAYLORSHOP performance

As stated in Hypothesis 4.6.1, it is assumed that exposure to the TAYLORSHOP scenario during learning leads to more scenario-relevant implicit knowledge compared to those participants who only learned from texts. In order to test this assumption, an analysis of variance with the two fixed factors learning mode and discussion mode would be desirable. Unfortunately, the data level does not permit this analysis, as the dependent variables profit and trend are not normally distributed. Differences in those two variables over the two factors are thus determined with two separate non-parametric Mann-Whitney U-Tests. The first one is performed comparing the central tendencies of TAYLORSHOP profit and trend between the two modes of individual scenario knowledge acquisition: learning from texts only versus learning from texts and from experimenting with the system presented in the text. The Mann-Whitney U test exhibits a marginally significant difference in performance scores for profit ($Z = -1.298, p = .10$, one-tailed), but not for trend ($Z = -0.944, p = .17$, one-tailed). Participants who had a scenario available during learning achieved an average profit of -22846 DM while those who only learned from texts made an average profit of -49519 DM. Experience learners achieved a profit with comparison to the previous month (trend) in 5.07 months on average, compared to 4.48 months for those who learned from texts only.

In order to test whether this slight superiority of experience during learning is based on implicit knowledge acquisition during learning, pre- and post-discussion declarative knowledge WIS-Test scores are compared between experience learners and text-only learners. WIS_{pre}, WIS_{post} and WIS_{Δ} display normal distributions (WIS_{pre}: $K-SZ = .749, p = .63$; WIS_{post}: $K-SZ = .793, p = .56$; WIS_{Δ}: $K-SZ = .983, p = .29$) and are thus z-transformed and compared with independent sample t-tests (compare Table 8.11). It turns out that the slight superiority of experience learners is not mirrored in superiority in declarative TAYLORSHOP scenario knowledge at any time. Thus, a dissociation between performance (in profit) and knowledge on individual level is visible in the data, induced by experience learning.

Table 8.11: Independent sample t-test between experience learners and text learners for the three declarative knowledge measures ($N = 134$)

	Knowledge acquisition								
	Texts only			Texts & experience					
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	t	df	p
ZWIS_{pre}	70	.09	1.02	64	-.02	.95	.628	132	.53
ZWIS_{post}	70	.09	1.05	64	-.05	.91	.879	132	.38
ZWIS_{Δ}	70	.01	1.01	64	-.06	1.01	.397	132	.69

In order to test whether an exposure to the TAYLORHOP microworld during discussion led to a further increase in CPS for those in the experience learning condition, a second Mann-Whitney test for TAYLORSHOP performance is carried out, comparing profit and trend between those groups who had a scenario present during the discussion and those who had had one during learning, but not during the discussion. The Mann-Whitney U Test does not show significant differences, neither for profit ($Z = -.744, p = .23$, one-tailed), nor for trend ($Z = -1.247, p = .11$, one-tailed). Similarly to the abovementioned results, a slight superiority of those who were subjected to experience learning and to a scenario during discussion compared to those who had no scenario available, neither during learning, nor during discussion, is visible for profit (Z = -1.341; p = .09) and trend (Z = -1.319, p = .09). All in all, the data provide marginal support for Hypothesis 4.6.1 with reference to learning mode and profit.

8.2.5 Knowledge acquisition and structural knowledge

As a marginally significant dissociation between TAYLORSHOP performance and verbalizable knowledge for experience learners has been established, this subsection analyzes whether this dissociation is accompanied by an increased structural implicitness (as assumed by Hypothesis 4.6.2). As structural implicitness is most closely operationalized with the weighted density measure (compare Sections 3.9.3 and 8.1.1), it is tested whether the weighted density exhibits significantly higher values for experience learners in contrast to participants who only learned from texts. The other five available AST-coefficients (number of concepts (degree), number of edges (size), clusters, diameter, and RMPL) will also be subjected to this analysis.

Only the variable graph size (number of edges in the PFNET) displays a normal distribution ($K - SZ = .585, p = .88$). Thus, this variable will be submitted to a t-test while the other values will be submitted to Mann-Whitney U-tests between groups that worked on the system during learning and those groups who did not. For the graph size, no significant increase in edges under the experience condition is observed ($t(131) = .84, p = .20$, one-tailed). The non-parametric comparisons for the other graph theoretic values are presented in Table 8.12.

Table 8.12: Mann-Whitney U test for AST-elicited structural knowledge coefficients between experience learning and learning from texts only ($N = 134$)

	Knowledge acquisition						
	Texts only		Texts & experience				
	n	Mean rank	n	Mean rank	U	Z	p^a
# of nodes	70	70.60	64	64.11	2023	-.968	.17
Cluster	70	72.81	64	61.69	1868	-1.670	$.08^+$
Diameter	70	72.24	63	61.17	1838	-1.852	$.06^+$
W. density	70	61.92	63	72.64	1849	-1.602	$.05^{*a}$
RMPL	70	66.06	63	68.04	2139	-.296	.38

[a] The significance value for the weighted density measure is one-tailed as Hypothesis 4.6.2 assumes more structural implicitness under the experience condition.

It becomes visible that the knowledge graphs of participants who learned from experience have less clusters and a smaller diameter. Both differences reach marginal levels of significance. At the the same time, knowledge graphs elicited under the experience condition are significantly denser than those networks from participants who learned from texts only. This finding is in line with the logical structure of the AST (compare Section 3.9.3): The weighted density depends negatively on the number of edges and associated concepts. As there are fewer associated concepts in the text-only condition, the density in the experience condition is larger. The fewer number of associated concepts in the text-learning condition is most probably also responsible for the smaller diameter and number of clusters. Thus, a more favorable density in the experience condition automatically leads to less favorable values among the other coefficients. This issue will be discussed further in Section 9.1.1, but overall, Hypothesis 4.6.2, postulating denser knowledge graphs under experience learning, is supported by the data.

The fact that experience with the scenario leads to denser networks also supports Hypothesis 3.9.5, stating that task performance involving implicit knowledge for completion (as established by the marginal superiority of experience learning and its dissociation from verbalizable knowledge), correlates with the AST-elicited weighted density measure.

8.2.6 Structural knowledge and CPS performance

Hypothesis 4.6.3 assumes a relationship between structural knowledge parameters and TAYLORSHOP performance. This is tested by calculating a non-parametric Spearman rho correlation between structural knowledge coefficients and the two variables that operationalize TAYLORSHOP performance, as both of them and most of the independent variables are not normally distributed (compare previous sections). The correlations are presented in Table 8.13, together with the correlations between structural knowledge parameters and WIS_{post} and BIS-K.

Table 8.13: Spearman rho correlations between structural knowledge parameters, TAYLORSHOP performance, declarative knowledge and reasoning (N = 134). $^{+}$ indicates marginal significance on 10% level (one-tailed), * indicates significance on 5%-level (one-tailed), ** on 1%-level (one-tailed).

	(1)	(2)	(3)	(4)	(5)	(6)
Profit	-.08	.02	-.08	.02	$.18^{*}$	.05
Trend	-.08	.02	-.08	.01	$.16^{*}$	.06
BIS-K	.07	$.14^{*}$	.02	-.02	-.01	-.12
WIS_{post}	.03	.10	.00	$.12^{+}$	.02	.00
1. # of nodes	–	$.44^{**}$	$.82^{**}$	$.38^{**}$	$-.78^{**}$	$-.88^{**}$
2. # of edges		–	$.30^{**}$	$.19^{*}$	$-.21^{**}$	$-.45^{**}$
3. Cluster			–	$.30^{**}$	$-.65^{**}$	$-.73^{**}$
4. Diameter				–	$-.43^{**}$	.04
5. W. density					–	$.62^{**}$
6. RMPL						–

Only one feature of structural knowledge, the weighted density, significantly correlates with both TAYLORSHOP performance measures profit and trend. At the same time, the weighted density displays near-zero correlations with the established determinants of TAYLORSHOP performance declarative knowledge (WIS) and reasoning (BIS-K). This indicates an additional variance explanation of the density measure on CPS performance; an assumption one would test by adding the weighted density measure to a hierarchical regression that includes the established determinants of CPS and then adds the weighted density measure. However, two assumptions of regression models are violated by the data: Normal distribution of error terms leading to a normal distribution of residuals (Statsoft Inc., 2006) and linear independence of regressors (Mansfield & Helms, 1982): The residuals of a regression involving not normally distributed variables will most likely not be normally distributed as well, and the other established predictors of CPS performance either correlate (BIS scores and knowledge) or are in direct linear dependency ($\text{WIS}_{post} = \text{WIS}_{pre} + \text{WIS}_{\Delta}$). These two problems will be handled through permutation tests and lasso regressions[1] in the following.

Not normally distributed residuals do not prevent the conduction of a regression analysis, but they render the p-values meaningless. The beta weights of a regression always minimize the quadratic error of the fit, but without a normal distribution of errors, the statistics underlying p-value computation in regression analysis do not follow a t- or F-distribution, which is assumed when calculating the p value. If the distributions of the t- and F-statistics are not known, no concrete p can be supplied. A workaround for this is the use of permutation tests for calculating a p value. Permutation tests were developed in the 1930s by Fisher and Pitman (Fisher, 1935; Pitman, 1937, 1938). In order to test whether two sample means originate from two (alternative hypothesis) or one underlying distribution(s) (null hypothesis), permutation tests pool the data from two samples A and B (of size n_A and n_B) into one sample C of size n_{A+B}. n_A cases are drawn from C and their average is computed and compared to the average of the remaining sample of size n_B. This process is repeated frequently until an estimation of the distribution of the mean difference given the null hypothesis is obtained. The one-sided p-value denotes the proportion of permutations where the difference in means is larger than the observed difference. The two-sided p-value employs the absolute values.

Anderson and Legendre (1999) summarized various permutation-based tests of significance of a single predictor variable in multiple regression, testing the null hypothesis that the predictor's beta coefficient equals zero. "An important application of such permutation tests is their use in canonical analysis of multivariate data..., where the data generally do not fulfill the assumptions required by traditional parametric testing procedures" (M. J. Anderson

[1] I am very grateful to Stephan Kolassa for his suggestions on those methods and for his help with the according R scripts.

& Legendre, 1999, p. 272). One of the approaches presented is the permutation of residuals under the full model as proposed by ter Braak (1992). It proceeds as follows:

A standard least-squares regression of the dependent variable Y with the independent variables X and Z of the type

$$Y = \beta_0 + \beta_{1-2}X + \beta_{2-1}Z + \epsilon$$

is performed. It delivers b_0, b_{1-2}, and b_{2-1} as estimates of the corresponding β values and residuals R corresponding to ϵ.

For a specific estimate of a β coefficient b, a reference value t_{ref} is obtained by calculating $t = (b-0)/se(b)$ where se denotes the estimated standard error of the partial regression coefficient. The residuals are permuted randomly, producing $R*$, and new values $Y*$ are created by calculating

$$Y* = b_0 + b_{1-2}X + b_{2-1}Z + R* \,.$$

"The new values $Y*$ are regressed on X and Z to obtain an estimate b^*_{2-1} and a value t^* under permutation. Here, $t^* = (b^*_{2-1} - b_{2-1})/se(b^*_{2-1})$" (p. 280), and the permutation of R, the calculation of Y^* and the new regression are repeated a large number of times, yielding a distribution of values of t^* under permutation. The two-sided significance value p denotes the proportion of absolute values in this distribution larger than $|t_{ref}|$, the one-sided significance the percentage of values larger than t_{ref}.

Anderson and Legendre (1999) employed simulation studies in order to demonstrate that this procedure is robust to several violations that underlie classic parametric calculation of significance values, including violations of the assumption of normal distribution of residuals and non-collinearity of regressors.

Collinearity of regressors also bears the risk of delivering unrealistic beta weights, as an increase of collinearity between columns of the design matrix increases the probability that the error terms, the statistical white noise, influence the calculation of betas (M. J. Anderson & Legendre, 1999). Subsampling and shrinkage are strategies in regression analysis that are robust against collinearity issues (Hastie, Tibshirani, & Friedman, 2001).

"In this approach, we retain only a subset of the variables, and eliminate the rest from the model. Least squares regression is used to estimate the coefficients of the inputs that are retained. There are a number of different strategies for choosing the subset" (ibid, p.55).

> By retaining a subset of predictors and discarding the rest, subset selection produces a model that is interpretable and has possibly

> lower prediction error than the full model. However, because it is a discrete process – variables are either retained or discarded – it often exhibits high variance, and so doesn't reduce the prediction error of the full model. Shrinkage methods are more continuous, and don't suffer as much from high variability (ibid, p. 59).

The least absolute shrinkage and selection operator (lasso) is a regression technique first proposed by Tibshirani (1996). It regularizes a least squares regression by imposing a penalty on the size of regression coefficients (Hastie, Taylor, Tibshirani, & Walther, 2007). It fits a linear model

$$f(x) = \beta_0 + \sum_{j=1}^{p} x_j \beta_j \tag{8.1}$$

by solving the optimization problem

$$min_\beta \sum_{i=1}^{N} \left(y_i - \beta_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \text{subject to} \sum_{j=1}^{p} |\beta_j| \leq s \tag{8.2}$$

The lasso method performs a continuous subset selection (Hastie et al., 2001), because "choosing an s can be thought of as choosing the number of predictors to include in a regression model" (Hastie et al., 2007, p. 2). If the tuning parameter s is set to 1, it delivers ordinary least square estimates, as the beta weights are scaled in such a way that they sum up to 1.

In summary, the lasso selects a subset of predictors that deliver a maximum in variance explanation of the full model including all predictors, but whose included parameters are below the threshold of a collinearity that violates regression model assumptions. If a regressor from the set of regressors "offered" to the lasso procedure is included in the model before s reaches 1, this indicates its importance for the prediction. The lasso thus produces a rank-order of regressors, sorted by relevance. According to Hastie (2007), this is due to the fact that lasso chooses always that regressor for inclusion in the next step that correlates highest with the residuals of the model of the current step.

In the following, two lasso regressions are performed with computer experience, economic knowledge, BIS scores, WIS_{pre}, WIS_{post}, WIS_{Δ}, and the weighted density measure; one with profit as dependent variable, the other with trend. The employed algorithms are described in Hastie, Taylor, Tibshirani and Walther (2007). For computation, the LARS package (Hastie & Efron, 2007) of the R software (R Development Core Team, 2007) was employed.

Lasso Regression for TAYLORSHOP Profit

A lasso regression with the dependent variable TAYLORSHOP Proft (Profit) is performed. It delivers the variable sequence reported in Table 8.14. The

Table 8.14: Sequence of lasso regression moves with TAYLORSHOP profit as dependent variable

Step	Variable #	Variable	*df*	RSS	C_p
0	-	-	1	1.954×10^{12}	37.063
1	5	WIS_{post}	2	1.819×10^{12}	24.477
2	3	BIS-K	3	1.792×10^{12}	27.125
3	2	Economic knowledge	4	1.689×10^{12}	20.364
4	6	WIS_{Δ}	5	1.658×10^{12}	19.738
5	1	Computer experience	6	1.616×10^{12}	18.145
6	7	AST weighted density	7	1.554×10^{12}	14.814
7	4	WIS_{pre}	8	1.450×10^{12}	8

standardized beta weights over the incrementing s parameter are plotted in Figure 8.3. Note that the weighted density is included before s – the tuning parameter that is gradually increased, compare Equation 8.2 – reaches .4. This indicates a certain significance of the weighted density, although its relevance appears to be small. The lasso indicates that the variable WIS_{Δ} is of least importance for prediction of the dependant variable. This is not surprising, as WIS_{Δ} is collinear with WIS_{pre} and WIS_{post}.

Due to its low significance (and its collinearity) as identified by the lasso, WIS_{pre} is dropped as a predictor. The remaining variables are inserted in a linear regression. As its residuals are not normally distributed as visible in the deviation from the main diagonal of the q-q plot (compare Figure 8.4), the one-tailed permuted p (calculated with ter Braak's method as explained above) is reported (compare Table 8.15). It accounts for 22.35% of the overall variance (multiple $R^2 = .223$, adjusted $R^2 = .184$). This indicates that the proposed model is able to explain a substantial fraction of the regressand variable profit, and that the weighted density makes a small but marginally significant contribution to this regression. Thus, the weighted density measure has a small and marginally unique explanatory power in predicting the TAYLORSHOP profit, and Hypothesis 4.6.3 receives marginal support for the profit variable. The strongest predictors of TAYLORSHOP profit in the lasso regression are post-discussion knowledge, reasoning, and prior economic knowledge.

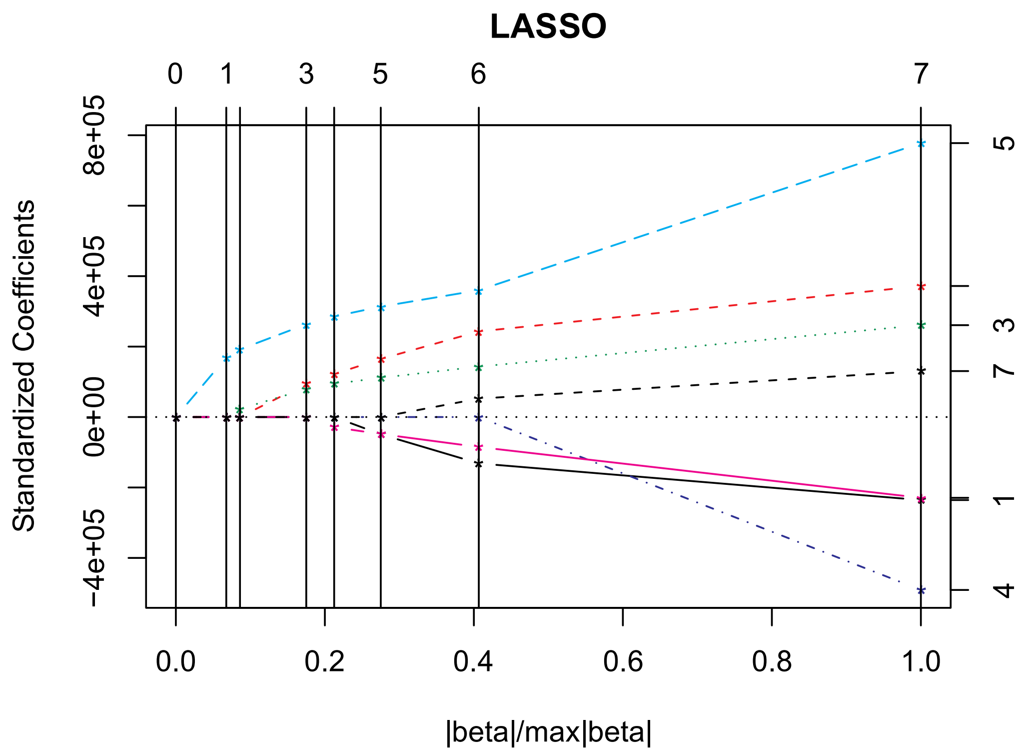


Figure 8.3: Plot of standardized β coefficients over the lasso tuning parameter s (denoted as $|\beta|/max|\beta|$) for TAYLORSHOP profit. The variable names according to variable numbers are presented in Table 8.14, the vertical lines represent the steps of the lasso regression given in Table 8.14. The weighted density is denoted by variable 7.

Lasso Regression for TAYLORSHOP Trend

A lasso regression with the regressors computer experience, economic knowledge, BIS-K, WIS_{pre}, WIS_{post}, WIS_{Δ}, and the weighted density measure and the trend variable as regressand is performed. It delivers the variable sequence reported in Table 8.16.

The standardized beta weights over the incrementing s parameter are plotted in Figure 8.5. Note that in this regression, the weighted density is included before s reaches .3. This indicates a stronger significance of the weighted density in predicting the trend measure, although its relevance appears to be small. The lasso indicates that the variable WIS_{Δ} is of low importance for prediction of the dependent variable. This is not surprising, as WIS_{Δ} is collinear with WIS_{pre} and WIS_{post}. For the prediction of the

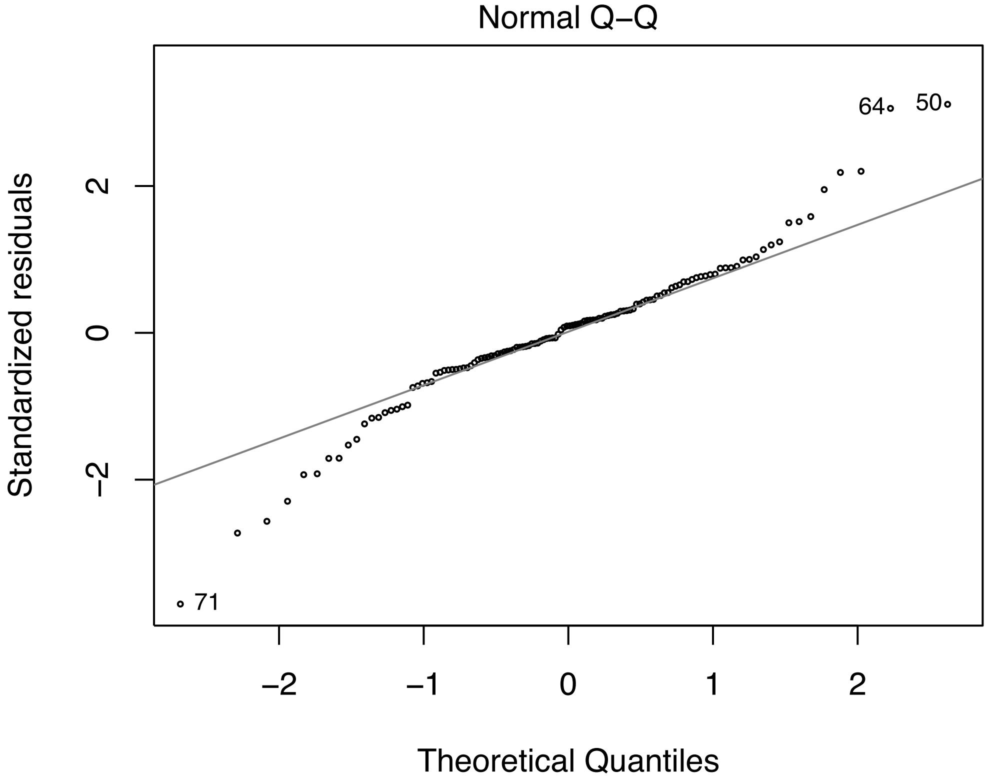


Figure 8.4: Q-Q plot of standardized residuals of the regression of profit on WIS_{post}, WIS_{delta}, BIS, computer exprience, and economic knowledge.

trend, the computer experience appears to be even less important than the collinear WIS_{pre}. The two variables are thus excluded, and a linear regression is performed.

Like in the previous case, its residuals are not normally distributed as visible in the deviation from the main diagonal of the q-q plot (compare Figure 8.6). Thus, instead of regular significance values, the one-tailed permuted p (calculated with ter Braak's method as explained above) values are reported (compare Table 8.17). The regression accounts for 26% of the overall variance (multiple $R^2 = .260$, adjusted $R^2 = .229$). As in the case of the profit, these results indicate that the proposed model is able to explain a substantial fraction of the regressand variable profit, and that the weighted density makes a small but marginally significant contribution to this regression. Again,

Table 8.15: Linear regression of TAYLORSHOP profit on WIS_{post}, BIS, economic knowledge, WIS_{Δ}, computer experience and AST weighted density.

	b	Std. error	t	p	f^2
(Intercept)	-353163	151466	-2.332	$.010^{**}$	-
WIS_{post}	4317	1398	3.089	$.001^{**}$	.142
BIS	2334	1591	1.467	$.072^{+}$	.026
Economic knowledge	30316	9900	3.062	$< .001^{**}$	.046
WIS_{Δ}	-6482	5705	-1.136	.12	.017
Computer experience	-25728	12957	-1.986	$.030^{*}$	.037
AST weighted density	48437	32909	1.472	$.068^{+}$	.018

Table 8.16: Sequence of lasso regression moves with TAYLORSHOP trend as dependent variable

Step	Variable #	Variable	*df*	RSS	C_p
0	-	-	1	1611.2	37.063
1	5	WIS_{post}	2	1474.4	24.477
2	3	BIS-K	3	1361.1	27.125
3	2	Economic knowledge	4	1294.6	20.364
4	6	WIS_{Δ}	5	1248.4	19.738
5	7	AST weighted density	6	1213.5	18.145
6	4	WIS_{pre}	7	1129.2	14.814
7	1	Computer experience	8	1126	8

the strongest predictors of TAYLORSHOP trend in the lasso regression are post-discussion knowledge, reasoning, and prior economic knowledge.

Table 8.17: Linear regression of TAYLORSHOP trend on WIS_{post}, BIS, economic knowledge, WIS_{Δ}, computer experience, and AST weighted density

	b	Std. error	t	p	f^2
(Intercept)	-8.18	4.22	-1.936	$.029^{*}$	-
WIS_{post}	.12	.39	3.627	$< .001^{**}$	.234
BIS	.08	.045	1.791	$.041^{*}$	.040
Economic knowledge	.62	.263	2.381	$.010^{**}$	.041
WIS_{Δ}	-.25	.16	-1.551	$.070^{+}$	.019
AST weighted density	1.27	.931	1.362	$.091^{+}$	.016

In summary, hypothesis 4.6.3 (individual CPS performance is influenced

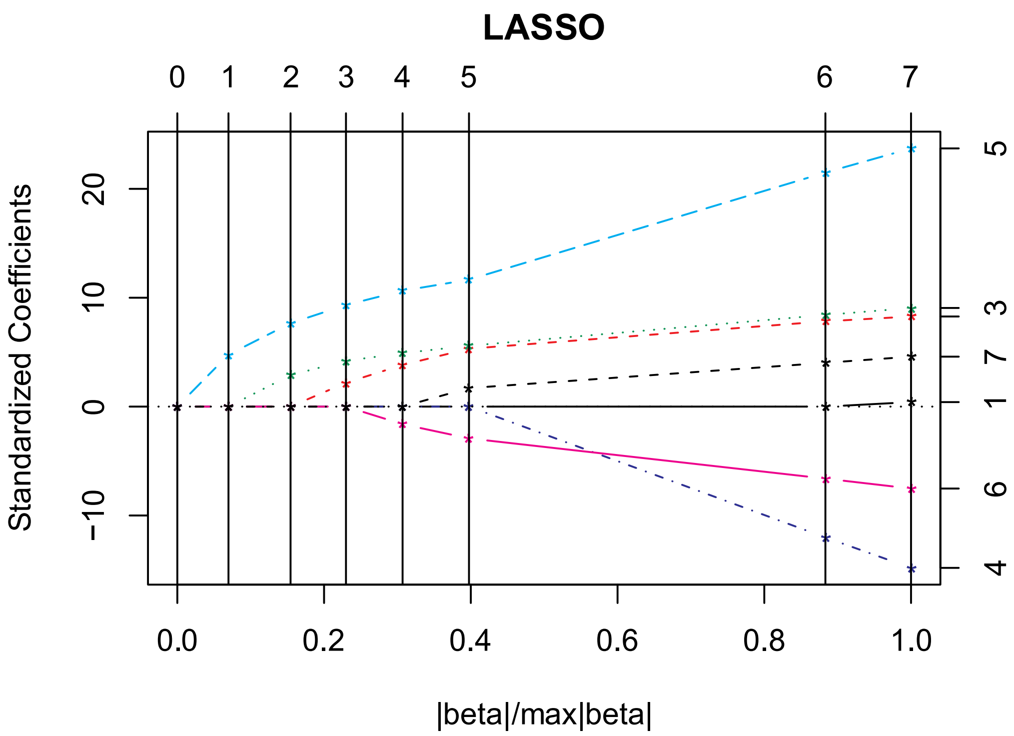


Figure 8.5: Plot of standardized β coefficients over the lasso tuning parameter s (denoted as $|\beta|/max|\beta|$) for TAYLORSHOP trend. The variable names according to variable numbers are presented in Table 8.16, the vertical lines represent the steps of the lasso regression given in Table 8.16. The weighted density is denoted by variable 7.

by scenario-relevant implicit structural knowledge) receives marginal support by the findings. Although significant correlations are visible, effects in comparison to the established predictors of CPS performance are at the brink of visibility.

8.2.7 Further findings: Self-evaluation and gender differences

Apart from the calculations with reference to the hypotheses, further effects are reported in this subsection. It is assessed whether the self-assessment of TAYLORSHOP knowledge on the ten knowledge elements has any connection with TAYLORSHOP performance and the other measurement variables. Furthermore, possible gender differences in TAYLORSHOP performance are

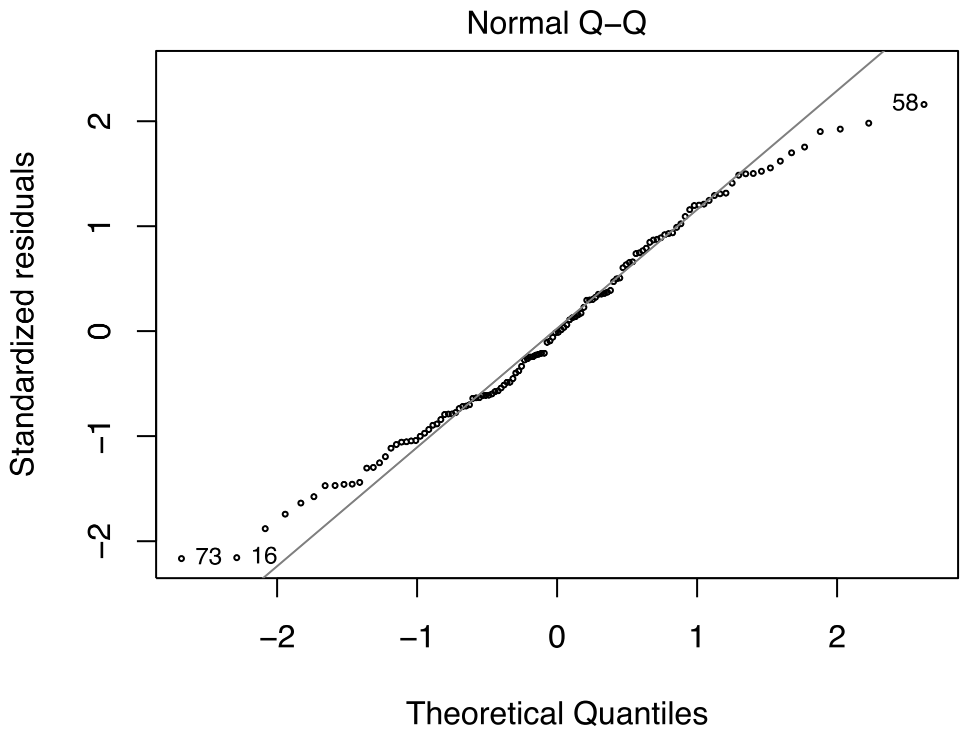


Figure 8.6: Q-Q plot of standardized residuals of the regression of trend on WIS_{post}, WIS_{delta}, BIS, computer exprience, and economic knowledge.

reviewed.

With reference to self-evaluation, the total self-assessment score displays a significant (two-tailed) correlation with prior economic knowledge ($r = .17, p = .05$) and AST diameter ($r = .21, p = .02$). All other correlations of self-assessment do not reach significant values, e.g., with TAYLORSHOP profit ($r = .10, p = .23$) and trend ($r = .07, p = .42$). Thus, self-evaluation of scenario knowledge is neither able to predict TAYLORSHOP performance, nor is it able to predict declarative knowledge scores. It can be seen as a reflection of general economic knowledge due to its small correlation with prior economic knowledge.

Since Wittman (1996) and Süß (1996) reported gender differences with respect to TAYLORSHOP scenario performance in such a way that men achieved significantly better performance scores than women, this effect is also checked. Therefore, a Mann-Whitney U test between male and female problem solvers and the two measures of TAYLORSHOP performance is employed. A significant difference for both overall profit ($Z = -4.939, p < 0.001$ (two-tailed)) and trend ($Z = -5.295, p < 0.001$ (two-tailed) is observed: Male participants scored substantially higher on both criterions: For profit, male participants achieved a median loss of -5301 DM while female participants achieved a median loss of -54056 DM. Male participants exhibit a median trend of seven, female participants exhibit a median trend of three. This finding requires a further analysis in order to determine whether the effects reported above are artifacts of this effect or are confounded with it, because male and female participants were not intentionally balanced over the cells of the experimental design (compare Section 7.2.1). Therefore, a χ^2 analysis is performed in order to test whether there are significantly more men than women in the experimental conditions of relevance for data analysis on individual level (differences in gender compositions of dyads on group level will be analyzed in the section on group results). The two findings on individual level where fixed factors could be contaminated by the random factor gender are:

1. Participants who learned from experience achieved marginally significantly higher profit scores in the dynamic scenario compared to those who learned from texts only.
2. Participants who learned from experience exhibited a marginally significant higher weighted density of AST-elicited pathfinder-adjusted knowledge networks (PFNETs).

Since the non-normal distribution of data prevents the use of an analysis of variance, a cross-tabulation with a Chi-Square test is performed. It reveals that there are no significant differences in the distribution of male and female participants over the factor learning mode (compare Table 8.18). It reveals a $\chi^2(N = 134) = 1.802, p = .18$; Fisher's exact test of significance (2-sided) = .225 Thus, there is no significant overrepresentation of any gender in the factor learning mode. Thus, the previously discovered effects are most probably not solely attributable to gender artifacts.

8.3 Laboratory Experiment Results on Group Level

In this section, the results of hypotheses testing on the group level will be reported. The level of analysis is shifted, and the unit of analysis will be

Table 8.18: Crosstabulation of participant gender and learning mode ($N = 134$)

	Gender		
	Female	Male	Total
From texts only	42	28	70
From texts and experience	31	33	64
Total	73	61	134

the dyad instead of the individual participant. Therefore, variables will be averaged. This also pays tribute to the nonindependence of the profit variable (compare Section 8.2.3). In this way, the following new variables are created: The dyadic average of the short version of the BIS scale K, "reasoning" (dBIS-K), average dyadic self assessment of computer experience (dCE) and prior economic knowledge (dEK), dyadic average of pre- and post-discussion WIS declarative TAYLORSHOP scenario knowledge (dWIS$_{pre}$ and dWIS$_{post}$), and post-discussion knowledge increase (dWIS$_{\Delta}$). Furthermore, there is the skillMap similarity measure *sksim*, calculated on dyadic level (compare Section 6.2) and the Euclidian distance of self-assessment, as explained in the following section. Structural knowledge coefficients are not calculated on dyadic level, because there is no hypothesis referring to them on group level. For the assessment of TAYLORSHOP scenario performance, the following coefficients are available: dProfit (dyadic average), dTrend (dyadic average), number of sold shirts (dyadic average), and dProfit per shirt (dyadic average).

8.3.1 Operationalization of scenario performance on group level

In order to obtain comparability of group level results with those on individual level, it would be desirable to use the same operationalizations for TAYLORSHOP profit on group level (dyadic average of profit and trend) as on individual level. Since some of the hypotheses require the testing of polynomial hypotheses, normal distribution of dependent variables is another desirable feature, because it would allow the use of variance analysis techniques from the general linear model. Thus, dyadic average of profit and trend are subjected to Kolmogornov-Smirnov tests for normal distributions. Both the dyadic average of profit ($Z = .637, p = .81$) and trend ($Z = .789, p = .56$) display a standard normal distribution. As participants were instructed to maximize their profit and the dyadic average profit (referred to as dProfit) exhibits significant correlations with the established outside criteria of CPS (compare Table 8.20), it is employed as the single measure for CPS performance on group level. For easier comparability, its z-transformed score will be employed as operationalization of TAYLORSHOP scenario performance on the dyadic level. All other variables on the dyadic level also display a normal distribution, except for computer experience (dCE, compare Table 8.19). The correlation matrix of the measurement variables on the group level is presented in Table 8.20.

Table 8.19: Kolmogorov-Smirnov Test for normal distribution of independent variables on the dyadic level

	dBIS-K	dCE	dEK	dWIS$_{pre}$	dWIS$_{post}$	dWIS$_{\Delta}$
$\bar{x}$	101	3.63	2.38	18.85	21.90	3.08
SD	5.57	.625	.831	6.38	6.70	3.76
Z	.516	1.501	1.182	.497	.620	.690
$p(Z)$	.953	.020	.950	.970	.840	.730

Table 8.20: Intercorrelation matrix of measurement variables on group level. * denotes significance on 5%, ** on 1%, and *** on .1% level. + denotes marginal significance on 10% level. All sigificance values are one-tailed.

	dBIS-K	dWIS$_{pre}$	dWIS$_{post}$	dWIS$_{\Delta}$	dEK	dCE[a]
dProfit	.25*	.24*	.37**	.25*	.15	.05
dBIS-K		.61***	.55***	-.04	.03	.19^{+}
dWIS$_{pre}$			.84***	-.22*	.06	.18^{+}
dWIS$_{post}$				.34**	-.07	.19^{+}
dWIS$_{\Delta}$					-.20*	-.03

[a] Correlations between dCE and other variables are Spearman rank-order correlations. All other correlations are Pearson correlations.

8.3.2 Comparability of knowledge distribution condition

Over the three conditions of knowledge distribution (shared, partially shared, unshared), a one-sample ANOVA reveals no statistically relevant differences in dyadic average scores for reasoning (BIS-K) ($F(2, 64) = 0.616, p = 0.543$) and pre-discussion TAYLORSHOP declarative scenario knowledge dWIS_{pre} ($F(2, 64) = 1.01, p = 0.367$). Possible differences for computer experience and economic foreknowledge were tested with the Kruskal-Wallis-Test revealing no statistically relevant difference between conditions for economic foreknowledge ($\chi^2 = 4.01, p = .260$). However, despite the random assignment of participants to dyads, a difference of computer experience between conditions was visible ($\chi^2 = 9.6, p = .020$): The participants assigned to the unshared condition showed less average computer experience than groups in the partially shared and shared condition. Since no significant correlation between the employed operationalization of computer experience and individual TAYLORSHOP profit was found ($r = -.03, p = .400$), this result can be neglected.

8.3.3 Subjective experience of knowledge-overlap

In order to test whether the manipulation of knowledge overlap was subjectively experienced, the Euclidian distance between group members in self-assessment on TAYLORSHOP expertise is calculated for each dyad. The Euclidian distance is a vector with ten elements; each element is the absolute value of the difference of the two individual responses to the same item on the self-assessment questionnaire in one dyad. One element of the vector can thus range from 0 to 3. The elements of the vector are summed up and divided by the possible maximum (30). In this way, the Euclidian distance in self-assessment is a number between 0 and 1. 1 indicates a maximum difference in knowledge self-assessment between the two group members, 0 indicates a perfect overlap of knowledge self-assessment. If the manipulation in terms of knowledge overlap inside the dyad was successful, groups in the unshared condition should display a larger Euclidian distance than groups in the partially shared condition, who again should display a larger Euclidian distance than groups in the shared condition. This pattern is visible in the data (compare Figure 8.7) and a one-way analysis of variance of the Euclidian distance over the factor knowledge distribution exhibits significant results ($F(2, 64) = 4.688, p = .013, \eta^2 = .132$).

If all ten self-assessment items are summed up into one total self-assessment score and this score is averaged on dyadic level, there is no significant correlation between this score and taylorshop profit on dyadic level (r = .05, p = .66). The skillMap similarity measure is not employed for the analysis of knowledge overlap, because its connection to the subjective experience of knowledge overlap is not determined at this point, as weighted semantic relations between the different items of the self-assessment questionnaire are

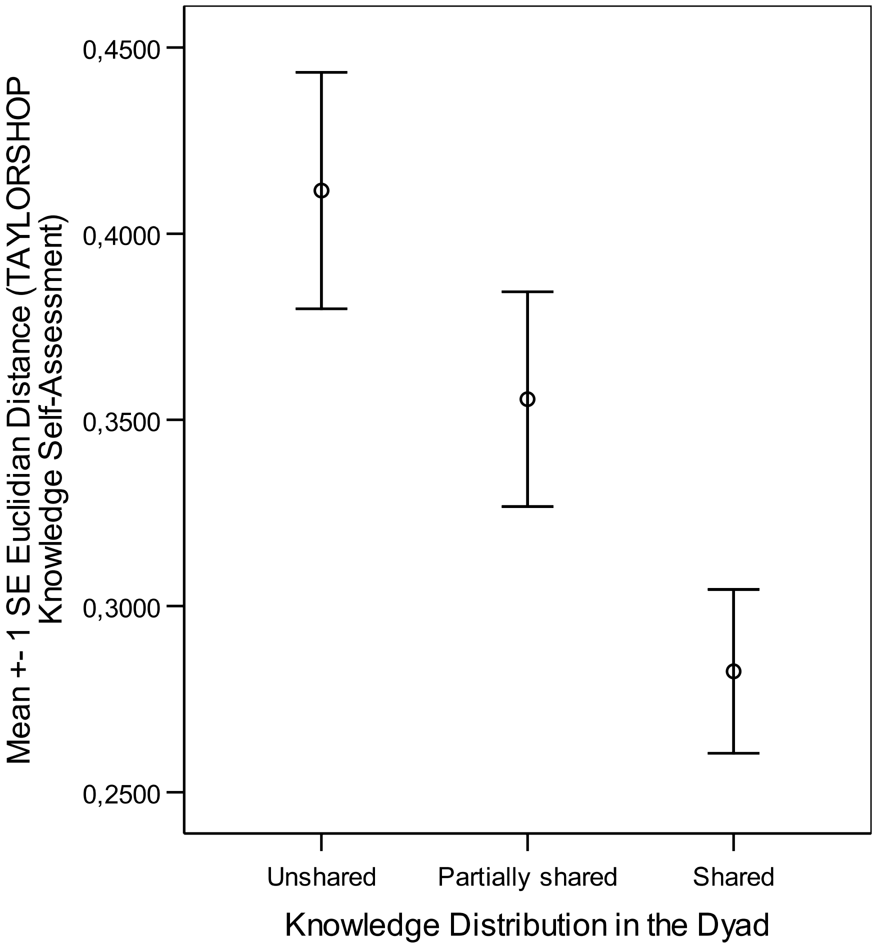


Figure 8.7: Euclidian Distance of TAYLORSHOP-relevant knowledge self-assessment in the dyad with regard to experimental conditions

employed in its calculation and its relation to cognitive group heterogeneity is hypothetical. The relationship between the knowledge dstribution in the dyad and the skillMap similarity measure will be tested in Section 8.3.7.

8.3.4 Knowledge distribution and knowledge increase

In order to test Hypothesis 5.5.1 (knowledge overlap among group members has a curvilinear n-shaped impact on knowledge increase), a one-way analysis of variance with linear and quadratic polynomial contrasts of the effect of knowledge distribution on post-discussion knowledge increase (dWIS_{Δ}) is performed. The combined model does not reach significance ($F(2, 69) =$

$1.684, p = .19, \eta^2 = .048$), nor does the linear term ($F(1, 69) = .048, p = .83, \eta^2 = .000$). The quadratic term exhibits a marginally significant effect ($F(1, 69) = 3.320, p = .073, \eta^2 = .047$). The pattern of the data follows the prediction of the n-shaped connection between knowledge overlap inside the dyad and knowledge increase (compare Figure 8.8). Hypothesis 5.5.1 thus receives marginal support.

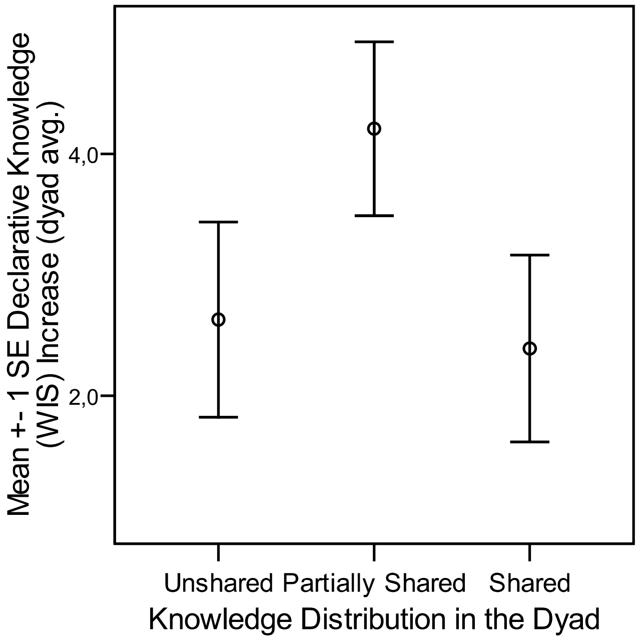


Figure 8.8: The effect of knowledge distribution in the dyad on average post-discussion knowledge increase ($dWIS_{\Delta}$)

8.3.5 Knowledge increase and group performance

Hypothesis 5.5.2 assumes a positive correlation between a group's knowledge increase and its CPS performance. The group average of declarative knowledge increase after discussion ($dWIS_{\Delta}$) correlates with average group TAYLORSHOP profit at $r = .25$ ($p = 0.26$ (one-tailed)). The Hypothesis is thus supported. Hypothesis 5.5.3 postulates an n-shaped relation between knowledge overlap and CPS performance, which is observed in the data (compare Figure 8.9). A one-way analysis of variance with linear and quadratic polynomial contrasts of the effect of knowledge distribution on TAYLORSHOP performance reveals a significant combined effect between groups ($F(2, 69) = 4.810, p = .011, \eta^2 = .125$), a marginally significant effect of the linear term ($F(1, 69) = 3.458, p = .067, \eta^2 = .045$), and a significant quadratic effect ($F(1, 69) = 6.162, p = .016, \eta^2 = .080$) as visible in Figure 8.9. Hypothesis 5.5.3 is thus supported.

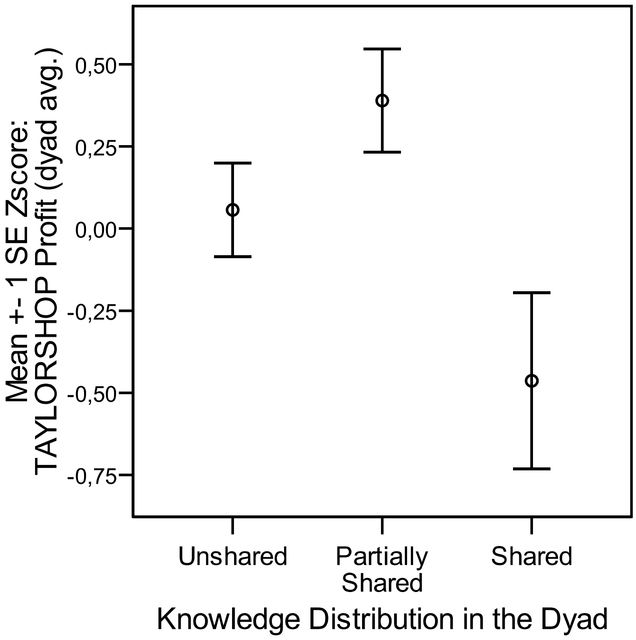


Figure 8.9: The effect of knowledge distribution in the dyad on average Complex Problem Solving Performance (dProfit)

8.3.6 Implicit knowledge on group level

In order to test the influence of implicit knowledge on TAYLORSHOP performance (Hypothesis 4.6.1) on the group level, the means and standard errors of the variable dProfit are plotted over the different levels of experience with the scenario (none / during learning / during discussion / during both), compare Figure 8.10. The highest level of exposure to the scenario prior to the actual control task leads to a higher performance on group level, but the difference is not significant as visible by the overlapping error bars. Hypothesis 4.6.1 is thus refuted on group level. A plot of the post-discussion dyadic average declarative knowledge (compare Figure 8.11) reveals that exposure to the scenario during learning leads to significantly less verbalizable knowledge compared to no exposure at all and exposure during learning and discussion. One might have expected that scenario exposure does not necessarily lead to an increase of declarative knowledge, buts its cooccurrence with a *decrease* in verbalizable knowledge is somewhat surprising. This finding will be further discussed in Section 9.2.3.

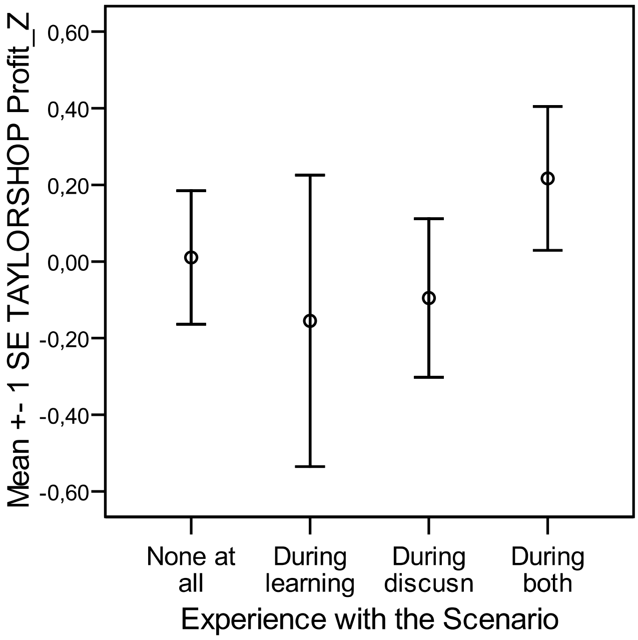


Figure 8.10: The effect of experience with the scenario on scenario performance (dProfit)

In order to test for interaction effects between knowledge distribution and experience with the scenario, a univariate analysis of variance of the effect of the two factors on dProfit is performed, including dBIS-K and dWIS$_{pre}$ variables as covariates (dWIS$_{post}$ and dWIS$_{\Delta}$ are not included due to their confoundation with the factor knowledge distribution). The model does not reach significance ($F(15, 69) = 1.216, p = .229$). It reveals only one significant effect of knowledge distribution on dProfit ($F(2, 69) = 3.636, p = .033, \eta_p^2 = .115$). Neither the main effect of experience on dProfit ($F(3, 69) = 0.271, p = .85$), nor the interaction effect between experience and knowledge distribution ($F(6, 69) = 0.462, p = .83$) reach significance. This provides further support for Hypothesis 5.5.2 stating that group average of declarative knowledge increase after discussion correlates with the group's complex problem solving performance.

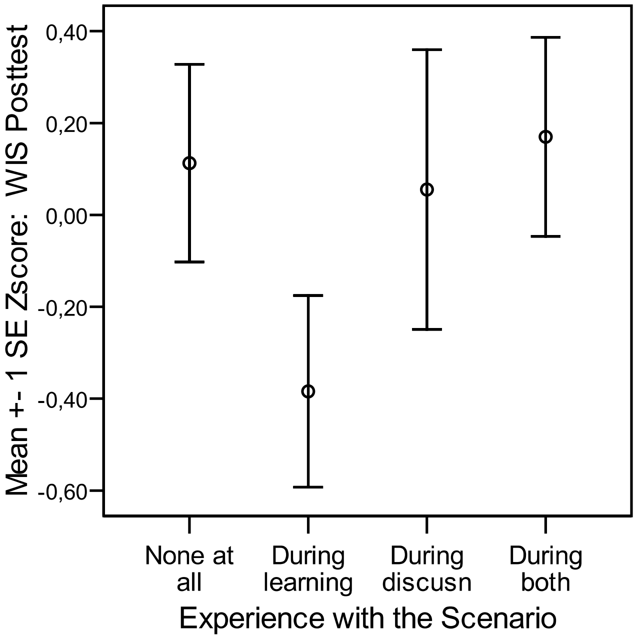


Figure 8.11: The effect of experience with the scenario on post-discussion verbalizable knowledge (dWIS$_{post}$)

8.3.7 skillMap similarity and group performance

Prior to hypothesis testing, the skillMap similarity data is screened for normal distribution. Due to technical problems during the similarity calculation, there are two invalid similarity measures, resulting in an sample size of 68 cases. A one-sample Kolmogorov-Smirnov-Test reveals that the skillMap similarity measure exhibits a normal distribution (K-S $Z = 1.062, p = .209$).

In order to test whether the proposed similarity measure for intra-group knowledge overlap displays a curvilinear n-shaped relation with CPS performance (Hypothesis 6.3.1), a quadratic regression of the groups' average CPS performance (dProfit) is performed on the skillMap similarity measure. Neither the linear term ($\beta = 10.19, p = .106$) nor the qadratic term ($\beta = -20.94, p = .339$) reach significance; overall variance explanation is low ($R^2 = .08$) and the overall model reaches marginal significance ($F(2, 64) = 2.874, p = .064$). Hypothesis 6.3.1 is thus refuted. Due to the marginal significance of the overall model, a correlation is calculated between dProfit and the skillMap similarity measure. The two variables correlate at $r = .26$ ($p = .032$, one-tailed). Thus, contrary to the hypothesis, a *linear* relationship between the similarity measure and CPS performance (dProfit) is observed.

Since dProfit exhibits an n-shaped distribution over experimental conditions, the similarity measure is also plotted over experimental conditions (compare Figure 8.12).It turns out that its distribution is very similar to the distribution of dProfit over experimental conditions: The similarity measure exhibits the highest values under the partly shared condition and has significantly lower averages in the other conditions as indicated by the non-overlapping error-bars. This surprising finding wil be further discussed in Section 9.2.5.

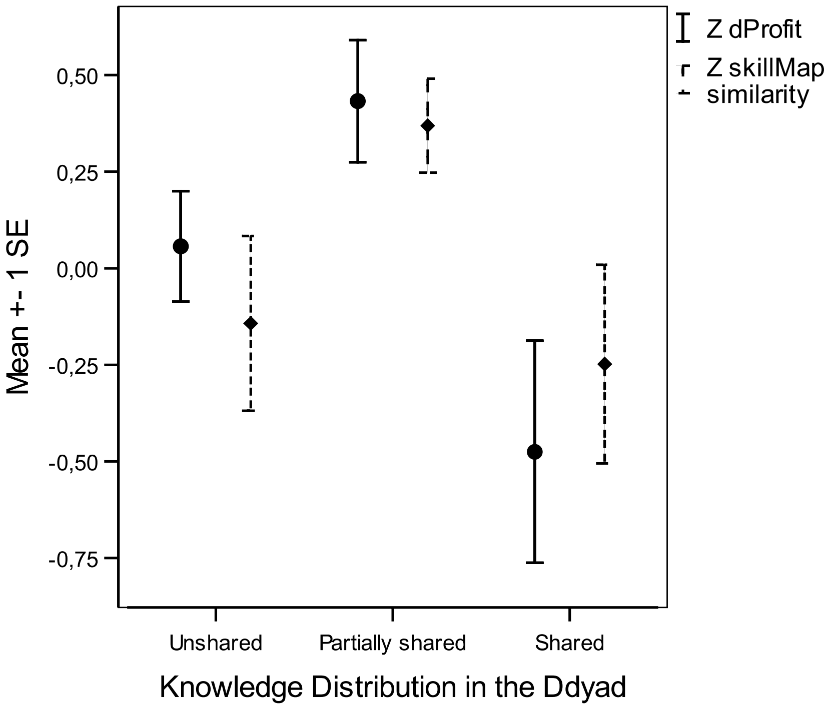


Figure 8.12: Error bar plot of the skillMap similarity measure and dyadic CPS performance (dProfit) over knowledge distribution condition

8.3.8 Gender effects

As gender effects were found on the individual level (compare Section 8.2.7), this effect is also examined on the group level. An error-bar plot of complex problem solving performance over the gender composition of the dyads shows that same-gender male dyads outperform mixed-gender dyads, which again outperform same-gender female dyads (compare Figure 8.13). A similar pattern is visible for dBIS-K (reasoning) scores and pre- and post-discussion declarative knowledge scores; the pattern of declarative knowledge increase

($dWIS_\Delta$) is less clear. One-way analyses of variance of the factor gender composition on these variables exhibit significant differences in means for dProfit ($F(2, 69) = 4.873, p = .011, \eta^2 = .071$), $dWIS_{pre}$ ($F(2, 69) = 4.283, p = .018, \eta^2 = .062$), and $dWIS_{post}$ ($F(2, 69) = 4.274, p = .018, \eta^2 = .062$). dBIS-K reaches marginal significance ($F(2, 69) = 2.812, p = .067, \eta^2 = .041$), and $dWIS_\Delta$ after discussion does not show significant differences ($F(2, 69) = .287, p = .752$). Thus, performance differences over different gender compositions might be due to higher dBIS-K and dWIS scores of dyads containing men. However, a partial correlation of gender composition (same gender female = 2, mixed gender = 3, same gender male = 4) and dProfit controlling for dBIS-K, $dWIS_{pre}$, $dWIS_{post}$, $dWIS_\Delta$, computer experience (dCE) and prior economic knowledge (dEK) still leads to a correlation of r = .22 (p = .020, one-tailed).

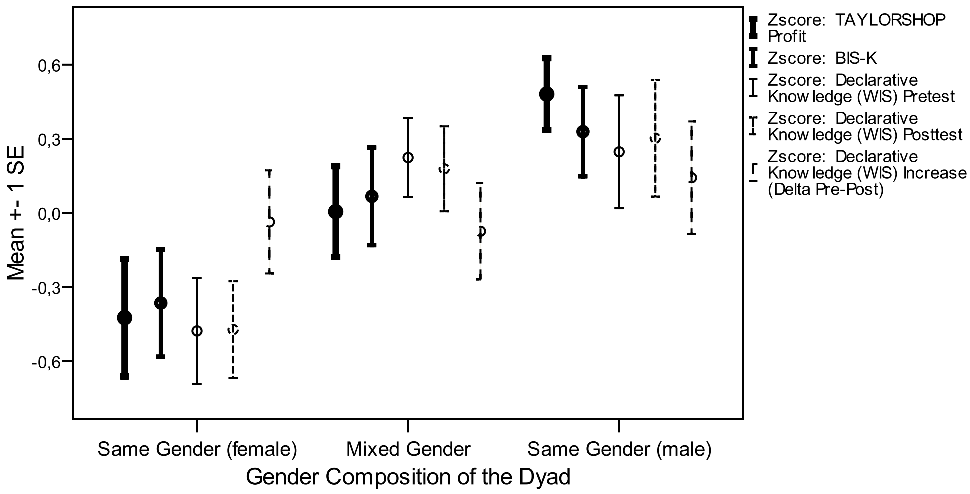


Figure 8.13: The effect of the gender composition of the dyad on TAYLORSHOP performance, reasoning, declarative knowledge pre- and post-discussion score and knowledge increase

As the assignment of gender to cells of the experimental design was not intentionally balanced (compare 7.2.1), it is checked whether the reported findings might be based on gender effects. A crosstable analysis of the factors knowledge distribution and gender distribution leads to $\chi^2(n = 70) = 5,604$ ($p = .231$). Thus, there are no significant over- or under representations of gender composition in the experimental design.

Chapter 9

Discussion

The second chapter of this thesis outlined the concept of knowledge and presented unconscious access to structural knowledge or structural implicitness as a part of individual implicit knowledge. The third chapter introduced a computer-based structural knowledge elicitation technique, the association structure test (AST), intended to elicit a part of this structural implicitness. The fourth chapter introduced the domain of complex problem solving, which encompasses highly relevant real-life tasks that managers and politicians face, and outlined how different kinds of knowledge and intelligence play a role in successful complex problem solving. The fifth chapter focused on complex problem solving tasks in groups, and presented theories on the relationship between group cognitive heterogeneity and group performance. The sixth chapter introduced a computer system from the realm of knowledge management, the skillMap, which aims at eliciting group knowledge and expertise through self-assessment. For this system, a method for quantifying dyadic similarity with regard to self-assessment of represented individuals was introduced. The following experiments, whose methodology was introduced in the seventh, and whose results were presented in the eighth chapter, aimed at testing several hypotheses put forward with regard to the ability of the introduced methods to measure knowledge and with regard to the connection between knowledge and performance in complex problem solving: It was assumed that the AST-elicited structural implicitness correlates with CPS performance and provides an additional and unique explanation of CPS variance in comparison to established predictors of CPS, such as declarative knowledge and intelligence. Complex problem solving performance of groups should display a n-shaped connection with group heterogeneity, and the skillMap similarity measure was supposed to predict group performance.

This chapter will discuss and integrate the results with regard to the influence of AST-elicited structural knowledge on performance, with regard to the connection between group heterogeneity and group performance, and with regard to the connection between the skillMap similarity measures and group performance. For each of the three fields, results will be integrated, and implications and issues will be discussed. The chapter ends with an outlook on future research and a conclusion.

9.1 Knowledge Elicitation with the AST

In this section, the results of the experiments that aimed at validating the association structure test (AST) are discussed and integrated. The discussion will try to answer the question whether it is possible to measure a part of unconsciously accessed individual structural knowledge with the AST test system in a valid way.

9.1.1 AST internal consistency

After an initial analysis of the logical structure of the AST coefficients number of nodes, number of edges, clusters, diameter, weighted density, and RMPL in Section 3.9.3, an inspection of empiric correlations among the coefficients reveals the same correlational structure as the logical structure: There are high levels of correlations among all coefficients and some display an inverse proportional relationship with each other. With regard to the overall high correlations among coefficients, it is visible that the three highest correlations occur between the number of nodes and three other coefficients: the relative maximum path length, the number of edges, and the weighted density. This illustrates the strong dependence of all AST coefficients on the number of nodes.

With regard to the inverse proportional relationship between coefficients, it becomes clear that all measures of the AST either have a conceptual correlation or are fully or partly mathematically deductible from each other. The main problem of this feature is the resulting negative correlation between the weighted density on one side and the number of edges and nodes on the other. The nodes are supposed to partly capture declarative knowledge, while the weighted density is supposed to capture structural knowledge and structure implicitness. A high weighted density will thus co-occur with a low number of edges. Translated into the intended underlying concepts, this would mean that a high level of structural implicitness in a certain domain co-occurs with low explicit knowledge in that domain and vice-versa. Although this may be plausible for some instances, it must not always be the case. A high degree of structural implicitness and a thorough explicit knowledge on a certain domain is very plausible. This inverse proportional relationship between two measurement variables that are supposed to capture concepts that are not inversely proportional to each other casts doubts on the AST's construct validity.

A factor analysis of empirical AST coefficient correlations (compare Section 8.1.1) reveals that five AST-coefficients (the number of nodes, the number of edges, the number of clusters, the weighted density measure, and the relative maximum path length (RMPL) are based on two underlying factors. Four of these coefficients, the number of nodes in the graph, the number of edges in the graph, the number of clusters, and the relative maximum path length, load on one factor, while the second factor exhibits only one substantially loading coefficient, the weighted density.

The number of nodes and the relative maximum path length (RMPL) measures were originally assumed to tap into aspects of declarative knowledge. Thus, their underlying factor is labeled as explicit knowledge. The number of edges was originally supposed to capture the implicit and explicit structural aspect of knowledge, but due to its logically tight coupling with the number of nodes, the number of edges also loads on the factor that is already loaded by the number of nodes. The number of clusters and the diameter were originally thought to range somewhere in between declarative knowledge and structural knowledge. Due to its tight logical connection with the number of associated concepts, the number of clusters also loads on the first factor. The diameter measure does not exhibit substantial loadings on the two extracted factors, which leaves its role and its interpretability unclear. This is especially unfortunate as other researches have credited the diameter of Pathfinder-elicited knowledge networks with a certain explanatory power (Eckert, 1998b).

Adjusting the number of edges to the graph degree (the number of nodes) by calculating the weighted density weakens the tight connection between the number of edges and the number of nodes: The absolute correlation between weighted density and number of nodes is only half as large as the absolute correlation between the number of edges and the number of nodes. The factor analysis shows that the weighted density measure is part of an underlying construct which is different from those constructs which are closely related to the number of nodes (number of edges, number of clusters, and the relative maximum path length RMPL). These two underlying constructs correlate, but they are not completely overlapping. As the weighted density is considered to be the coefficient that captures structural features best (compare Section 3.9.1), the second factor is labeled "structural knowledge including structural implicitness". The observed medium correlation between the factors supports the assumption that the different types of knowledge are interrelated (compare Chapter 2). The identification of this second factor thus supports the original assumption that the AST is able to capture a certain aspect of structural implicitness.

In the overall assessment of the AST, despite a somewhat encouraging underlying factor structure, two problems remain: The ambiguity of the diameter variable and, more severe, the inverse proportional relationship between the weighted density measure on the one hand and the number of nodes and edges on the other. These issues render the AST imperfect. Whether they also render it unusable for diagnostic purposes will be analyzed in the following subsections.

9.1.2 Labeled and unlabeled AST testing

The AST is supposed to elicit structural knowledge (Jonassen et al., 1993) that is unconsciously accessed (compare Chapter 2). It tries to do so by minimizing the cognitive load when asking for the existence of connections (edges) between previously associated concepts in a pairwise manner without displaying the entire graph and by asking for a quick and intuitive judgment on the existence of a relation between two concepts based on 'gut-feeling'. Such a mode of elicitation is supposed to include both those edges that can be explicitly labeled and those that the participant cannot define. If this mode of edge elicitation is compared to a mode where edges have to be explicitly labeled, it should contain significantly more edges (compare Section 3.9).

This hypothesis is supported by the results: Given a comparable amount of associated concepts, quick and intuitive placements of edges lead to a larger amount of placed edges than the placement of edges that have to be labeled (compare Table 8.3). Note that this holds true for the *adjusted* PFNETS (compare Section 3.7) in which edges that violate the triangulation assumption have been removed. The increased amount of edges placed under the non-explicit condition also influences the diameter of the structural knowledge graph and its measure of relative maximum path length (RMPL) to similar extents. The weighted density measure, which is conceptualized as operationalizing structural implicitness to the closest extent, remains below levels of significance, but displays the predicted trend, i.e. a higher density for graphs elicited under the non-explicit condition.

These results indicate that a quick and intuitive placement of connections between concepts captures some inter-concept relations that the participants cannot label. This discovery is able to explain Rothe's and Warning's findings (1991) that individuals have difficulties with consistently assigning labels to edges: Individuals are just not able to name all edges they assume to be present. Labels attached to edges that a person cannot properly name in the first place will be arbitrary and inconsistent.

Nonetheless, these results have to be treated with a certain degree of caution. The employed explicit test mode of the AST forced participants to enter a written label for an edge if placed. This procedure might have been very tiresome, and it is possible that participants did omit labels they could have possibly placed in this mode due to a lack of motivation and/or compliance, which again limits the task performance (as predicted by Goal Theory, compare Locke & Latham, 1990). Although this possibility cannot be ruled out, the magnitude of the observed effect (compare Table 8.3) renders this possibility as the sole explanation improbable, and it is very likely that experiment participants could not label at least some of the edges placed in the quick and intuitive condition. The fact that there are some unnamable edges is not a sufficient condition for the AST's capability to tap into the implicit, but it is a sine qua non and justifies further analyses.

9.1.3 AST sensitivity to learning

The AST is supposed to correspond to learning: If learning takes place, the changes in knowledge organization invoked by the learning process are supposed to be mirrored by the according changes of the AST's parameters (compare Hypothesis 3.9.2).

In the seminar task, four AST-parameters changed significantly over the course of a weekly three-month seminar: The number of associated concepts, the number of edges, the number of clusters, and the relative maximum path length (RMPL). The coefficients for diameter and weighted density did not exhibit significant changes. As theory asserts that academic learning, as occurred in this specific case, is a process of knowledge construction and thus a process of knowledge organization (Phye, 2002), the fact that four of the six AST-elicited parameters changed supports the validity of the assumption that the AST is able to capture changes in structural knowledge. However, the domain of the experiment, an academic seminar, does not suggest that much implicit learning took place, because implicit learning "is the acquisition of knowledge that takes place largely independent of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired" (Reber, 1996, p. 5) – these conditions are hardly met in the context of the experiment. Thus, the absence of significant changes of the diameter and density could be due to their sensitivity to implicit learning, which might not have taken place here. This assumption is in line with a study analyzing changes in knowledge structures in a domain where implicit learning was likely to occur: Schvaneveldt et al. (1985) compared structural knowledge scores of novice and senior fighter pilots. They found significant differences in the number of edges, but contrary to the expected levels, as novice pilots placed more edges than expert pilots between expert-selected stimulus concepts representing typical fighter maneuvers. One possible explanation for this finding is the higher level of automation among expert pilots, which co-occurs with forgetting of knowledge that once was explicit. This observation also might have been caused by the fact that the changes occurring in structural knowledge during implicit learning are not only of a quantitative, but also of a qualitative nature not captured in the employed metrics. Winitzky, Kauchak, and Kelly (1994) found that expert teachers' knowledge did not change quantitatively, but displayed trends towards qualitatively different knowledge structures as compared to novice teachers. However, the fact that the RMPL measure, which is a measure for the organization of the knowledge organization (compare Section 3.9.1), did not change, casts doubts on the latter possibility.

The factor analysis of the AST coefficients (compare Section 8.1.1) justifies the assumption of sensitivity to implicit learning for the weighted density measure as it is the only coefficient with a minor loading on the same factor as those who are undoubtedly connected to explicit knowledge, such as the number of associated concepts. The connection between the diameter and implicit learning is less clear as it is not loading on the two factors that resulted from the factor analysis; its role will be discussed later in this section.

Another reason for the lack of change of these two coefficients might of course lie in their indifference towards learning in general. Whether this is the case can be established by linking them to outside criteria, which will be discussed in Section 9.1.5.

In summary, four of the five AST-coefficients that are thought to capture explicit parts of knowledge (compare Section 8.1.1) display significant changes after an academic seminar. Changes in the construct led to significant changes of the measurements, which indicates that the AST measures what it is supposed to measure.

9.1.4 AST convergent validity

The third experiment aimed at establishing a measure for convergent validity of the AST test system by comparing AST scores, obtained both under the standard condition and under the explicit condition with labeled edges, with the Structure Laying Technique (Scheele & Groeben, 1984, 1988). Hypothesis 3.9.3 assumed that the number of edges, the weighted density, the diameter, and the relative maximum path lengths of SLT-elicited knowledge graphs and AST-elicited knowledge graphs correlate. The results of this experiment are unsuitable to support this hypothesis for the standard AST system due to an insufficient sample size. With only four cases in the subsample containing the participants who worked on the AST in its standard mode, correlations are too unstable. The only correlation that can be interpreted with utmost caution is the one between AST scores aggregated from both test modes (standard and force-labeled edges) and SLT scores. The fact that the number of elicited edges correlates near zero between AST and SLT in the aggregated sample indicates that in term of edges, the AST measures something different than the SLT. However, AST and SLT exhibit a certain level of overlap for the diameter measures, which partly rely on the number of edges for calculation. The density measures of the AST and the SLT also display a medium correlation, but it fails to reach levels of statistical significance. The fact that one coefficient significantly overlaps indicates at least some degree of overlap between the two methods: Little for the number of elicited edges and compactness, medium overlap with regard to the diameter. These results are surprising for two reasons: Firstly, the diameter relies heavily on the number of edges inside the graph, which show no significant overlap. Secondly, the RMPL value does not exhibit significant crorrelations between AST and SLT, but it is also a measure for graph structure that relies heavily on the number of edges, the diameter (which overlaps) and the number of nodes (which is kept constant).

All in all, the results for this experiment are inconsistent due to their internal contradictions and due to the their limited sample size and the resulting questionable correlation coefficients.

9.1.5 AST congruent and discriminant validity

The fourth experiment assessed whether AST-elicited structural knowledge measures obtained for a certain domain correlate with performance measures indicating achievement in that particular domain (Hypothesis 3.9.4), and whether they correlate with verbal IQ scores. Both findings will be discussed in this subsection.

AST and exam grades

In the comparison between the AST and the SLT, the correlations between exam grades on the one side and the relative maximum path length (RMPL) and the number of edges on the other reaches a significant level in the expected direction: More compact networks (indicated by smaller RMPL values) and networks with more edges correlate with higher grades. The number of concepts and the number of clusters reach marginally significant correlations with the exam grade in the expected direction. These four correlations provide a certain support for the AST's congruent validity.

The diameter does not exhibit significant correlations with the grade. On top of that, the weighted density variable correlates contrary to the expected direction (a higher density correlates with a lower grade). At first glance, the latter finding may appear odd, but it is in line with the logical dependencies among the AST coefficients (compare Section 3.9.3): As the weighted density is in an inverse proportional relationship with the number of nodes and edges of a graph (compare Figure 3.7), it *must* exhibit a negative correlation with grades, since the number of concepts and edges exhibit a positive correlation with grades. This finding outlines the dilemma of the logical structure of the AST: a high correlation of explicit measures with a criterion conditions a low correlation of the weighted density with that criterion (and vice-versa).

In this case, the external criterion was of a rather explicit nature, and its correlation with all the four AST-coefficients that load on the explicit factor (see above) speaks in favor of the AST's ability to capture explicit knowledge structures. This finding also demonstrates that implicit structural knowledge scores do not perform well in predicting the performance of tasks that rely heavily on the correct reproduction and codification of explicit knowledge. An experiment where the importance of other forms of knowledge other than declarative knowledge reproduction in an intellectual task plays a role is required to determine whether AST-elicited structural knowledge is of predictive value. It will be discussed in the next section. Up to this point, the ascertainment is that those AST coefficients that exhibit high loadings on the declarative knowledge factor correlate with an outside criterion that measures performance in a task that is based on declarative knowledge. This argument is supported by findings by Rothe and Ceglarek (2006). They found that the number of associated concepts in word association tasks (compare Section 3.3.1) correlate with performance measures in a variety of different tasks that are mainly based on declarative knowledge reproduction.

AST and verbal IQ

The comparison of AST scores with WST-elicited verbal IQ scores reveals overlap between two AST coefficients loading on the explicit factor and WST IQ-scores. The correlation between AST and grades falls below significant levels if controlled for WST scores. Note that the two coefficients that exhibit the highest loadings on the declarative knowledge factor, the number of associated concepts and the number of clusters, also exhibit the highest correlation with WST scores, while the weighted density measure shows the lowest correlation with WST scores. Explicit knowledge and verbal IQ scores are thus highly correlated, AST-elicited structural knowledge that includes unlabeled edges and verbal IQ are not.

In this particular experiment, the AST is unable to provide an additional significant variance explanation of the exam grade beyond the correlation between WST scores and exam grades. This could be due to the characteristics of the employed sample. In their theory of expertise and intelligence, Horn and Masunaga (2006) argued that non-experts' performance is primarily determined by application of abilities from the domain of general intelligence to the current situation, while experts' performance in their field of expertise is largely uncorrelated with scores of general intelligence. The fact that the average grade of the sample was rather poor indicates that the participants are not experts in the surveyed domain. Thus, their performance scores are quite possibly influenced by their general level of intelligence, whose sample mean is one standard deviation below the expected population mean. This indicates that the relatively poor performance in terms of grades might have been completely determined by subjects' below-average intelligence. Horn and Masunaga (2006) also state that the ability to acquire knowledge is highly correlated with cognitive abilities (see also Jäger et al., 1997, in Section 2.5). Taken together, the students in the sample might have had very little acquired knowledge (indicated by poor grades and low IQ scores) and the performance that they did exhibit might have been primarily been a function of their intelligence. Under this condition, it is not surprising that additional explanation of variance delivered by the AST is low.

Generally, an overlap between AST-elicited explicit knowledge scores and intelligence is not surprising, as knowledge and intelligence are overlapping concepts (compare Section 2.5). Reber (1996) argues that this is due to similar underlying processes of learning (knowledge acquisition) and cognitive processes: "The basic principles of cognitive induction and abstraction on one hand and conditioning and associative learning on the other share a common process – the detection of covariation between events" (p. 4).

In summary, the correlation between the AST coefficients that load on the factor explicit knowledge and an external criterion supports the AST's convergent validity in a rather explicit task while discriminant validity has not been fully established due to sample issues. These results so far allow no conclusion with regard to the proposed ability of the AST to elicit implicit knowledge as all knowledge domains in the past four experiments were of a predominantly verbal and explicit kind. However, the findings so far indicate that the AST could be used as a tool for knowledge elicitation and knowledge diagnosis in declarative fields. One could of course argue that the same level of predictive performance can be achieved with a simpler tool such as the WST, but the AST allows a further level of analysis beyond the descriptive level as so far employed in this thesis: It is suited for an analysis of the content and structure of individual associations on a qualitative level: Teachers and trainers can use the tool to assess what concepts their students remember, and how they are interrelated. On top of that, they can obtain parameters that allow a classification of students and performance prediction for explicit tasks. The AST's ability for performance predictions in tasks that rely on implicit knowledge is discussed in the next subsection.

9.1.6 AST-elicited structural knowledge in complex problem solving

After the preliminary tests delivered some indications that the AST could be suitable for knowledge elicitation purposes, its use for performance prediction in a complex problem solving setting is discussed. The integration of findings from Section 8.2 will establish whether the elicitation of unconsciously available structural knowledge is a suitable measure for individual implicit knowledge used in complex problem solving (CPS).

This assumption was tested in a laboratory experiment involving a complex problem-solving task. The first challenge of this experiment lay in the identification of a suitable measure for complex problem solving, as both Funke (1983) and Süß (1996) encountered various methodological issues when trying to use the profit measure provided by the here-employed TAYLORSHOP scenario: Funke was unable to find correlations with the IQ scores he employed, and the scenario was too difficult for Süß's subjects who were unable to achieve a profit in the simulation. In the present study, these issues were not encountered, as the profit measure correlates with BIS-elicited reasoning scores and several participants were able to obtain a profit; the overwhelming majority of participants performed better than they would have performed if they had not made any changes to the system. Two reasons might have played a role in the establishment of suitability of the profit variable: Firstly, participants in the present study were considerably older than the high-school pupils who participated in Süß's experiment. As the development of individual expertise and skills reaches its full blossom in adulthood (Horn & Masunaga, 2006), the older participants might have been more skillful complex problem solvers than the pupils. Secondly, computer experience might have played a smaller role than ten years ago: In Süß's sample of pupils (Süß, 1996), computer experience was a strong predictor for scenario performance, whereas it does not show significant correlations with CPS performance in the present study. This could be due to a generally higher level of computer experience in today's student population.

The second challenge after the establishment of a valid measure for CPS performance lay in the induction of implicit knowledge. The fact that an exposure to the dynamic system during learning led to an increased performance in system control but not to an increased amount of verbalizable knowledge compared to those participants who only learned from texts indicates a dissociation between performance increase and explicit knowledge during practical scenario experience. As such dissociations have been treated as proof for the occurrence of implicit knowledge in system control by Berry and Broadbent (Berry, 1984; Berry & Broadbent, 1984; Berry, 1991; Berry & Broadbent, 1995; Broadbent et al., 1986), its occurrence is taken as such in this experiment.

Since a valid measure of CPS performance and the occurrence of implicit knowledge can be assumed, AST-elicited structural knowledge parameters can be put in relation with the dissociation between performance and verbalizable knowledge. The practical experience with the system during learning did not only lead to implicit knowledge, but also to a higher weighted density of the AST-elicited structural knowledge graph. The fact that weighted density also correlates with complex problem solving performance and provides a small but visible unique contribution to the regression of all measurement variables on complex problem solving performance (compare Section 8.2.6) suggests that the weighted density measure of the AST seems to capture a certain although small amount of implicit knowledge. At the same time, the weighted density does not mirror explicit learning, because it does not correlate with tests for declarative scenario knowledge (compare Table 8.13).

In the only experiment of the thesis where the presence of implicit knowledge can be assumed due to the observation of a dissociation between verbalizable knowledge and performance among experience learners, the weighted density mirrors this dissociation by exhibiting marginally significantly higher values for experience learners. The weighted density measure is at the same time the only AST parameter that correlates with CPS performance and does not correlate with BIS scores for intelligence.

Given the positive correlation between the weighted density measure and CPS performance, the logical structure of the AST coefficients conditions a negative correlation between the profit on one side and the number of graph nodes and edges on the other. The number of nodes does show a negative correlation with the profit, while the number of edges exhibits a near-zero correlation with the profit variable. Both correlations do not reach levels of statistical significance, which is only caused by the small correlation between the weighted density and the profit. If this correlation had been higher, a negative correlation between the edges, the number of nodes, and profit would have been observed.

Despite this problematic relationship between the weighted density and nodes and edges (compare Section 9.1.1), the results from the laboratory experiment support the assumption that the weighted density measure is an indicator for structural knowledge including structural implictness as conceptualized previously. This would explain a number of reported findings in connection with that parameter: The small but present correlations with performance scores in the TAYLORSHOP scenario might be small because implicit knowledge only appears to play a small role in CPS performance in the reported results, and the AST is in turn supposed to only measure a fraction of it. The dissociation between experience and performance is rather small (compare Table 8.12), which could be seen as as an indication for the small importance of implicit knowledge in CPS. Furthermore, the small amount of available time for practice (ten minutes) might have also reduced the amount of available implicit knowledge. This will be discussed further in Section 9.2.3.

If these assertions should hold true, the connection between AST-elicited weighted density scores and performance-relevant implicit access to structural knowledge is very weak. The effect is only visible for one of the two performance indicators of CPS (profit), it is only visible on individual level, the effect size is very small (compare the lasso analysis in Section 8.2.6), and it only occurs in experience learning (not in discussion with experience). The effect of learning by doing on structural knowledge density might be so small that it only occurs in individual practice. In the discussions, group processes could have obstructed this effect. Such an obstruction would be in line with findings by Endres and Putz-Osterloh (1994), where different levels and styles of group interaction in complex problem solving did not have an effect other than an increase of the probability to perform poorer than individuals – an indication that group processes can interfere with problem solving strategies on the individual level.

In summary, one has to ascertain that the link between the AST-elicited weighted density of knowledge graphs and performance-relevant structural knowledge is weak. Whether this weakness is caused by a conceptually weak connection between the AST coefficients and implicit knowledge or by the small role of implicit knowledge in the experiment cannot be determined. However, the research assumption that a fraction of individual implicit knowledge can be psychometrically assessed through structural knowledge elicitation techniques has received some support by the body of evidence.

9.1.7 Overall assessment of AST validity

Despite some encouraging results from the sensitivity experiment (compare Section 9.1.3), the overall validity of the AST as a whole with regard to the general claims of the method (assessment of explicit knowledge and structural implicitness regardless of knowledge domain) can be questioned for two reasons.

Firstly, the four AST coefficients that load the factor explicit knowledge do not show a uniform pattern of significant correlations with measures that elicit explicit knowledge in the TAYLORSHOP experiment. Only the diameter shows a small significant correlation with the WIS_{post} test for declarative scenario knowledge (compare Table 8.13), but the diameter measure is not loading on the explicit knowledge factor to a large extent (compare table Table 8.2). The four AST-coefficients nodes, edges, cluster, and RMPL do not capture specific explicit knowledge in the TAYLORSHOP experiment. The only coefficient that displays a consistent and expected pattern in the context of the TAYLORSHOP experiment is the weighted density measure: it only correlates significantly with scenario performance and might thus actually tap into the structural implicitness. The diameter is the only AST parameter that exhibits a significant correlation with specific declarative knowledge scores in the TAYLORSHOP experiment, yet it is the only AST-parameter that does not load substantially on the two identified factors. It is thus not clear what exactly the diameter measures.

Secondly, the weighted density measure displays a negative correlation with exam grades in the vocational school experiment; poor grades and high densities co-occur. This behavior is rooted in the logical structure of the AST: The significant correlation between the number of nodes and exam grades conditions this counterintuitive negative correlation. This logical connection explains the odd behavior of the weighted density measure in the school experiment quite well, but it also underlines the fact that in this specific context, the weighted density measure is useless.

However, four of the other five AST-variables apart from the weighted density measure displayed significant correlations with exam scores as expected. Thus, where the weighted density variable resulted in useless correlations, the other variables produced reasonable measures, while this pattern was reversed in the TAYLORSHOP experiment (see above).

In summary, the results indicate that the AST is not the general-purpose tool that it was designed to be, but some of its indicators can be of predictive use in specific contexts. The question which variables are useful in what contexts, and, more interestingly, which processes moderate these connections, has been only slightly tapped. The logical structure of the AST's coefficients however leads to a situation where the usefulness of one coefficient, i.e. a positive correlation between explicit knowledge scores and exam grades, conditions the uselessness of another (the weighted density) and vice versa: The weighted density proofed to be a significant predictor in the TAYLORSHOP experiment, where the number of nodes and edges turned out to be without any predictive value. The resolution of this contradiction has to be in the focus of further research. In its present state, the AST may be an imperfect tool, but results are also encouraging.

9.1.8 Methodological issues

The attempt to capture different types of individual knowledge with the AST is not without methodological problems. The small effect of AST-elicited structural knowledge scores on CPS performance is difficult to interpret, because three weak links (or any combination of them) in the underlying argumentation are possible: Firstly, unconscious access to structural knowledge may not be a part of implicit knowledge in the strict sense. Secondly, a method that relies on verbal stimuli may be unable to capture anything that is not explicitly available. Thirdly, implicit knowledge may play a less prominent role in complex problem solving performance than expected. The small effect of AST-elicited structural knowledge scores on CPS performance is thus difficult to interpret.

Furthermore, the AST's internal structure proved to be problematic due to certain logical dependencies as described in the previous sections of this chapter.

Apart from this conceptional problem, there were issues in the planning and conduction of experiments. Throughout all of the validation experiments, very little attention was placed on an analysis of the type of acquired knowledge. None of the experiments took place in a domain where the importance of implicit knowledge was established. Their suitability for the validation of an instrument which is primarily aimed at implicit knowledge elicitation is thus questionable. Participant feedback in these experiments also revealed that the AST is seen as a very tiring procedure, especially the second part of pairwise concept ratings. Due to this negative feedback, the AST was not implemented twice in the complex problem solving experiment, because it was feared that participants would not maintain the level of concentration and motivation necessary for successful AST completion twice during the same experiment. This led to a lack of retest reliability analysis which in turn prevented a monomethod analysis of validity as requested by Campbell and Fiske (1959). On the other hand, individual knowledge is hardly a stable construct over time, and a method for its elicitation will thus hardly display high retest reliability over time.

The poor quality of the sample in the vocational school experiment as outlined in Section 9.1.6 was another unanticipated issue that rendered some results with regard to the validity of the weighted density measure difficult to interpret (compare Section 9.1.7).

Another methodological issue lies in the complex problem solving experiment that aimed at establishing a connection between unconscious access to structural knowledge (structural implicitness) and complex problem solving performance (Hypotheses 4.6.1 – 4.6.3). The AST's weighted density measure is a mixture of both consciously and unconsciously available structural knowledge, as edges with consciously available labels and edges without consciously available edges enter its calculation (compare Section 8.1.2). The two constructs are thus confounded and a thorough test of the hypothesis would have involved a further experimental condition, in which the explicit mode of the AST had been employed. If the explicit weighted density showed no correlation with CPS performance, but the one elicited under standard conditions did, only then one could accept the hypothesis without any doubt. As such a design would have added another factor to the already complex experimental design, this was not feasible in the scope of the study.

The complex and long experimental design also prevented the inclusion of other scales that would have increased variance explanation of complex problem solving. For example, participants' motivation was not surveyed, although it is known to play an important part in task performance in a variety of tasks (Locke & Latham, 1990), also on group level (Schauenburg, 2004).

In the same experiment, two other methodological flaws become apparent. Firstly, the gender of participants was not intentionally balanced over experimental conditions. The fact that gender affects performance could have been anticipated due to earlier findings by Wittman and Süß (Süß, 1996; Wittmann et al., 1996). Although it turned out that male and female participants were not distributed over conditions in a significantly disproportional way, the small effects of structural knowledge on CPS performance could have been influenced by this artifact.

Finally, the available time span of ten minutes for learning and practicing might have been much too short for proper formation of implicit knowledge. Participants in the experience condition had to read the text on successful scenario control and experiment with the scenario during this time. That was probably too little and could have led to contrary effects, as short periods of scenario experience can lead to confusion instead of solid experience (Kluwe, Haider, & Misiak, 1990). This will be discussed further in Section 9.2.3.

9.1.9 Implications for research

The findings on the individual level with regard to implicit and structural knowledge and with regard to complex problem solving can be integrated with several existing bodies of research: With complex problem solving research, with works on structural knowledge, and with works in the field of knowledge and cognitive psychology.

First of all, the results replicate the known findings on the connection between complex problem solving (CPS) performance and its established predictors: CPS performance correlates with reasoning scores, elicited with the Berlin Intelligence Structure Test (BIS), and with scenario-specific declarative knowledge as in other studies (J. Funke, 1992; Kersting, 1999; Süß, 1996). The lasso procedure (compare Section 8.2.6) extended the state of research by delivering a rank order of the independent variables, sorted by relevance. According to those results, scenario-specific knowledge (WIS_{post}) has the strongest influence on CPS performance, followed by reasoning (BIS-K) and the self-assessment of prior economic knowledge. These predictors are followed by a knowledge increase through learning as the next relevant independent variable, which indicates that the ability to acquire new scenario-specific knowledge, the participants' ability to learn, also plays an important role in CPS performance. The weighted density of AST-elicited structural knowledge is the last independent variable to enter the lasso, which indicates that it has the smallest influence on CPS performance among the set of predictors. However, all these predictors together only account for less then a quarter of the overall variance of the dependent variable. This indicates that there are further determinants of complex problem solving abilities beyond the scope of this thesis. The strong tie between computer experience and TAYLORSHOP performance that Süß (1996) observed was not visible in the data. As discussed earlier, the generally high level of computer experience among today's' university students may have evened out CPS-relevant differences present more than ten years ago.

Secondly, the results support the assumption that implicit knowledge plays a part in CPS performance. Broadbent and Berry (1984; 1984) were the first to assume that implicit knowledge is responsible for their observed dissociation between CPS performance increase through scenario experience and no visible increase in verbalizable knowledge. Their argument has been challenged repeatedly (J. Funke, 1995). One of the main criticisms was that their employed systems were simple finite state automata (J. Funke, 2001a) that require only simple strategies for successful control (compare Section 4.2.2). If a small "path" through the several finite stages on the system is sufficient for its successful control, so the argument, than the other states of the system off this path do not need to be examined and no further verbalizable knowledge of the system will develop during its control. Thus, the dissociation between performance and verbalizable knowledge is an effect of the system's structure. The present results did discover a small dissociation between CPS performance and verbalizable knowledge after (little) scenario experience in some conditions. This effect was obtained with the TAYLORSHOP system that cannot be modeled as a finite state automat, but as a linear structural model (J. Funke, 2001a). Thus, dissociation between CPS performance and verbalizable knowledge cannot solely be explained as an artifact of the use of finite state automata systems.

Thirdly, with regard to the concepts from knowledge psychology, the findings provide weak support for one assumption of the dimensional model of knowledge types presented in Chapter 2: It is possible to attribute unconscious access to structural knowledge to the domain of implicit knowledge. Especially the ascertainment that implicit memory is not a different storage in memory, but a specific kind of access to memory contents (Markowitsch, 1999, p. 25) has proven to be viable in the development of the AST and its application in complex problem solving. The fact that the two underlying factors beneath the AST-elicited constructs correlate at a moderate level also supports the notion that implicit and explicit knowledge are intertwined (Nonaka & Konno, 1998).

On top of that, the presented results are in line with some of the aspects of Dienes and Perner's theory of implicit and explicit knowledge (1999). They postulated that the structure between explicit knowledge elements can be implicit and referred to this as structural implicitness. The finding that participants place conceptual links between explicit knowledge elements that they cannot label (compare Section 8.1.2) supports this claim.

Finally, this study provides further support for the usefulness of graph-based structural knowledge approaches in knowledge assessment that were previously proposed by Eckert (1998a) and Bonato (1990). Their results showed moderate success in performance prediction with graph based coefficients on small samples; it has been possible to extend the usefulness of such approaches to a larger setting in the context of the presented experiments.

9.1.10 Implications with regard to practice

Davis, Curtis, and Tschetter (2003) suggested the use of structural assessment techniques for the evaluation of cognitive training outcomes; Beatty and Gerace (2002; 2002) suggested computer-based word association tasks for student assessment, and both studies called for further research in their respective fields. The presented results support these suggestions by providing further proof for the usefulness of both approaches. As the experiments in this thesis have shown, AST-elicited structural assessment parameters correlate with grades among students, and with simulation performance in a staged assessment center (although correlations are small in the latter setting and correlations with verbal IQ scores were observed in the former). Based on these results, the use of the AST in recruitment processes could be considered: It delivers not only quantitative measures describing individual knowledge quantity and organization, but also allows a qualitative view of an individual knowledge structure by providing a unique visualization of someone's knowledge and its internal connectedness. The benefit of the AST test system in this regard is its context-independent structure: The tool remains the same regardless of the knowledge domain in question. With classic pen-and-pencil tests that aim at knowledge assessment, e.g. multiple choice exams, each knowledge domain requires the design of a specific test, which can be costly and time consuming. The AST does not require such efforts and may be employed as a universal test for knowledge structures. This benefit may outweigh the fact that AST administration is a rather time-consuming process, and that participant acceptance of the test procedure during pairwise comparison declines if the test is too long.

Furthermore, the elicitation of implicit constructs, such as personal construct systems and/or mental models is seen as an important task in organizational knowledge management by some researchers (Clases, 2007), as it allows to make views of the organization underlying organizational work practice explicit. The AST could provide another tool for this task, assisting practitioners in a comprehensive diagnosis of the organization's state.

With further efforts in AST design, especially in the development of an analysis module, the AST could be introduced into personnel screening and selection processes, e.g. during assessment centers, given that the abovementioned problems are overcome and the mediating factors that determine the predictive use of AST coefficients in specific contexts are further investigated and understood. Then, the AST could prove to be useful in organizational and instructional practice as computer-based personnel selection methods tend to have high acceptance among participants (Kersting, 1999).

9.2 Laboratory Experiment Results on the Group Level

After the results on the individual level with regard to the AST and the connections between structural knowledge, implicit knowledge and complex problem solving performance have been discussed in the previous section, this section focuses on discussing the results on group level: The connection between the knowledge distribution in the dyad and its complex problem solving performance, and the possibility of predicting group performance based on group self-assessment with the skillMap system. As in the previous section, the results will be integrated across several hypotheses, and issues and implications will be discussed.

9.2.1 Data considerations

The established suitability of the dependent variable profit (compare Section 9.1.6) is even more visible when averaged on the dyadic level (dProfit): It not only exhibits the same features in terms of satisfactory correlations with outside criteria and statisfactory levels of difficulty, but it also displays a normal distribution which allows the use of parametric statistics in its analysis.

Experiment participants experienced the manipulation of knowledge overlap as expected: the Euclidian distances of knowledge-self assessment clearly show the anticipated pattern. The use of the skillMap similarity measure for a treatment check was not advised at this point, because the measure had not been validated so far and relies on the semantic relations between stimulus concepts for calculation.

Self-assessment of expertise can be highly unstable, depending on the context of the self assessment. For example, a meta-analysis of studies on self-assessment of medical doctors revealed “that in a majority of the relevant studies, physicians do not appear to accurately self-assess. Weak or no associations between physicians’ self-rated assessments and external assessments were observed” (D. A. Davis et al., 2006, p. 1100). In the experiment described in Chapter 7, students were required to assess their skill level with regard to their expertise in the TAYLORSHOP scenario. Students exhibit a medium correlation between self assessment capabilities in intellectual tasks and external rating; a meta-analysis of 51 studies comparing students self-assessment with teacher assessment resulted in an average r of .39 (Falchikov & Boud, 1989, p. 420). It was thus assumed that at least for the current experimental setting, an appropriate level of validity could be assumed for self-assessment scales. This turned out to be true and false at the same time: With regard to the calculation of knowledge similarity inside the dyads with the Euclidian distance, the self-assessment proved to be a viable measure, judged by its predicted distribution. However, the aggregated self-assessment score of the dyads did not correlate with scenario performance, and the skillMap similarity measure, which is based on the self-assessment scores, did not exhibit the predicted quadratic relation with CPS performance (compare Section 8.3.7). On the other side, its use in the skillMap similarity calculation delivered valid prognostic results as the skillMap similarity predicted average performance of groups under the different experimental conditions and correlated with CPS performance (more on this in Section 9.2.5).

The finding that the aggregated self-assessment does not exhibit significant correlations with the participants' performance is in contrast with the correlations expected from Falchikov's meta-analysis (1989). A possible explanation for this can be seen in the nature of the tasks: The studies included in the meta analysis surveyed students who made self-assessments on material they had studied for a longer period of time. Adequate self-assessment of knowledge requires a feeling of knowing and a judgment of learning (Nelson, 1996), both of which are more accurate if the learned material resides fully in long-term memory and tends to be biased if learned content is still active in short-term memory (ibid), which might have been the case in this experiment.

9.2.2 Knowledge distribution and group CPS performance

In dyads where each member receives seven learning items with varying overlap, the degree of scenario-specific knowledge overlap inside the dyad exhibits a curvilinear n-shaped (i.e. quadratic) effect on the average dyad's complex problem solving performance. The degree of scenario-specific knowledge overlap exhibits a similar but weaker effect on the average knowledge increase inside the dyad, which correlates with the dyad's pooled performance. The degree of overlap, which is confounded with the number learning items inside the dyad, also displays a linear effect on CPS performance, but the quadratic polynomial effect dominates over the linear effect as visible in the higher effect sizes of the polynomial analyses of variance. The effect of knowledge overlap on performance is n-shaped in the way that partial overlap leads to higher performance values than complete overlap or almost no overlap. Dyads with no overlap containing twelve knowledge elements performed better than fully overlapping dyads containing only seven elements, but fell short of dyads with partially overlapping knowledge containing a total of ten elements. Note that a constant number of knowledge elements were assigned to each individual participant. Thus, the obtained effects can be explained with the observation of a higher knowledge increase inside the dyad through discussion in the partially overlapping condition. It is the degree of heterogeneity among group members that leads to broader knowledge within the group, but an increase in heterogeneity will also ultimately lead to an increase of process losses.

The results show that the average group performance in this experiment is not primarily a function of the amount of knowledge inherent in the dyad. However, the observation that dyads operating under the unshared condition outperformed dyads working under the shared condition indicates that the amount of knowledge contained in the dyad is not negligible. Since the number of knowledge elements in the unshared groups was almost twice the number of elements compared to the groups under the shared condition, this is not surprising. The simultaneous alteration of knowledge overlap and amount of knowledge per dyad does not allow an exact comparison between the strength of these two effects but it seems obvious that the curvilinear overlap effect is stronger because the highest amount of knowledge cannot equalize the n-shaped curvilinear effect in the no overlap versus the partial overlap condition. In numbers: the curvilinear effect of knowledge distribution on CPS performance ($\eta^2 = .080$) is medium ($.05 < \eta^2 < .013$, cf. Cohen, 1988, 1992) and is almost twice as large as the corresponding small linear effect ($\eta^2 = .045$).

For complex problem solving, these results are in line with the findings from the domain of simple problem solving or memory tasks that identified a superiority of partially overlapping items for group performance over full overlap (Ohtsubo, 2005; Tindale & Sheffey, 2002): as for simple problems, a partial overlap of knowledge in the group is superior to no and full overlap. Thus, when discussing complex problem solving strategies, partially overlapping knowledge is superior to fully overlapping knowledge – under the condition that the amount of individual knowledge of group members is roughly equal, as established in Section 8.2.2.

The clear visibility of the n-shaped relationship between knowledge overlap in the group and the group's complex problem solving performance is remarkable for two reasons. First, the scale for knowledge overlap was very small; it only ranged between two and twelve overlapping items. A visible drop in performance between different levels of knowledge overlap might have occurred outside of the scope of the experiment. Secondly, experiment participants were not supposed to combine knowledge from different fields of expertise, but from a common domain (TAYLORSHOP scenario control). This might have created a similarity between knowledge elements conceived as being different from each other in the design of the experiment. The fact that an effect was observed despite these two reasons can be seen as an indication of the robustness of the effect.

However, this effect is only partially visible in average dyadic knowledge increase after discussion (compare Figure 8.8) where the curvilinear effect is only marginally significant. This may be due to a feature of the experiment: In order to maintain comparability between pre- and post-discussion knowledge tests, subjects were not allowed to use their notes in the post-discussion knowledge test, but they were allowed to use their notes in the TAYLORSHOP microworld task. In this way, individual cognitive limitations might have prevented the use of all knowledge exchanged in the discussion in the unshared condition. If this was the case, dyads from the unshared condition should exhibit a high performance when allowed to use their notes. They were allowed to do so during the scenario control. Since – counterintuitively – dyads in the unshared condition performed significantly poorer than those from the shared condition, a performance decrease in the WIS post-discussion test due to the unavailability of notes does not seem to be the case.

9.2.3 Implicit knowledge on group level

The finding that neither trial-and-error experiences during learning nor the availability of a scenario during discussion for strategy evaluation influenced average group scenario performance in any measurable way is not in line with the findings on the individual level (compare Sections 9.1.6 and 9.1.9). The small magnitude of the effect on the individual level probably led to its statistical disappearance in the process of averaging results on dyadic level. The available time for scenario experience during learning and interaction might have been too limited to allow the formation of implicit knowledge to larger extents. Despite the ascertainment that the influence of implicit knowledge on scenario control may be smaller than originally expected (Kluge, 2004), the time issue appears to be of high importance. According to Kluwe, Haider and Misiak (Kluwe et al., 1990), contrary effects – i.e. a performance decrease after experience with a scenario – can occur if the scenario was presented too briefly. As Süß (1996) found effects only for male participants after 30 minutes of scenario exposure, the here-employed interval of ten minutes might have been too short. The significantly decreased amount of verbalizable knowledge in those dyads that learned from texts and from scenario experience during the initial learning period (see Figures 8.10 and 8.11) compared to those who learned from texts only indicates that ten minutes was not enough for text reading and scenario elaboration.

Finally, there are no previous studies that explicitly analyzed implicit knowledge in TAYLORSHOP control. Thus, it may be possible that implicit knowledge plays a smaller role in this paticular microworld than in the control of others. A combination of the above-mentioned aspects may have led to the findings, but a satisfactory answer cannot be drawn from the reported findings.

9.2.4 Gender effects

The effect of gender composition of the dyad on complex problem solving performance is surprising in its magnitude, both on the individual level (compare Section 8.2.7) and when averaged on the group level (compare Section 8.3.8). Gender effects in CPS performance have also been reported elsewhere: Wittman and colleagues (1996) were able to identify higher prior economic knowledge of male participants and their superior computer experience as causes for the effect. In the present study, this does not seem to be the case: Average prior economic knowledge self-assessment and average computer experience self assessment do not show a significant correlation with dyadic CPS performance. Furthermore, a partial correlation of gender and performance, controlled for all other measurement variables, still leads to a significant correlation between gender and CPS performance. Thus, none of the measured variables were able to explain the differences completely. Possible reasons on the individual and the group level, such as influences of personal traits (e. g. Nair & Ramnarayan, 2000) or less concurrence seeking of men (which leads to increased group performance, compare Schulz-Hardt, Jochims, & Frey, 2002; Smith, Petersen, Johnson, & Johnson, 1986) due to more fights for rank order (Buss, 2003), are speculative but worthwhile to follow up. A recent stream of research has identified the so-called stereotype threat as responsible for gender-specific differences in performance in challenging mathematical and/or technical tasks (Martens, Johns, Greenberg, & Schimel, 2006; Marx & Stapel, 2006). The Stereotype-Threat paradigm assumes that members of a subgroup or minority will perform poor in tasks they perceive as especially difficult for members of their respective group, as this perception conditions feelings of anxiety and frustration which again impair task performance (Marx & Stapel, 2006). It appears plausible that the rather technical display of the TAYLORSHOP scenario (compare Figure 7.1) along with its rather technical content (operating machinery) has caused stereotype threat among the female participants. Stereotype threat in complex problem solving might thus be of interest in future research.

9.2.5 skillMap similarity algorithm

The results with reference to complex problem solving and the skillMap similarity measure contain a peculiarity that requires further discussion: The similarity measure was intended to deliver a measure between 0 and 1 increasing in proportion to the similarity between two individuals represented in the skillMap system (compare Section 6.2.2). It was assumed that this similarity measure would display an n-shaped connection with CPS performance (Hypothesis 6.3.1), which was not the case. However, the similarity measure displayed the same distribution over experimental conditions as the CPS performance: Those groups that scored high on complex problem solving performance also exhibit a high similarity measure. This finding is mirrored by the significant correlation between CPS performance and the similarity measure. While the correlation coefficient might be inflated due to some groups that exhibited no similarity at all, the n-shaped distribution of the similarity value over experimental conditions is very clear both in the error bar plot (compare Figure 8.12) and in the following variance analysis with quadratic contrasts.

The similarity measure and CPS performance do not exhibit an n-shaped relation, but a linear relation. This finding implies that groups operating under the shared condition received a higher skillMap similarity measure than those under the fully shared condition. Contrary to that, the Euclidian distance, which does not take the semantic structure of expertise concepts into account, exhibited a linear relation to the experimental condition (compare Section 8.3.3), while the similarity measure exhibits an n-shaped relation. Note that the Euclidian distance and the similarity measure are calculated from the same data (self-assessment scores on ten items with reference to TAYLORSHOP expertise). Thus, the different trajectories displayed by the two measures can only be explained by the underlying calculations. The similarity measure seems to overestimate similarity values at medium overlap while it seems to underestimate similarity values under high overlap when operating on the experimental data. This cannot be due to technical flaws in the calculations, as test calculations (compare Section 6.2.2) did not exhibit any abnormalities. Thus, the reason for the overestimation of similarities and the deviation from the trajectory of the Euclidian distance must be rooted in the semantic structure of the employed expertise graph connecting the skill nodes: A medium overlap in self-assessment led to more paths than equal or different assessments and presumably to inflated similarity values. Thus, for this specific instance, the theory of an n-shaped connection between performance and heterogeneity was "built in" to the skillMap similarity calculation, while it did not occur in the calculation of the Euclidian distance. This finding indicates a conceptional flaw in the similarity measure, because it does not measure similarity of knowledge self-assessment as it is supposed to. Its calculation does not seem to capture similarity in a linear proportional way, but delivers an n-shaped distribution. Unintentionally, it is thus a measure for the mediocracy of overlap; a linear measure for the propitiousness of overlap and thus a linear logarithmic scale for the increasing trajectory of knowledge growth in Scholl's model of team effectiveness (compare Figure 5.1). Thus, higher values of the similarity measure denote a more favorable position towards the maximum of the parabola.

In this way, the skillMap did provide a prediction of group performance on an aggregated level as it predicted which of the three experimental conditions would produce the highest average complex problem solving performance. Thus, the elicitation, visualization and analysis of competencies of group members with an IT system have proven to be of value as they allow performance prediction of groups – at least in the employed context and contrary to the originally assumed way. This is in line with findings from Trier (2005) who also pointed out the use of such systems. His findings are somewhat augmented by the current results because the system he proposed relies on data mining for the establishment of competencies whereas the current project has demonstrated that the use of self-assessments can serve as a viable source for predictions as well.

The proposed similarity measure employs a simple, path-based similarity measure. In graph theory, much more sophisticated ways of similarity identification based on structural similarities exist (e. g. Le et al., 2003). It would be desirable to develop more accurate similarity measures for future applications that rely on such algorithms that take further structural information into account.

It is however questionable whether the findings from the present study can be generalized to domains other than complex problem solving. The employed skillMap uses a very task-specific vocabulary, which might not necessarily be present in other tasks. Generally, the quality of prediction relies on the accuracy of self-assessment and on the accuracy and precision of the underlying semantic structure. In the presented experiment, the experimenter specified the latter. It is however possible to attribute the task of generating a semantic structure of expertise or knowledge elements to the users of such a system, as proposed in the skillMap architecture (compare Section 6.1.6). Such jointly user-generated structures are referred to as folksonomies (Catutto et al., in print) and have proven to be both highly accurate and practical (ibid). However, one has to bear in mind that this accuracy has not been tested in the present experiment.

The performance prediction that the skillMap delivered is also a further indicator of the usefulness of graph-based knowledge management systems aiming at facilitating linkage between knowledge elements (Trier, 2005). In the broader perspective, the success of such systems, which all include a social networking feature and belong to the category of third generation knowledge management systems (Schütt, 2003), is in line with findings from a Delphi-study on the future of knowledge management (Scholl, König, Meyer, & Heisig, 2004) that predicted a success of such systems.

Another issue challenging the generalizability of the laboratory experiment's results is the laboratory setting itself: Skill self-assessment and problem solving took place under artificial conditions during a very short time span, which is especially unlikely in real-world project work. Therefore, further field studies need to establish the accuracy of the employed similarity measure. Apart from the conceptional and practical issues, the similarity algorithm was able to deliver useful predictions based on these self-assessments. This encourages further studies.

9.2.6 Methodological issues

On the group level, the study faces some methodological issues that should not remain unaddressed. Firstly, the findings on group level with regard to the influence of knowledge distribution on group performance can be questioned due to the fact that the nominal group technique was employed. Measures for group performance consisted of average individual scores of participants who had engaged in a group task, the discussion, but who had performed the CPS task on their own. The group was thus not totally artificial, but it is questionable whether an average score of coacting individuals represents the same as a score of a task that was jointly carried out by all group members together. Examples from research on brainstorming (Diehl & Stroebe, 1987; Gallupe et al., 1992) indicate that nominal group performance and actual group performance can differ and further research has to establish whether the identified results hold also true for joint acting on complex problems. Nominal group technique however had to be employed in this particular setting, as specific hypotheses on the influence of group information exchange processes on group performance had been put forward. It was thus necessary to separate information exchange and task performance. The observed strong effect of information overlap on group performance justifies this procedure, but future studies should avoid such artificial separation of group processes.

The analysis of hypotheses on the individual and the group level in the same experiment is another issue faced in the study. Although the experiment was designed in a way that would keep hypotheses on the two levels separate, the discovery of a few nonindependencies between individual and group measures (compare Section 8.2.3) shows that embedding individuals in discussing dyads did not leave the results on the individual level untouched. Despite the fact that individual scores cannot be exclusively represented as effects on individual level, the subtraction of group level effects shows that the remaining individual-level effects are in the same scope as the non-adjusted values (compare Table 8.10).

9.2.7 Implications with regard to research

Despite the above issues, this study indicates that group performance in complex problem solving varies with knowledge overlap inside the group and that a partial overlap is more beneficial for group performance than full or no overlap. This corroborates Scholl's theory of team effectiveness (1996, 2005) in that respect.

The results of the laboratory experiment have implications for research on team mental models (compare Section 5.2.2). In an experiment by Mathieu et al., dyads had to perform a computer-simulated air-combat simulation that classifies as a complex problem. At several points in time, participants' task mental model was assessed through similarity ratings of expert-identified concepts representing task knowledge (note that this operationalization is almost identical to the operationalization of structural knowledge as employed in this thesis). The similarity of the group members' structural knowledge was assessed with a simple graph-theoretic measure; contrary to the authors' expectation, it did not exhibit a significant positive correlation with task performance.

This finding can be explained by the results presented in this study: The authors upheld the implicit underlying hypothesis that cognitive homogeneity, i.e. homogeneity of structural task knowledge and task performance in a computer-simulated complex task, displays a linear relationship, whereas the results presented in the previous chapter suggest that an n-shaped relation exists between the two. Generally, the findings presented in the previous chapter contradict the notion of a linear relation between task mental model homogeneity and team performance. Instead, they can be seen as an answer to van Knippenberg's (2007) request for further studies on curvilinear effects in (informational) group diversity.

9.2.8 Implications with regard to practice

This study indicates that with regard to the influence of knowledge, group performance in complex problem solving varies with knowledge overlap inside the group and that a partial overlap is more beneficial for group performance than full or no overlap. This has practical implications for the assembly of groups and teams in organizational practice that are supposed to deal with complex problems: If the team members are from similar domains and are thought to possess a comparable magnitude of individual expertise, e.g. due to similar length of employment or similar careers, a partial overlap of knowledge and expertise among team members will be likely to produce better results than teams of persons with identical backgrounds and teams with greater differences in individual knowledge. If a project should require participants with very different backgrounds, a larger time horizon for seeking mutual understanding could be helpful as well as a facilitator for increasing effective exchange (Schimansky, 2006).

The skillMap could prove to be a valuable tool for organizational practice and offers various uses in knowledge management. For example, a human resource department could use the similarity measure in the process of assembling project teams or work groups for certain complex tasks. Since "work groups are increasingly used to complete a variety of organizational tasks"(Levesque, Wilson, & Wholey, 2001, p. 135), this could be of value in management. The system could also be used to visualize the organizational context of employees, which could potentially increase understanding and knowledge exchange among organizational members, because context awareness is a key factor for successful knowledge sharing (Nonaka & Konno, 1998).

Furthermore, the results of the similarity identification could be extended to other practical uses. For example, groups with a high average similarity measure and strong ties among all members who originate from different organizational departments or teams could be automatically classified as communities of practice. Since the leverage of such communities in organization is seen as a key strategic advantage (Saint-Onge & Wallace, 2003), their identification and support with automatic means could lead to business advantages.

9.3 Outlook

The attempt to construct a computer-based test system for the elicitation of individual structural knowledge that also captures structural implicitness, the AST, can be viewed as partially successful. Although AST-scores correlate with knowledge-related performance indices such as as exam grades and complex problem solving performance in certain contexts, these correlations do not appear to apply universally. In explicit knowledge domains, some measures correlate better than others, wheras in complex problem solving, only one measure delivered small correlations with only one of two available outside criteria. The results indicate that knowledge elicitation with a context-free structural elicitation technique is possible, but many questions remain open that require further research.

First of all, the predictiveness of the different AST measures has to be clarified; the validity of the AST is not fully established and further experiments appear necessary in order to assess the stability of knowledge structures over a certain period of time. Although knowledge cannot be seen as a stable trait like personality dimensions, a certain amount of stability between two measurements can be expected if no learning or forgetting has taken place. Re-test assessments need to be performed in order to establish retest reliability.

Secondly, as mentioned earlier, the AST can only be viewed as measurig implicit knowledge in CPS if explicitely available edges are excluded from the set of edges contained in the elicited PFNET. The results reported in this thesis contained a mixure of labeled and unlabeled edges. A further experiment, contrasting standard AST-scores with labeled AST scores in CPS performance, is required to fully assess the role of implicitly available edges in complex problem solving and their effect.

The fact that correlations between the AST weighted density measure and CPS performance were dominated by dyad-level (shared) effects requires further research on the individual level. A replication of the AST-experiment in complex problem solving with a strict individual design, two instances of AST-use (for retest assessment) and the contrasting of the unlabeled with the force-labeled AST version (in order to subtract explicit structural knowledge from the elicited mixture of structural implicitness and explict structural knowledge) could address all above-mentioned issues.

With regard to results on the group level, the degree of scenario-specific knowledge overlap in dyads in which a constant number of knowledge elements was assigned to each individual participant had a larger effect on the average dyad members' performance in a complex scenario than the amount of knowledge inside the dyad. The effect of knowledge overlap on performance is n-shaped (cf. 9.2.2); it seems to be partially due to a higher knowledge increase inside the dyad through discussion in the partially overlapping condition. For the domain of complex problem solving, these results do not contradict the findings from studies from the domain of simple problem solving or memory tasks that identified a superiority of partially overlapping items for group performance (Ohtsubo, 2005; Tindale & Sheffey, 2002). Further studies can now analyze whether there exists a similar connection between knowledge overlap and performance in complex problem solving as predicted in Zajonc's model of group performance (Zajonc & Smoke, 1959). The role of gender effects and the role of implicit knowledge also require further analyses, as the role of implicit knowledge and the mechanisms underlying the observed gender differences in performance remained unclear in this study.

As team mental model research assumes that certain aspects of team mental models are of a procedural type which partly belongs to the category of non-explicit knowledge in the conceptual model presented in Chapter 2, it can be reasoned that team mental models include implicit aspects that are partly measured with the AST. It would thus be a promising endeavor to try and assess team mental models with the AST, because AST-elicited models would be broader than those elicited by Mathieu et al. (2000) as they include less consciously available concept relations. Predictions of team performance could then incorporate the findings from the group level by assuming an n-shaped relation between group member homogeneity and team performance. If the homogeneity was to be accessed with the AST instead of the skillMap, this requires the development and/or employment of more complex similarity measures than those employed in Pathfinder scaling, because AST-elicited networks differ in number of nods, node labels, and structure, while Pathfinder-elicited networks differ in structure only. The use of such similarity measures to AST-elicited graphs and their application to shared team mental models is thus one of the promising future projects.

Another possible continuation of the studies presented in this thesis lies in the conversion of the two separate computer tools that were presented (AST and skillMap). The AST elicits individual knowledge elements via term association and concept comparisons. The skillMap visualizes and analyzes the skills and knowledge of several individuals and connects individuals and knowledge elements. However, the assignment of knowledge elements to individuals is purely based on subjective self-assessment; a method delivering results of varying accuracy (see above). These self-assessments could be replaced by AST-elicited concepts and the skillMap similarity algorithm could be extended to calculating the similarity between AST-elicited knowledge structures, combining the two methods into one powerful device for knowledge analysis, visualization, and management on both the individual and the group level.

9.4 Summary

This thesis has dealt with the concept of knowledge, its forms, its representation, possibilities of its elicitation on the individual and the group level, and with its influence on both individual and group performance in complex problem solving.

In the first and second chapter, the importance of knowledge for todays' complex economies has been outlined, as knowledge creation and exchange are seen as key drivers for organizational innovation, which drives organizational success. Based on this ascertainment, the concept of knowledge was analyzed, shaped and defined. Broadly, it can be conceptualized as structural connectivity patters whose contents have proven to be viable for the attainment of certain goals. On a more specific level, findings from the field of memory science and cognitive psychology were linked, which resulted in a two-dimensional model of knowledge types with the dimensions consciousness of use and 'codifiability'. Based on this model, it was argued that structural implicitness, the implicit connection between explicit concepts, is a part of individual implicit knowledge, which is seen as an important factor in organizational knowledge creation by some scholars. The second chapter concluded with the research assumption that the elicitation of individual structural knowledge could provide an opportunity to psychometrically assess a portion of individual structural knowledge, which could in turn be employed for performance prediction.

The third chapter discussed multiple methods for structural knowledge elicitation. One technique, Pathfinder-scaled networks, was assessed to be of high usefulness for structural knowledge assessment, but it was also argued that this method contains two weaknesses: It relies on expert referent structures of questionable validity and presents participants with concepts that the participants are supposed to 'know' without checking this assumption. Thus, a new computer-based structural knowledge assessment technique is presented: The Association Structure Test (AST). It employes participants' free term associations to a stimulus concept, recording of thinking times between associations, and quick, intuitive, and unlabeled similarity ratings between previously associated concepts for the construction of a Pathfinder-adjusted minimum spanning tree of the participants' knowledge structure. Each participant's knowledge structure is quantified with six graph-theoretic variables, thought to represent different aspects of explicit knowledge and structural implicitness.

It was then argued that the AST must be employed in a task domain where implicit knowledge acquisition and application is important for task performance in order to assess its usefulness for performance prediction. One area of problem solving research in which the role of implicit knowledge has been debated extensively is complex problem solving (CPS), which was introduced in the fourth chapter, along with established determinants of CPS performance. Complex problems are characterized by their ill-definedness (the goal state of the system is not clear), their complexity and connectedness (different interconnected variables need to be considered), their opaqueness (variable interconnections are unclear), and their dynamics (the system autonomously changes over time). Real-world problems faced by managers and politicians are classified as complex problems. In laboratory research, complex problems are modeled with dynamic computer scenarios that adhere to the above criteria. General knowledge, scenario-specific knowledge and intelligence, especially reasoning, are known to determine individual complex problem solving, and one strand of research also sees implicit knowledge as an important determinant in complex problem solving, and CPS was thus identified as task domain for AST use.

The fifth chapter shifted the focus from the individual level to the group level. It was argued that the assignment of complex problems to groups is seen as a way to overcome individual information processing limitations in today's knowledge-intensive organizations. Research on information exchange among groups and on team mental models has come up with two competing theories on how knowledge distribution in groups effects the groups' general performance. Information exchange researchers claim that a certain degree of cognitive heterogeneity increases team performance, wheras team mental model researchers advocate cognitive homogeneity as a success factor for group performance. The effect of cognitive heterogeneity in groups on group complex problem solving performance is thus not clear. However, a specific model of team effectiveness assumes a quadratic relation between group heterogeneity and group performance and it was thus argued that such an n-shaped relationship will be discovered in a complex problem solving experiment.

The sixth chapter described the skillMap IT-system for organizational knowledge management that stores and visualizes employees' knowledge self-assessment and the semantic structure of the knowledge elements (referred to as skills). For this system, a similarity algorithm was suggested that calculates knowledge similarity (i.e., the level of homogeneity of knowledge self-assessment) that takes the semantic structure between knowledge elements into account. The algorithm is supposed to deliver a discrete measure of group cognitive heterogeneity, based on self-assessment.

The seventh chapter presented a methodology for testing the various hypotheses that were put forward in the preceding chapters: Four small experiments aimed at validating some of the AST's central claims, and a laboratory experiment aimed at testing whether (a) individual implicit knowledge plays a part in CPS, operationalized with the TAYLORSHOP scenario, (b) AST-elicited structural knowledge predicts CPS performance, (c) group knowledge overlap and group CPS performance exhibit an n-shaped quadratic relation, and (d) the skillMap's similarity measure also displays a curvilinear relationship with group CPS performance.

The results as presented in the eighth chapter indicated that the AST does capture concept-relations that participants are unable to label explicitely, supporting the research assumption put forward in the second chapter. Some of the AST's coefficients exhibit sensitivity to learning, and four AST coefficients correlate with a knowledge-based outside performance criteria (exam grades). These results appear to be somewhat unstable, as AST scores also correlate with IQ-scres elicited with the WST vocabulary test. The AST's weighted density parameter, which is thought to capture structural implicitness to the strongest extent, loads a different factor than the other explicit-knowledge-based AST-coefficients, supporting the original claim that structural implicitness correlates with CPS performance.

In the CPS experiment, a slight dissociation between CPS performance increase through scenario experience and CPS performance indicates a (small) influence of AST-elicited structural implicitness on CPS performance, but the connection is weak: It only occurs under certain conditions and is only visible on the individual level, but not on the group level. Thus, implicit knowledge plays a small role in TAYLORSHOP scenario performance and its mode of action seems to be influenced by situational and contextual factors. AST-elicited weighted density measures correlate weakly with complex problem solving performance indicators. However, the AST-coefficients that were thought to capture explicit knowledge do not correlate with measures of explicit CPS scenario knowledge, the influence of the weighted density on CPS performance is primarily due to dyad-level (shared) effects and probably influenced by gender effects. In summary, the results indicate that implicit knowledge assessment is theoretically possible under certain conditions, but most probably not in the general context free way as intended. Further research appears necessary.

In order to test the hypotheses with regard to the influence of knowledge distribution in groups on group CPS performance, the score of coacting dyads that stemmed from the same experiment as the AST results were obtained. The dyads learned and worked on the CPS scenario under three different conditions with regard to knowledge overlap: In the overlapping condition, both group members received the same information on scenario control, in the partially shared condition, they received partially overlapping information and completely divergent information in the non-overlap condition. Participants' individual cognitive load was equal under all conditions. As expected, dyads working under the partially shared condition outperformed the other dyads, both in terms of CPS performance and in knowledge acquisition. Contrary to expectations, the skillMap similarity measure did not exhibit a curvilinear relation with group performance, but a linear one. It turned out that proposed the similarity measure has a built-in bias delivering increased similarity values for medium-similar persons. The results indicate that performance-predictions with IT-systems are generally possible, but the employed similarity measure requires further revision before practical implementation. The curvilinear relationship between knowledge overlap in a dyad and the dyad's complex problem solving performance is probably the most resilient finding of this thesis. It has both theoretical and practical implications: With regard to shared information research, it extends known findings on the curvilinear relation between group information overlap and group performance for simple problems to complex problem solving. The findings of this thesis can also explain why in team mental model research, cognitive homogeneity did not correlate with team performance in some studies. With regard to organizational practice, the findings indicate that project teams that have to tackle real complex problems will probably be most effective if a certain amount of knowledge heterogeneity is present among team members. Further research appears necessary in order to extend the current findings from the laboratory into the organizational environment, and to further clarify the relationship between the amount of knowledge inside a group on the one hand and knowledge overlap on the other hand on group performance, as the amount of knowledge inside the group was not completely controlled in the conducted experiment.

Bibliography

Aamodt, A., & Nygard, M. (1995). Different roles and mutual dependencies of data, information, and knowledge – an AI perspective on their integration. *Data and Knowledge Engineering*, *16*, 191–222.

Abdi, H. (2007). Metric multidimensional scaling. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics.* (pp. 598–605). Thousand Oaks (CA): Sage.

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*, *86*(2), 303–311.

Alavi, M., & Leidner, D. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107–136.

Amelingmeyer, J. (2004). *Wissensmanagement: Analyse und Gestaltung der Wissensbasis von Unternehmen* [Knowledge management: Analysis and design of the knowledge base of organizations] (3rd ed.). Wiesbaden: Dt. Univ.-Verl.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R. (1983). A speading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295.

Anderson, J. R., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1050.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *The Journal of Statistical Computation and Simulation*, *62*, 271–303.

Argyris, C., & Schön, D. A. (1996). *Organizational learning II.* Reading, MA: Addison-Wesley.

Bales, R. F., & Strodtbeck, F. L. (1951). Phases in group problem-solving. *Journal of Abnormal and Social Psychology*, *46*, 485–495.

Ballsteadt, S.-P., & Mandl, H. (1985). *Diagnosis of knowledge structures in text learning* (Tech. Rep.). Tübingen: Deutsches Institut für Fernstudien an der Universität Tübingen, Forschungsbericht Nr. 37.

Beatty, I. D. (2000). *Conmap: Investigating new computer-based approaches to assessing conceptual knowledge structure in physics.* Unpublished doctoral dissertation, University of Massachusetts.

Beatty, I. D., & Gerace, W. J. (2002). Probing physics students' conceptual knowledge structures through term association. *American Journal of Physics*, *70*(7), 750–758.

Beatty, I. D., Gerace, W. J., & Dufresne, R. J. (2002). *Measuring and modeling physics students' conceptual knowledge structures through term association times.* Online paper. Retrieved 2008-05-21, from `http://arxiv.org/pdf/physics/0702111v1`

Beckmann, J. F., & Guthke, J. (1995). Complex problem solving, intelligence, and learning ability. In P. A. Frensch & J. Funke (Eds.), *Complx problem solving: the european perspective.* Hillsdale, NJ: Lawrence Erlbaum.

Benta, M. I. (2003). *AGNA (Applied Graph & Network Analysis).* Computer Software. Department of Sociology, University College Cork, Ireland. Retrieved 2007-12-04, from `http://www.geocities.com/imbenta/agna/`

Berry, D. C. (1984). *Implicit and explicit knowledge in the control of complex systems.* Unpublished doctoral dissertation, University of Oxford, United Kingdom.

Berry, D. C. (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology*, *43*, 881–906.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *36A*, 209–231.

Berry, D. C., & Broadbent, D. E. (1995). Implicit learning in the control of complex systems. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving - the european perspective* (pp. 131–150). Hillsdale, NJ: Lawrence Erlbaum.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bliese, P. D. (2006). *multilevel: Multilevel functions. R package version*

2.2. Computer software online manual. Retrieved 2007-12-04, from `http://cran.r-project.org/doc/packages/multilevel.pdf`

Bliese, P. D., & Halverson, R. R. (1998). Group size and measures of group-level properties: An examination of eta-squared and icc values. *Journal of Management*, *24*, 157–172.

Bonato, M. (1990). *Wissensstrukturierung mittels Struktur-Lege-Techniken: eine graphentheoretische Analyse von Wissensnetzen* [Knowledge structurig with structure-laying-techniques: A graph-theoretical analysis of knowledge networks]. Frankfurt am Main; New York: Peter Lang.

Bonett, D. G., & Wright, T. A. (2000). Sample size requrements for estimating pearson, kendall and spearman correlations. *Psychometrika*, *65*(1), 23–28.

Boose, J. H. (1985). A knowledge acquisition program for expert systems based on personal construct psychology. *International Journal of Man-Machine Studies*, *23*, 495 - 525.

Boose, J. H., & Bradshaw, J. M. (1987). Expertise transfer and complex problems: Using aquinas as a knowledge-acquisition workbench for knowledge-based systems. *International Journal of Man-Machine Studies*, *26*, 3–28.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet: Software for social network analysis.* Computer Software. Analytic Technologies.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*(6), 807–842.

Broadbent, D. E., FitzGerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, *77*, 33–50.

Buchner, A., Funke, J., & Berry, D. C. (1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks? *Quarterly Journal of Experimental Psychology*, *48A*, 166–187.

Busch, P. A., Richards, D., & Dampney, C. K. (2001). Visual mapping of articulable tacit knowledge. In P. Eades & T. Pattison (Eds.), *Australian symposium on information visualisation, (invis.au 2001)* (Vol. 9, pp. 37–47). Sydney, Australia: ACS.

Bush, A. A., & Tiwana, A. (2005). Designing sticky knowledge networks. *Communications of the ACM*, *48*(5), 67–71.

Buss, D. M. (2003). *Evolutionary psychology: The new science of the mind* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitraitmultimethod matrix. *Psychological Bulletin*, *56*(2), 81–96.

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., & Eskridge, T. (2004). Cmaptools: A knowledge modeling and sharing environment. In A. J. Cañas, J. D. Novak, & F. M. Gonzalez (Eds.), *Concept maps: Theory, methodology, technology. proceedings of the first international conference on concept mapping* (Vol. I, pp. 125–133). Pamplona, Spain: Universidad Publica de Navarra.

Catutto, C., Schmitz, C., Baldassarri, A., Servedio, V. D. P., Loreto, V., Hotho, A., et al. (in print). Network properties of folksonomies. *AI Communications Journal, Special Issue: Network Analysis in Natural Sciences and Engineering.*

Ceglarek, P. (forthcoming). *Der Wortassoziationsversuch als wissensdiagnostisches Instrument im arbeitspsychologischen Kontext* [The word-association experiment as an instrument for knowledge diagnosis in the context of work psychology]. Unpublished doctoral dissertation, Institute of Psychology, University of Potsdam, Germany.

Ceglarek, P., Meyer, B., Kermas, M., & Rothe, H.-J. (2006). *Verfahren zur Diagnose von (strukturellem) Wissem im Vergleich* [Comparison of methods for (structural) knowledge assessment]. Paper presented at the 48. Tagung experimentell arbeitender Psychologen (TEAP), 26.-29.03.2006, Universität Mainz.

Chisholm, R. M. (1957). *Perceiving: A philosophical study.* Ithaca, NY: Cornell University Press.

Clases, C. (2007). Implicit knowledge in personal construct systems. In N. Gronau (Ed.), *Proceedings of the 4th conference on professional knowledge management, 28th-30th March, 2007* (Vol. 2, pp. 285–292). Berlin: GITO.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cooke, N. J., Salas, E., & Stout, R. (2000). Measuring team knowledge. *Human Factors*, *42*, 151–173.

Cross, R., & Parker, A. (2004). *The hidden power of social networks.* Cambridge, MA: Harvard Business School Press.

Damerow, P., & Lefèvre, W. (1998). Wissenssysteme im geschichtlichen Wandel [Knowledge systems in historic change]. In F. Klix & H. Spada (Eds.), *Enzyklopädie der Psychologie* (Vol. 6: Kognition, p. 77 - 114). Göttingen: Hogrefe.

Dansereau, F., Alutto, J. A., & Yammarino, F. J. (1984). *Theory testing in organizational behavior: The varient approach.* Englewood Cliffs, NJ: Prentice-Hall.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organi-*

zations manage what they know. Cambridge, MA: Harvard Business School Press.

Davis, D. A., Mazmanian, P. E., Fordis, M., Harrison, R. van, Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*, *296*(9), 1094–1102. Available from `http://jama.ama-assn.org/cgi/reprint/296/9/1094`

Davis, J. H. (1980). Group decision and procedural justice. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 157–230). Hillsdale, NJ: Lawrence Erlbaum.

Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating cognitive training outcomes: validity and utility of structural knowledge assessment. *Journal of Business and Psychology*, *18*(2), 191–206.

Day, E. A., Jr, W. A., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, *86*(5), 1022–1033.

Dearholt, D., & Schvaneveldt, R. (1990). Properties of pathfinder networks. In R. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 1–30). Norwood, NJ: Ablex.

Despres, C., & Chauvel, D. (Eds.). (2000). *Knowledge horizons*. Boston: Butterworth-Heinemann.

Dick, M., & Wehner, T. (2002). Wissensmanagement zur Einführung: Bedeutung, Definition, Konzepte [Introduction to knowledge management: Relevance, definitions, concepts]. In W. Lüthy, E. Voit, & T. Wehner (Eds.), *Wissensmanagement - praxis* (pp. 7–27). Zürich: vdf Hochschulverl. an der ETH.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*, 497–509.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735–808. Available from `http://www.acsu.buffalo.edu/~emiii/target.pdf`

Dorfman, J., Shames, V. A., & Kihlstrom, J. F. (1996). Intuition, incubation, and insight: implicit cognition in problem solving. In G. Underwood (Ed.), *Implicit cognition* (pp. 257–298). Oxford: Oxford University Press.

Dörner, D. (1979). Kognitive Merkmale erfolgreicher und erfolgloser Problemlöser beim Umgang mit sehr komplexen Systemen [Cognitive features of successful and unsuccessful problem solvers in dealing with very complex problems]. In H. Ueckert & D. Rhenius (Eds.), *Komplexe menschliche Informationsverarbeitung* (pp. 185–196). Bern: Hans Huber.

Dörner, D. (1980). On the difficulty people have in dealing with complexity. *Simulation & Games*, *11*, 87–106.

Dörner, D. (1987). *Problemlösen als Informationsverarbeitung* [Problem solving as information processing] (3rd ed.). Stuttgart: Kohlhammer.

Dörner, D. (1994). Problemlösen und Gedächtnis [Problem solving and memory]. In D. Dörner & E. van der Meer (Eds.), *Das Gedächtnis* (pp. 295–320). Göttingen: Hogrefe.

Dörner, D., & Kaminski, G. (1987). *Gewusst wie : Handeln, Problemlösen, Entscheiden* [Know how: Acting, problem solving, decision making]. Radio Program: Funkkolleg Psychobiologie: Verhalten bei Mensch und Tier / Veranst.: Westdt. Rundfunk. Wiss. Leitung: Klaus Immelmann; Kollegstunde 17. Köln: Westdeutscher Rundfunk.

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Komplexität* [Lohhausen. On dealing with uncertainty and complexity]. Bern: Huber.

Drucker, P. (1993). *Post-capitalist society.* New York: Econ.

Eckert, A. (1998a). Die "Netzwerk Elaborierungs Technik (NET)" - Ein computerunterstütztes wissensdiagnostisches Instrumentarium [The "Network Elaboration Technique (NET)" - A computer-supported knowledge diagnosis instrument]. *Diagnostica*, *44*(4), 220–224.

Eckert, A. (1998b). *Kognition und Wissensdiagnose: die Entwicklung und empirische Überprüfung des computerunterstützten wissensdiagnostischen Instrumentariums Netzwerk-Elaborierungs-Technik (NET)* [Cognition and knowledge diagnosis: The development and empiric analysis of the computer-supported method for knowledge diagnosis Network-Elaboration-Technique (NET)]. Lengerich: Pabst Science Publishers.

Eigler, G. (Ed.). (1990). *Plato: Werke in acht Bänden. Bd. 6: Theaitetos, Der Sophist, Der Staatsmann, griech./dt., bearb. von P. Staudacher* [Plato: Works in eight volumes. Vol. 6: Theaitetos, The Sophist, The Statesman, Greek/German editing by P. Staudacher] (Vol. 6). Darmstadt: Wissenschaftliche Buchgesellschaft.

Endres, J., & Putz-Osterloh, W. (1994). Komplexes Problemlösen in Kleingruppen: Effekte des Vorwissens, der Gruppenstruktur und der Gruppeninteraktion [Complex problem solving in small groups: Effects of previous knowledge, of group structure, and of group interaction]. *Zeitschrift für Sozialpsychologie*, *25*(1), 54–70.

Engelkamp, J. (1991). *Das menschliche Gedächtnis: das Erinnern von Sprache, Bildern und Handlungen* [Human memory: Remembering speech, pictures, and actions] (2nd ed.). Göttingen: Hogrefe.

Engelkamp, J. (1998). *Memory for actions.* Hove: Psychology Press.

Engelkamp, J., & Pechmann, T. (Eds.). (1993). *Mentale Repräsentation* [Mental representations]. Bern: Huber.

Engelkamp, J., & Zimmer, H. D. (1994). *Human memory : a multimodal approach.* Seattle: Hogrefe & Huber.

European Council. (2000). *Lisbon European Council 23 and 24 March 2000: Presidency Conclusions.* Online. Retrieved 2003-09-24, from `http://www.europarl.eu.int/summits/lis1_en.htm`

European Foundation for the Improvement of Living and Working Conditions. (2004). *Advancement of the knowledge society.* Online report. Retrieved 2004-06-04, from `http://www.eurofound.eu.int/publications/files/EF0409EN.pdf`

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), pp. 395-430.

Fillenbaum, S., & Rapoport, A. (1971). *Structures in the subjective lexicon.* New York: Academic Press.

Fisher, R. (1935). *Design of experiments.* Edinburgh: Oliver and Boyd.

Fitts, P. M., & Posner, M. I. (1979). *Human performance.* Westport, CT: Greenwood.

Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW). (2001). *Management von nicht-explizitem Wissen: noch mehr von der Natur lernen* [Management of non-explicit knowledge: Learning even more from nature]. Online report. Retrieved 2005-05-20, from `http://www.faw.uni-ulm.de/deutsch/publikationen/bmbf\discretionary{-}{}{}studie/BMBF_Studie_Teil_1.pdf`

Forsyth, D. R. (1990). *Group dynamics* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Forsyth, D. R. (2006). *Group dynamics* (4th ed.). Belmont, CA: Thomson/Wadsworth.

Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: the european perspective* (pp. 4–25). Hillsdale, NJ: Lawrence Erlbaum.

Fromm, M. (2004). *Introduction to the repertory grid interview.* Chichester: Wiley.

Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? [A few comments on the problems in problem solving research or: Is test intelligence a predictor?] *Diagnostica*, *29*, 283–302.

Funke, J. (1986). *Komplexes Problemlösen* [Complex problem solving]. Berlin: Springer.

Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung* [Knowledge on dynamic systems: Acquoisition, representation, and use]. Berlin: Springer.

Funke, J. (1995). Experimental research on complex problem solving. In P. A. Frensch (Ed.), *Complex problem solving : the european perspective* (pp. 243–268). Hillsdale, NJ: Erlbaum.

Funke, J. (2001a). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, *7*(1), 69–89.

Funke, J. (2001b). Neue Verfahren zur Erfassung intelligenten Umgangs mit komplexen und dynamischen Anforderungen [New methods for registering intelligent dealing with complex and dynamic requirements]. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung* (pp. 89–109). Lengerich: Pabst Science Publishers.

Funke, J. (2004). Komplexes Problemlösen: Möglichkeiten deduktivistischen Vorgehens [Complex problem solving: Possibilities of deductivistic approaches]. In E. Erdfelder & J. Funke (Eds.), *Allgemeine Psychologie und deduktivistische Methodologie* (pp. 281–300). Göttingen: Vandenhoeck & Ruprecht.

Funke, U. (1993). Computergestützte Eignungsdiagnostik mit komplexen dynamischen Szenarios [Computer-supported diagnostics with complex dynamic systems]. *Zeitschrift für Arbeits- und Organisationspsychologie*, *37*, 109–118.

Gaines, B. R., & Shaw, M. (1992). Knowledge acquisition tools based on personal construct psychology. *Knowledge Engineering Review*, *8*, 49–85.

Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker, J. F. (1992). Electronic brainstrorming and group size. *The Academy of Management Journal*, *35*(2), 350–369.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*, 121–123.

Glasersfeld, E. von. (1996). *Radikaler Konstruktivismus: Ideen, Ergebnisse, Probleme* [Radical constructivism: Ideas, results, problems]. Frankfurt a. M.: Suhrkamp.

Glasersfeld, E. von. (1997). *Wege des Wissens: konstruktivistische Erkundungen durch unser Denken* [The ways of knowledge: Constructivist explorations through our thinking]. Heidelberg: Carl-Auer-Systeme.

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 75–87). Nowood, NJ: Ablex.

Goldsmith, T. E., & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt (Ed.), *Pathfinder associative* *networks : studies in knowledge organization* (pp. 241–254). Norwood, NJ: Ablex.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, *83*(1), 88–96.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23.

Griffin, D., & Gonzales, R. (1995). Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, *118*(3), 430–439.

Gronau, N. (2002). A procedure model for evaluating knowledge management systems. In H. Arabnia, M. Youngsong, & B. Prasad (Eds.), *Proceedings of the international conference on information and knowledge engineering 2002* (pp. 78–83). Athens, Georgia: CSREA Press.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275.

Güldenberg, S. (1999). *Wissensmanagement und Wissenscontrolling in lernenden Organisationen* [Knowledge management and kmowledge contolling in learning organizations]. Wiesbaden: Gabler / Dt. Univ. Verl.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

Hansen, M., Nohria, N., & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review*, *77*(2), 106–116.

Hartmann, M. (1996). *Problemlöseprozesse in Kleingruppen* [Problem solving processes in small groups]. Hamburg: Kovac.

Hastie, T., & Efron, B. (2007). *lars: Least angle regression, lasso and forward stagewise.* R package version 0.9-7. Retrieved 2007-10-15, from `http://www-stat.stanford.edu/~hastie/Papers/`

Hastie, T., Taylor, J., Tibshirani, R., & Walther, G. (2007). Forward stagewise regression and the monotone lasso. *Electronic Journal of Statistics*, *1*, 1–29.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning : Data mining, inference, and prediction.* New York: Springer.

Hausmann, R. G. (2005). *Elaborative and critical dialog: Two potentially effective problem-solving and learning interactions.* Unpublished doctoral dissertation, University of Pittsburgh, Department of Philosophy. Available from `http://etd.library.pitt.edu/ETD/available/etd-08172005-144859/unrestricted/RGMHausmannETD2005.pdf`

Heer, J. (2004). *Prefuse: a software framework for interactive information visualization.* Unpublished master thesis, Computer Science

Division, University of California, Berkeley, U.S.A. Retrieved 2005-01-31, from `http://jheer.org/publications/2004-Heer-prefuse-Masters.pdf`

Heer, J., Card, S. K., & Landay, J. A. (2005). *Prefuse: a toolkit for interactive information visualization.* Online proceedings of CHI 2005, Human Factors in Computing Systems. Available from `http://jheer.org/publications/2005-prefuse-CHI.pdf`

Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in american life.* New York: Free Press.

Hirshman, E., & Wallendorf, M. R. (1982). Free-response and card-sort techniques for assessing cognitive content: Two studies concerning their stability, validity and utility. *Perceptual and Motor Skills*, *54*, 1095–1110.

Hoole, E. (2006). *Integrating and evaluating mathematical models of assessing structural knowledge: comparing associative networking methodologies.* Unpublished doctoral dissertation, Center for Assessment and Research Study, James Madison University, Harrisonburg, Virginia.

Horn, J., & Masunaga, H. (2006). A merging theory of expertise and intelligence. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The cambridge handbook of expertise and expert performance* (pp. 587–611). New York: Cambridge University Press.

Hussy, W., & Klinck, D. (1990). Wissenserwerb beim Lösen komplexer Probleme, der Einfluß von Kontext und Vorwissen [Knowledge acquisition in complex problem solving, the influence of conext and previous knowledge]. *Trierer psychologische Berichte*, *17*(9), 1–30.

Igata, N., Tsuda, H., Katayama, Y., & Kozakura, F. (2003). *Semantic groupware and its application to KnowWho using RDF.* Paper presented at the 7th IEEE International Symposium on Wearable Computers, October 21-23, 2003, White Plains, NY.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS-Test)* [Berlin intelligence structure test (BIS-Test)]. Göttingen: Hogrefe (Testzentrale).

Jansen, D. (2004). *Networks, social capital and knowledge production* (FÖV discussion papers ; 8). Speyer: Forschungsinstut für Öffentliche Verwaltung.

Johnson, P. J., Goldsmith, T. E., & Teague, K. W. (1995). Similarity, structure, and knowledge: A representational approach to assessment. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 221–249). Hillsdale, NJ: Lawrence Erlbaum.

Jonassen, D. H. (1993). Changes in knowledge structure from building semantic net versus production rule representations of subject content.

Journal of Computer-Based Instruction, *20*(4), 99–106.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge : techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum.

Juslin, P., & Olsson, H. (2005). Capacity limitations and the detection of correlations: comment on Kareev (2000). *Psychological Review*, *112*(1), 256–267; discussion 280–285.

Kahnemann, D., Slovic, P., & Tversky, A. (1982). *Jugdment under uncertainty: Heuristics and biases*. Cambridge, MA: Cambridge University Press.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review*, *107*(2), 397–402.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 451–477). New York: Cambridge University Press.

Kelly, G. A. (1955). *The psychology of personal constructs: A theory of personality*. New York, NY: Norton.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Kersting, M. (1991). *Wissensdiagnostik beim Problemlösen* [Knowledge diagnosis in problem solving]. Unpublished doctoral dissertation, Institute of Psychology, Free University Berlin, Germany.

Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien? Zur Kriteriumsvalidität von Problemlöseszenarien und Intelligenztests* [Diagnostics and employee selection with computer-supported problem solving scenarios? On the criterion validity of problem solving scenarios and intelligence tests]. Göttingen: Hogrefe.

Kersting, M., & Süß, H.-M. (1995). Kontentvalide Wissensdiagnostik und Problemlösen: Zur Entwicklung, testtheoretischen Begründung und empirischen Bewährung eines problemspezifischen Diagnoseverfahrens [Content valid knowledge diagnosis and problem solving: On the the development, test-theoretical explanation, and empiric approval of a problem-specific diagnostic method]. *Zeitschrift für Pädagogische Psychologie*, *9*(2), 83–93.

Klimecki, R., & Thomae, M. (2000). *Zur Bedeutung von Strukturmodellen für das Wissensmanagement* [On the importance of structural models for knowledge management]. Presentation at the Workshop der Kommission Organisation, 3rd and 4th of March, 2000. Zürich.

Klix, F. (1984). *Gedaechtnis, Wissen, Wissensnutzung* [Memory, kowledge, knowledge use]. Berlin: VEB Deutscher Verlag der Wissenschaften.

Klocke, U. (2004). *Folgen von Machtausübung und Einflussnahme für Wissenszuwachs und Effektivität in Kleingruppen* [The effects of the exertion of power and influence on knowledge increase and effectiveness in small groups]. Berlin: Dissertation.de.

Kluge, A. (2004). *Wissenserwerb für das Steuern komplexer Systeme* [Knowledge acquisition for the control of complex systems]. Lengerich: Pabst.

Kluwe, R. (1991). Zum Problem der Wissensvoraussetzungen für Prozess- und Systemsteuerung [On the problem of knowledge requirements for process and system control]. *Zeitschrift für Psychologie*, *Supplement 11*, 311–324.

Kluwe, R. (1996). Effects of type of learning on control performance. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (Vol. 2, pp. 81–88). Aldershot: Ashgate.

Kluwe, R. (1997). Acquisition of knowledge in the control of a simulated technical system. *Le Travail Humain*, *60*, 61–85.

Kluwe, R. (2006). Informationsaufnahme und Informationsverarbeitung [Information acquisition and information processing]. In B. Zimolong & U. Konradt (Eds.), *Enzyklopädie der Psychologie: Wirtschafts-, Organisations- und Arbeitspsychologie* (Vol. 2: Ingenieurpsychologie, pp. 35–71). Göttingen: Hogrefe.

Kluwe, R., Haider, H., & Misiak, C. (1990). Learning by doing in the control of a complex system. In H. Mandl, E. Bennett, E. de Corte, & H. F. Friedrich (Eds.), *Learning and instruction* (pp. 197–218). Oxford: Pergamon Press.

Knippenberg, D. van, De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, *89*(6), 1008–1022.

Knippenberg, D. van, & Schippers, M. (2007). Work group diversity. *Annual Review of Psychology*, *58*, 515–541.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities and the replication of technology. *Organization Science*, *3*(3), 383–397.

Laughlin, P. (1980). Social combination processes in cooperative problem-solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 127–155). Hillsdale, NJ: Erlbaum.

Laughlin, P. (1988). Collective induction: Group performance, social combination processes, and mutual majority and minority influence. *Journal of Personality and Social Psychology*, *54*, 254–267.

Le, S. Q., Ho, T. B., & Phan, T. H. (2003). A novel graph-based similarity measure for 2d chemical structures. *Genome Informatics*, *14*(2), 82–

91.

Lee, S., Choi, K. S., & Choe, I.-S. (2002). *Representation of tacit knowledge in organizations.* Online Proceedings of the third European conference on organizational knowledge, learning and capabilities OKLC 2002 5th & 6th April 2002, Athens, Greece. Retrieved 2004-06-04, from `http://www.alba.edu.gr/OKLC2002/Proceedings/pdf_files/ID121.pdf`

Levesque, L. L., Wilson, J. M., & Wholey, D. R. (2001). Cognitive divergence and shared mental models in software development project teams. *Journal of Organizational Behavior*, *22*(2), 135–144.

Li, M., & Gao, F. (2003). Why Nonaka highlights tacit knowledge: a critical review. *Journal of Knowledge Management*, *7*(4), 6–14.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance.* Englewood Cliffs, NJ: Prentice Hall.

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometirka*, *12*, 89–104.

Mansfield, E., & Helms, B. (1982). Detecting multicollinearity. *The American Statistician*, *36*(3), 158–160.

Marescaux, P., Luc, F., & Karnas, G. (1989). Modes d'apprenticage selectif et connaissances acquises au control d'un processus: Evaluation d'un modele simule [Selective modes of learning and knowledge acquisition in process control: Evaluation of a simulation model]. *Cahiers de Psychologie Cognitive*, *9*, 239–264.

Markowitsch, H.-J. (1992). *Neuropsychologie des Gedächtnis* [The neuropsychology of memory]. Göttingen: Hogrefe.

Markowitsch, H.-J. (1999). *Gedächtnisstörungen* [Memory disorders]. Stuttgart: Kohlhammer.

Markowitsch, H.-J. (2002). *Dem Gedächtnis auf der Spur: vom Erinnern und Vergessen* [Tracing memory: on remembering and forgetting]. Darmstadt: Wissenschaftliche Buchgesellschaft.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, *42*, 236–243.

Marx, D. M., & Stapel, D. A. (2006). It's all in the timing: measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology*, *36*(5), 687-698.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273–283.

McDougall, S. J. P., Curry, M. B., & Brujin, O. de. (2001). The effects of visual information on users' mental models: An evaluation of pathfinder analysis as a measure of icon usability. *International Journal of Cog-*

nitive Ergonomics, *5*(1), 59–84.

McGrath, J. E., Berdahl, J. L., & Arrow, H. (1995). Traits, expectations, culture, and clout: The dynamics of diversity in work groups. In S. E. Jackson & M. N. Rudermann (Eds.), *Diversity in work teams, research paradigms for a changing workplace* (pp. 17–47). Washington: American Psycological Aassociation.

Meyer, B., Ceglarek, P., Kermas, M., Lattke, T., & Rothe, H.-J. (2006). *AST: Ein computerbasiertes Verfahren zur Wissensstrukturdiagnose und seine Anwendung* [AST: A computer-based method for knowledge diagnosis and its application]. Paper presented at 48. Tagung experimentell arbeitender Psychologen (TEAP), 26.-29.03.2006, Universität Mainz.

Meyer, B., & Spiekermann, S. (2006a). skillMap: dynamic visualization of shared organizational context. In F. Lehner, H. Nösekabel, & P. Kleinschmidt (Eds.), *Multikonferenz Wirtschaftsinformatik 2006* (Vol. 2, pp. 528–540). Berlin: GITO.

Meyer, B., & Spiekermann, S. (2006b). skillMap: Wissenstransparenz für verteilte Organisationen [skillMap: knowledge transparency for distributed organizations]. In N. Gronau, P. Pawlowsky, P. Schütt, & M. Weber (Eds.), *Mit Wissensmanagement besser im Wettbewerb!* (pp. 75–83). Poing: CMP-WEKA.

Meyer, B., Spiekermann, S., & Hertlein, M. (2005). *skillMap: identification of unrecognized parallel developments and of Communities of Practice in distributed organizations.* Presentation at the The First World Congress of the International Federation for Systems Research (IFSR2005), November 14-17th. Kobe, Japan.

Meyer, B., & Sugiyama, K. (2007). The concept of knowledge in KM: a dimensional model. *Journal of Knowledge Management*, *11*(1), 17–35.

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary bounfaries. *Journal of Organizational Behavior*, *22*, 89–106.

Mohammed, S., & Ringseis, E. (2001). Cognitive diversity and consensus in group decision making: The role of input, processes, and outcomes. *Organizational Behavior and Human Decision Processes*, *85*, 310–335.

Nair, K., & Ramnarayan, S. (2000). Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality*, *34*(3), 305–328.

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychol-*

ogist, *51*, 102–116.

Nonaka, I., & Konno, N. (1998). The concept of "ba": building a foundation for knowledge creation. *California Management Review*, *40*(3), 40–54.

Nonaka, I., & Reinmoeller, P. (2000). Dynamic business systems for knowledge creation and utilization. In C. Despres & D. Chauval (Eds.), *Knowledge horizons* (pp. 89–112). Boston: Butterworth-Heinemann.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: how Japanese companies create the dynamics of innovation.* New York: Oxford University Press.

Nonaka, I., Toyama, R., & Konno, N. (2000). Seci, ba and leadership: a unified model of dynamic knowledge creation. *Long Range Planning*, *33*(1), 5–34.

Novak, J. D. (1998). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations.* Mahwah, NJ: Lawrence Erlbaum Associates.

Novak, J. D., & Cañas, A. J. (2006). *The theory underlying concept maps and how to construct them.* Technical Report IHMC CmapTools 2006-01. Florida Institute for Human and Machine Cognition. Retrieved 2006-11-08, from `http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf`

Novak, J. D., & Gowin, D. (1996). *Learning how to learn.* New York: Cambridge University Press.

Ohtsubo, Y. (2005). Should information be redundantly distributed among group members? Effective use of group memory in collaborative problem solving. *Journal of Applied Cognitive Psychology*, *19*, 1219–1233.

Opwis, K., & Lüer, G. (1996). Modelle der Repräsentation von Wissen [Models of knowledge representation]. In D. Albert & K.-H. Stapf (Eds.), *Enzyklopädie der Psychologie* (Vol. 5: Gedächtnis, pp. 337–432). Göttingen: Hogrefe.

Phye, G. D. (Ed.). (2002). *Handbook of academic learning: construction of knowledge* (3rd ed.). San Diego: Academic Press.

Pitman, E. (1937). Significance tests which may be applied to samples from any population I and II. *Journal of the Royal Statistical Society*, *Supp 4*, 119-130, 225-232.

Pitman, E. (1938). Significance tests which may be applied to samples from any population: III. *Biometrika*, *29*, 322–335.

Polanyi, M. (1958). *Personal knowledge.* London: Routledge & Kegan Paul.

Polanyi, M. (1966). *The tacit dimension.* Garden City, NY: Doubleday.

Polanyi, M. (1983). *The tacit dimension* (Reprinted ed.). Gloucester, MA: Smith.

Presse- und Informationsamt der Bundesregierung. (2007). Ressource

Wissen nutzen [Using knowledge as a ressource]. *e.conomy*, *5*(46). Available from http://www.bundesregierung.de/Content/DE/EMagazines/economy/046/sp-1-ressource-wissen-nutzen.html

Preußler, W. (1996). Zur Rolle expliziten und impliziten Wissens bei der Steuerung dynamischer Systeme [On the role of explicit and implicit knowledge in complex system control]. *Zeitschrift für Experimentelle Psychologie*, *43*(3), 399–434.

Preußler, W. (1998). Strukturwissen als Voraussetzung für die Steuerung komplexer dynamischer Systeme [Structural knowledge as requirement for complex system control]. *Zeitschrift für Experimentelle Psychologie*, *45*(3), 218–240.

Probst, G. J. B., Raub, S., & Romhardt, K. (1999). *Wissen managen* [Managing knowledge]. Wiesbaden: Gabler.

Punin, J., & Krishnamoorthy, M. (2001). *XGMML (eXtensible Graph Markup and Modeling Language) 1.0 draft specification.* Online draft. Available from http://www.cs.rpi.edu/~puninj/XGMML/draft-xgmml.html

Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg [On the relationship between test intelligence and problem solving performance]. *Zeitschrift für Psychologie*, *189*, 79–100.

Putz-Osterloh, W. (1993). Strategies for knowledge acquisition and transfer of knowledge in dynamic tasks. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 331–350). Amsterdam: North-Holland.

Putz-Osterloh, W., Bott, B., & Houben, I. (1988). Beeinflusst Wissen über ein realitätsnahes System dessen Steuerung? [Does knowledge about a realistic system influence its control?] *Sprache & Kognition*, *7*, 12–24.

Putz-Osterloh, W., & Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest [On the predictability of complex problem solving performance by IQ-test results]. *Zeitschrift für experimentelle und angewandte Psychologie*, *28*, 309–334.

Quesada, J., Kintsch, W., & Gomez, E. (2002). A computational theory of complex problem solving using latent semantic analysis. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual conference of the cognitive science society Fairfax, VA* (pp. 750–755). Mahwah, NJ: Lawrence Erlbaum.

Quillian, M. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 216–270). Cambridge, MA: MIT Press.

Quillian, M. (1988). Semantic memory. In A. M. Collins & E. E. Edwards

(Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 80–101). San Mateo, CA: Morgan Kaufmann.

R Development Core Team. (2007). *R: A language and environment for statistical computing.* Computer Software, available at http://www.R-project.org. Vienna, Austria. Retrieved 2007-12-05, from `http://www.R-project.org`

Rapoport, A., Livant, W. P., & Boyd, J. (1966). A study of lexical graphs. *Foundations of Language*, *2*, 338–376.

Rauchhaupt, U. von. (2005). *Wittgensteins Klarinette - Gegenwart und Zukunft des Wissens* [Wittgenstein's clarinet - present and future of knowledge]. Berlin: Berliner Taschenbuch Verlag.

Reber, A. S. (1996). *Implicit learning and tacit knowledge : an essay on the cognitive unconscious.* New York, NY: Oxford University Press.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual difference and IQ. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*(5), 888–896.

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence*, *16*, 259–272.

Rehäuser, J., & Krcmar, H. (1996). Wissensmanagement im Unternehmen [Knowledge management in the company]. In G. Schreyögg & P. Conrad (Eds.), *Managementforschung 6: Wissensmanagement* (pp. 1–40). Berlin: de Gruyter.

Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird [Inert knowledge: if what is learned is not being used]. *Psychologische Rundschau*, *47*, 78–92.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357.

Rost, J. (1980). *Gedächtnispsychologische Grundlagen naturwissenschaftlichen Denkens* [Memory-psychological foundations of scientific thinking]. Weinheim: Beltz.

Rothe, H.-J., & Ceglarek, P. (2006). Wort-Assoziations-Experimente und Struktur-Lege-Techniken als wissensdiagnostische Methoden [Word-association experiments and structure laying techniquues as methods for knowledge diagnosis]. In B. Bergmann & A. Pohlandt (Eds.), *Entwicklung von Kompetenz und Innovationsfähigkeit in Unternehmen* (pp. 151–162). Dreseden: saxOprint.

Rothe, H.-J., & Warning, J. (1991). Zur Identifikation von semantischen Relationen im fachspezifischen Gedächtnisbesitz von Experten vermittels Analogiebildung [On the identification of semantic relations in experts' subject-specific memory]. *Zeitschrift für Psychologie*, *Supplement 11*, 375–384.

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly*, *6*(3), 243–255.

Ruggles, R. (1999). The state of the notion: Knowledge management in practice. In J. W. Cortada & J. A. Woods (Eds.), *The knowledge management yearbook 1999-2000* (pp. 295–305). Boston, MA: Butterworth-Heinemann.

Rulke, D. L., & Galskiewicz, J. (2000). Distribution of knowledge, group network structure, and group performance. *Management Science*, *46*(5), 612–625.

Ruzgis, P. (1994). Thurstone, L. L. (1887-1955). In R. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 1081–1084). New York: Macmillian.

Saint-Onge, H., & Wallace, D. (2003). *Leveraging communities of practice for strategic advantage*. Amsterdam: Butterworth-Heinemann.

Sanderson, P. (1989). Verbalizable knowledge and skilled task performance: Associations, dissociations, and mental models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 729–747.

Satzger, W., Fessman, H., & Engel, R. R. (2002). Die Äquivalenz dreier deutscher Vokabeltests zur deutschen Version des HAWIE-R [The equivalence of three German vocabulary tests and the German version of the Wechsler adult intelligence scale-revised (HAWIE-R)]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *23*(2), 159–170.

Sauer, J., Felsing, T., Franke, H., & Rüttinger, B. (2006). Cognitive diversity and team performance in a complex multiple task environment. *Ergonomics*, *49*(10), 934–954.

Schauenburg, B. (2004). *Motivierter Informationsaustausch in Gruppen: Der Einfluss individueller Ziele und Gruppenziele* [Motivated information exchange in groups: The influence of individual goals and group goals]. Unpublished doctoral dissertation, Georg-August Universität Göttingen, Germany.

Scheele, B., & Groeben, N. (1984). *Die Heidelberger Struktur-Lege-Technik (SLT): Eine Dialog-Konsens-Methode zur Erhebung Subjektiver Theorien mittlerer Reichweite* [The heidelberger structure-laying-technique (SLT): A dialogue-consensus method for the elicitation of medium-range subjective theories]. Weinheim: Beltz.

Scheele, B., & Groeben, N. (1988). *Dialog-Konsens-Methoden zur Rekonstruktion subjektiver Theorien: die Heidelberger Struktur-Lege-Technik (SLT)* [Dialogue-consensus methods for the reconstruction of subjective theories: The Heidelberg structure laying technique]. Tübingen: Francke.

Schimansky, A. (2006). *Moderation als Strukturierungsansatz effektiver Gruppenarbeit – eine feldexperimentelle Evaluation.* [Moderation as structuring approach of effective group work – a field-experimental evaluation]. Lengerich: Pabst.

Schindler, R. (2002). *Knowledge sharing by means of lessons learned repositories: How to assist understanding, reflection and transfer.* Presentation at the International Conference on the Future of Knowledge Management, Berlin, 08.-10.03.2002.

Schmidt, F. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173–1181.

Schmidt, K.-H., & Metzler, P. (1992). *Wortschtztest (WST)* [Vocabulary test (WST)]. Weinheim: Beltz.

Schnotz, W. (1994). *Aufbau von Wissensstrukturen: Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten* [The contruction of knowledge structures: Analyses of the formation of coherence in knowledge acquisition from texts]. Weinheim: Beltz.

Scholl, W. (1990). Die Produktion von Wissen zur Bewältigung komplexer organisatorischer Situationen [The production of knowledge for dealing with complex organizational situations]. In R. Fisch (Ed.), *Vom Umgang mit Komplexität in Organisationen* (pp. 107–128). Konstanz: Universitätsverlag.

Scholl, W. (1992). Informationspathologien [Information pathologies]. In E. Frese (Ed.), *Handwörterbuch der Organisation* (Vol. 3, pp. 900–912). Stuttgart: Poeschel.

Scholl, W. (1996). Effective teamwork - a theoretical model and a test in the field. In E. Witte & J. H. Davis (Eds.), *Understanding group behavior* (Vol. 2: Small group processes and interpersonal relations, pp. 127–146). Mahwah, NJ: Lawrence Erlbaum.

Scholl, W. (1997). Gruppenarbeit: Die Kluft zwischen sozialpsychologischer Theoriebildung und organisationspsychologischer Anwendung [Group work: The barrier between the construction of social-psychological theories and organizational practice]. *Gruppendynamik*, *28*(4), 381–403.

Scholl, W. (2004). *Innovation und Information. Wie in Unternehmen neues Wissen produziert wird* [Innovation and information: How new knowledge is produced in companies]. Göttingen: Hogrefe.

Scholl, W. (2005). Grundprobleme der Teamarbeit und ihre Bewältigung: Ein Kausalmodell [Elementry problems of group work and their overcoming: A causal model]. In M. Högl & H. G. Gemünden (Eds.), *Management von Teams* (3rd ed., pp. 33–66). Wiesbaden: Gabler.

Scholl, W., König, C., Meyer, B., & Heisig, P. (2004). The future of knowledge management: An international delphi study. *Journal of Knowledge Management*, *8*(2), 19–35.

Schoppek, W. (2002). Examples, rules, and strategies in the control of dynamic systems. *Cognitive Science Quarterly*, *2*, 63–92.

Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, *88*, 563–586.

Schütt, P. (2003). *Die dritte Generation des Wissensmanagements.* Online-Paper (KM-Journal). Retrieved 2003-04-11, from `http://www.km-a.net/file_upload/km-journal_1-03_schuett.pdf`

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: studies in knowledge organization.* Norwood, NJ: Ablex.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. J., Tucker, R. G., et al. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, *23*, 699–728.

Shavelson, R. J. (1972). Some aspects of the corrspondance between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, *63*, 225–234.

Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. *Journal of Research in Science Teaching*, *10*, 101–111.

Shavelson, R. J., & Stanton, G. C. (1975). Construct validation: Methodology and application to three measures of cognitive structure. *Journal of Educational Measurement*, *12*, 67–85.

Sinz, R. (1979). *Neurobiologie und Gedächtnis: Neuronennetzwerke und Informationsspeicherung im menschlichen Gehirn* [Neurobiology and memory: neural networks and information storage in the human brain]. Stuttgart: Fischer.

Smith, K. A., Petersen, R. P., Johnson, D. W., & Johnson, R. T. (1986). The effects of controversy and concurrence seeking on effective decision making. *The Journal of Social Psychology*, *126*, 237–248.

Snowden, D. (2002). *Complex acts of knowing - paradox and descriptive self-awareness.* Online paper. IBM Global Services. Retrieved 2002-09-07, from `http://www-1.ibm.com/services/files/complex.pdf`

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Spender, J. C. (1996). Making knowledge the basis of a dynamic theory of the firm. *Strategic Management Journal*, *17*, 45–62.

Squire, L. R., & Frambach, M. (1990). Cognitive skill learning in amnesia. *Journal of Psychobiology*, *18*, 109–117.

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, *44*(1), 453–495.

Stanley, W. B., Mathews, R., Buss, R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction and practice on a simulated process control task. *Quarterly Journal of Experimental Psychology*, *41*, 553–577.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*, 1467–1478.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology*, *53*, 81–93.

Statsoft Inc. (2006). *Electronic statistics textbook.* Tulsa, OK: StatSoft.

Stehr, N. (1994). *Knowledge societies.* London: Sage.

Steiner, I. D. (1972). *Group process and productivity.* New York, NY: Academic Press.

Steiner, I. D. (1976). Task-performing groups. In J. W. Thibaut, J. T. Spence, & R. C. Carson (Eds.), *Contemporary topics in social psychology* (pp. 393–422). Morristown, NJ: General Learning Press.

Sternberg, R. (1982). *Handbook of human intelligence.* Cambridge, MA: University Press.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., et al. (2000). *Practical intelligence in everyday life.* Cambridge, MA: University Press.

Strauß, B., & Kleinmann, M. (1995). *Computersimulierte Szenarien in der Personalarbeit* [Computer-simulated szenraios in HR]. Göttingen: Verlag für Angewandte Psychologie.

Strohschneider, S. (1990). *Wissenserwerb und Handlungsregulation* [Knowledge acquisition and performance regulation]. Wiesbaden: Deutscher Universitäts Verlag.

Strohschneider, S. (1991). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme [Problem solving and intelligence: On the effects of making complex problems more concrete]. *Diagnostica*, *37*, 353–371.

Strube, G. (1984). *Assoziation: der Prozeß des Erinnerns und die Struktur des Gedächtnisses* [Association: The process of remembering and the structure of memory]. Berlin: Springer.

Strube, G. (1985). *Knowing what's going to happen in life I: A model of* *biographical knowledge* (Tech. Rep.). Munich: Max Planck Institute for Psychological Research. Paper 13/85.

Strube, G., & Schlieder, C. (1998). Wissensrepräsentation im Symbolverarbeitungsansatz [Knowledge representation in the symbol processing approach]. In F. Klix & H. Spada (Eds.), *Enzyklopädie der Psychologie* (Vol. 6: Kognition, pp. 501–530). Götingen: Hogrefe.

Sugiyama, K. (2002). *Graph drawing and applications for software and knowledge engineers.* Singapore: World Scientific.

Sugiyama, K., & Misue, K. (1996). A generic compound graph visualizer/manipulator: D-abductor. In F.-J. Brandenburg (Ed.), *Graph Drawing, Symposium on Graph Drawing, GD '95, Passau, Germany, September 20-22, 1995, Proceedings. Lecture Notes in Computer Science* (Vol. 1027, pp. 500–503). Heidelberg: Springer.

Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen* [Test intelligence, knowledge, and problem solving: cognitive requirements for successful performance in computer-simulated problems]. Göttingen: Hogrefe.

Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen [Test intelligence and complex problem solving]. *Psychologische Rundschau*, *50*, 220–228.

Süß, H.-M., & Faulhaber, J. (1990). *Berliner Version des Szenarios Schneiderwerkstatt* [Berlin version of the taylorshop scenario]. Computer Software. Freie Universität, Fachbereich Erziehungs- und Unterrichtswissenschaften, FPS "Intelligenz und Wissen".

Süß, H.-M., Kersting, M., & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen [Test intelligence and knowledge as performance predictors in computer-simulated complex problems]. *Diagnostica*, *37*(4), 334–352.

Süß, H.-M., Kersting, M., & Oberauer, K. (1993). Zur Vorhersage von Steuerungsleistungen an computersimulierten Szenarien durch Wissen und Intelligenz [On the prediction of control performance in computer-simulated scenarios through knowledge and test intelligence]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *14*, 189–203.

Takeuchi, H., Nonaka, I., & Hitotsubashi-Daigaku (Eds.). (2004). *Hitotsubashi on knowledge management.* Singapore: Wiley.

ter Braak, C. J. F. (1992). Permutation versus bootstrap significance tests in multiple regression and anova. In K.-H. Jöckel, G. Rothe, & W. Sendler (Eds.), *Bootstrapping and related techniques* (pp. 79–86). Berlin: Springer.

Thomson, G. H. (1939). *The factorial analysis of human ability.* Boston: Houghton Mifflin.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, Methodological*, *58*, 267–288.

Tindale, R. S., & Sheffey, S. (2002). Shared information, cognitive load, and group memory. *Group Processes & Intergroup Relations*, *5*(1), 5–18.

Trier, M. (2005). A tool for IT-supported visualization and analysis of virtual communication networks in knowledge communities. In O. K. Ferstl, E. J. Sinz, S. Eckert, & T. Isselhorst (Eds.), *Wirtschaftsinformatik 2005* (pp. 963–983). Heidelberg: Physica.

Tsuda, H. (2003). *Advanced "KnowWho" with semantic web technology: Human knowledge navigator.* Online paper. Fujitsu Laboratories Limited. Retrieved 2005-09-21, from `http://www.labs.fujitsu.com/en/techinfo/okar/sw/knowwho_catalog200307en.pdf`

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1995). Organization of memory: Quo vadis. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 839–847). Camebridge, MA: MIT Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *1*(53), 1–25.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301–306.

Vernon, P. E. (1950). *The structure of human abilities.* London: Methuen.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *27*(1), 92–114.

Wagener, D., Hochholdinger, S., Conrad, W., & Wittmann, W. W. (1997). Empirische Validierung des komplexen Szenarios FSYS im Rahmen einer Personalentwicklungsmaßnahme [Empiric validation of the complex scenario FSYS during a personnel development intervention]. *Berichte des Lehrstuhls Psychologie II der Universität Mannheim*, *10*, 1–35.

Wenger, E. (2003). *Communities of practice: learning, meaning, and identity* (Reprinted ed.). Cambridge, MA: Cambridge University Press.

West, L. H. T., Fensham, P. J., & Garrard, J. E. (1985). Describing the cognitive structures of learners following instruction in chemistry. In

L. H. T. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 29–49). Orlando, FL: Academic Press.

White, R. T. (1985). Interview protocols and dimensions of cognitive structure. In L. H. T. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 51–59). Orlando, FL: Academic Press.

Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. *Organizational Behavior*, *20*, 77–140.

Winitzky, N., Kauchak, D., & Kelly, M. (1994). Measuring teachers' structural knowledge. *Teaching and Teacher Education*, *10*(2), 125–139.

Wittgenstein, L. (2001). *Philosophische Untersuchungen. Kritisch-genetische Edition* [Philosophic analyses. critical-genetical edition]. Frakfurt a. M.: Wissenschaftliche Buchgesellschaft.

Wittmann, W. W., Süß, H.-M., & Oberauer, K. (1996). Determinanten komplexen Problemlösens. *Berichte des Lehrstuhls Psychologie II der Universität Mannheim*, *9*, 1–25.

Wyman, B. G., & Randel, J. M. (1998). The relation of knowledge organization to performance of a complex cognitive task. *Applied cognitive psychology*, *12*(3), 251–264.

Zajonc, R. B., & Smoke, W. H. (1959). Redundancy in task assignment and group performance. *Psychometirka*, *24*, 361–369.

Appendix A

Additional Taylorshop Experimental Material

A.1 TAYLORSHOP Interconnection Scheme

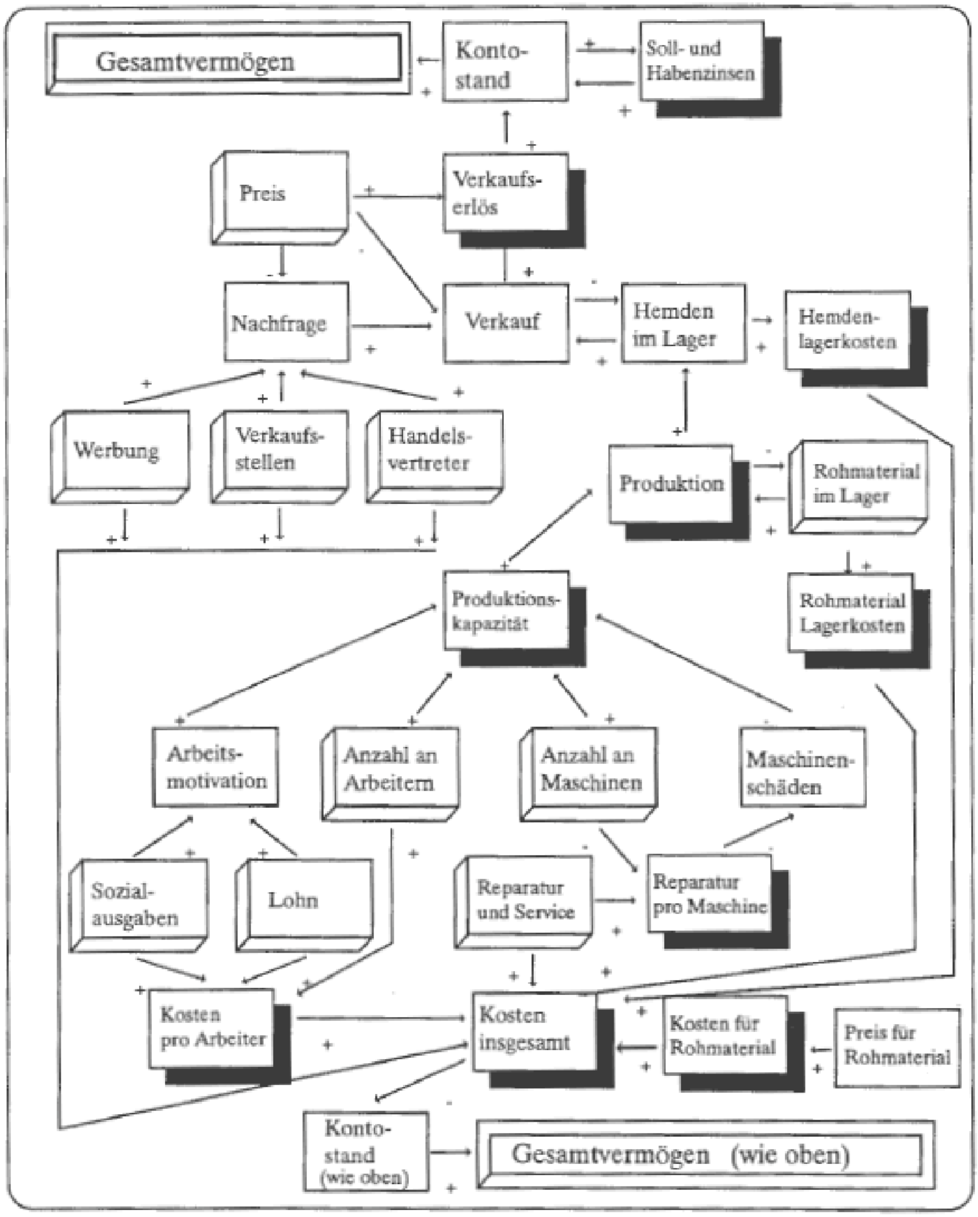


Figure A.1: German TAYLORSHOP scenario variable interconnection scheme (reprinted with permission from Süß, 1996, p. 102)

A.2 German TAYLORSHOP Knowledge Elements

Table A.1: Knowledge elements for TAYLORSHOP scenario control that were embedded in instructional texts based on Klocke (2004) in the original German language version

Element	Title	Content
G1	Stabiles System	Im Vergleich zu anderen computersimulierten Aufgaben ist die Hemdenfabrik ein relativ stabiles System. Kleine Eingriffe bewirken auch nur kleine Effekte. Man braucht also bei seinen Maßnahmen nicht zu vorsichtig vorzugehen, da die Fabrik nicht so leicht aus den Fugen gerät. Die Bank gewährt der Fabrik großzügigen Kredit. Wenn das Geld gewinnbringend investiert wird, ist es daher unproblematisch, Schulden zu machen. Verkaufsstellen oder Maschinen sollte man also nicht verkaufen, bloß weil man einen negativen Kontostand hat.
G2	Geichgewichtszustand	um das Gesamtvermögen zu maximieren, ist es einerseits sinnvoll, das System in einen Gleichgewichtszustand zu bekommen. Das erreicht man z. B., indem man dafür sorgt, dass sich Nachfrage und Angebot an Hemden bzw. Anzahl an Maschinen und Arbeitern entsprechen. Andererseits sollte man versuchen, den Betrieb zu vergrößern und den Verkauf zu steigern.
A	Rohmaterial und Lager	Der Rohmaterialpreis ist von sämtlichen Systemvariablen unabhängig. Er unterliegt allerdings deutlichen, marktbedingten Schwankungen. Man sollte darauf achten, dass immer ausreichend Rohmaerial vorhanden ist, um die Produktion sicherzustellen. Ein häufiger Fehler besteht darin, das Rohmateriallager nicht wieder aufzufüllen.
B	Hemden im Lager	Die Lagerhaltung ist mit Kosten verbunden. Deshalb sollten nicht zu viele Hemden auf Halde produziert werden. Gleichzeitig sollte aber soviel produziert werden, dass die Nachfrage befriedigt werden kann.
C	Angebot und Nachfrage	Die Anzahl der fertigen Hemden im Lager und die Menge Rohmaterial im Lager müssen zusammen mindestens so hoch sein wie die Nachfrage. Wenn Angebot und Nachfrage auseinander klaffen, ist es sinnvoller, eines von beiden zu steigern als eines von beiden zu senken.
D	Allgemeines	Damit sich die Investitionen in in Maschinen und Personal rentieren, sollte sowohl die Arbeitsauslastung als auch die Maschinenauslastung möglichst hoch sein.

Table A.1: (continued)

Element	Title	Content
E	Maschinen-auslastung	Die Maschinenauslastung gibt an, wie viel Prozent der Hemden, die mit der Maschine produziert werden können, auch tatsächlich produziert werden. Wenn Sie z. B. vier 100er Maschinen besitzen und tatsächlich etwa 400 Hemden produzieren, dann ist die Maschinenauslastung 100 %. Wenn die Maschinenauslastung deutlich unter 100% ist, werden Ihre Investitionen nicht optimal genutzt. Der Grund könnte sein, dass Sie zu wenig Arbeiter (2) für Ihre Maschinen (3) eingestellt haben, dass Ihre Arbeiter unmotiviert (1) sind oder dass die Maschinen Schäden (7) aufweisen. Vielleicht haben Sie auch einfach vergessen, genug Rohmaterial einzukaufen (12).
F	Arbeitsaus-lastung	Die Arbeitsauslastung gibt an, wie viel Prozent der Hemden, die die Arbeiter produzieren können, sie tatsächlich produzieren. Wenn Sie drei Arbeiter für drei 100er Maschinen eingestellt haben und diese auch etwa 300 Hemden produzieren, dann ist die Arbeitsauslastung 100 %. Wenn die Arbeitsauslastung niedriger ist, dann beschäftigen sich Ihre Arbeiter (2) womöglich mit anderen Dingen als dem Herstellen von Hemden. Vielleicht haben Sie es versäumt, ihnen genug intakte Maschinen (3) oder Rohmaterial (9) zur Verfügung zu stellen oder sie sollten sich um ihre Motivation (1) kümmern. Man sollte also nicht mehr Arbeiter als Maschinen haben. Wenn es mehr Arbeiter als Maschinen gibt, ist es jedoch besser, Maschinen zu kaufen (3) als Arbeiter zu entlassen (2). Schließlich wollen Sie Hemden verkaufen (17), um Ihr Gesamtvermögen zu steigern und dazu brauchen Sie Ihre Arbeiter. Die Arbeitsmotivation ist vom Lohn und von den Sozialausgaben abhängig. Beides zusammen erhöht aber auch die Kosten pro Arbeiter. Ein weiterer Grund dafür, warum Sie nur 100er-Maschinen betreiben sollten.
G	Investitions-zeitpunkt	Maschinen sollten Sie in den ersten Monaten kaufen, damit sich die Anschaffung lohnt. Sie sollten nur 100er Maschinen betreiben, um die vorhandene Arbeitskraft optimal zu nutzen.

Table A.1: (continued)

Element	Title	Content
H	Maschinen-schäden	Maschinenschäden (7) reduzieren die Produktionskapazität (6) und damit die tatsächliche Produktion (8). Durch Ausgaben für Wartung und Reparatur (5) können Schäden unabhängig vom Alter der Maschinen wirksam reduziert werden. Die Wartungsausgaben sollte man nie auf Null setzen. Wenn die Maschinenschäden mehr als 10 % betragen, sollte man die Wartungsausgaben erhöhen. Je mehr Maschinen man hat, desto mehr Geld muss man für Reparaturen ausgeben, da die Wartungsausgaben pro Maschine (4) dann abnehmen. Die Maschinenschäden (7) werden ausschließlich durch die Wartungsausgaben pro Maschine beeinflusst. Die Arbeitsmotivation (1) hat keinen Einfluss auf die Maschinenschäden.
I	Nachfrage	Die Nachfrage bestimmt, wie viele Hemben Sie pro Monat am Markt absetzen können. Die Ausgaben für Werbung erhöhen die Nachfrage, ebenso die Anzahl der Verkaufsstellen und die Anzahl der Handelsvertreter. Ein höherer Hemdenpreis senkt die Nachfrage. Der Preis hat einen stärkeren Einfluss auf die Nachfrage als jeweils Werbung, Verkaufsstellen und Vertreter.
J	Kosten	Die Posten Werbung, Anzahl der Verkaufsstellen und die Anzahl der Handelsvertreter erhöhen die monatlichen Kosten. Die monatlichen Kosten werden auch durch Rohmaterial-Lagerkosten, Kosten pro Arbeiter (Lohn- und Sozialaufwendungen) und durch evtl. Sollzinsen erhöht.

A.3 TAYLORSHOP Instructions (Example)

Die Hemdenfabrik

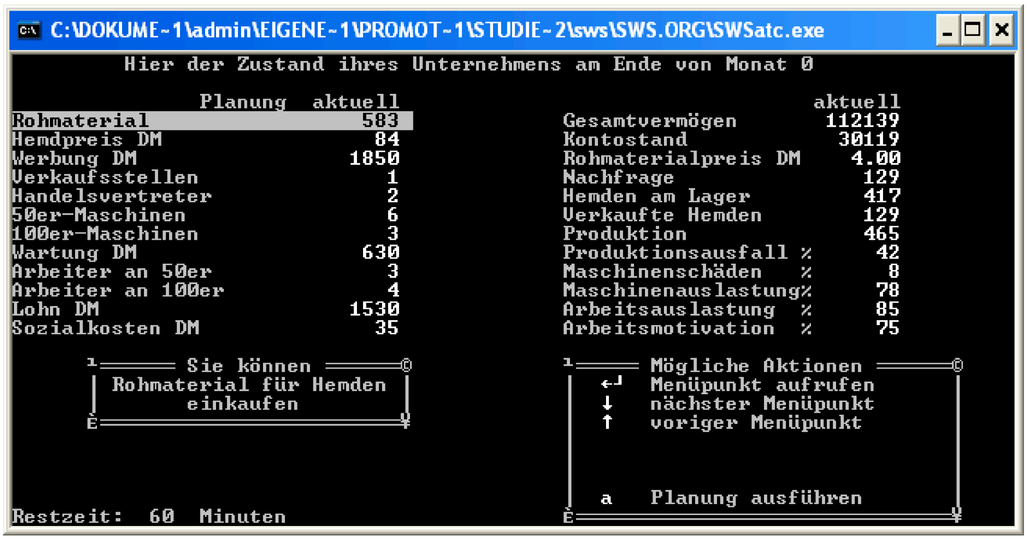
Stellen Sie sich vor, Sie hätten sich für den Managerposten in einer Hemdenfabrik in Berlin beworben und nun tatsächlich die Aufgabe bekommen, dieses unternehmen ein Jahr lang zu leiten. Ihre Leistung wird am Ende der Dienstzeit an der Höhe des Gesamtvermögens gemessen. Das Gesamtvermögen setzt sich zusammen aus dem Geld, das auf dem Firmenkonto bei der Bank liegt (Kontostand) und dem Wert der Maschinen, der Verkaufsstellen, des Rohmaterials und der Hemden im Lager.

Zur Leitung des Unternehmens stehen Ihnen über die Tastatur des Computers verschiedene Eingriffsmöglichkeiten zur Verfügung.

Das Unternehmen befindet sich in einem schlechten Zustand. Ohne Ihre Eingriffe ginge es bankrott. Es bleibt also Ihrem Managertalent überlassen, es vor dem Ruin zu retten.

Auf dem Bildschirm werden Sie eine Zustandsbeschreibung des Unternehmens und eine Aufstellung Ihrer Handlungsmöglichkeiten sehen.

In der obersten Zeile sehen Sie, in welchem Monat Sie sich gerade befinden. Darunter folgt auf dem Bildschirm die Beschreibung des Zustandes Ihrer Firma. An allen Variablen auf der linken Seite können Sie direkte Eingriffe vornehmen:



Sie können Rohmaterial einkaufen, den Hemdenpreis verändern, die Ausgaben für Werbung verändern, Verkaufsstellen neu eröffnen oder schließen, Maschinen kaufen oder verkaufen, und anderes mehr.

Falls Sie Maschinen kaufen oder Verkaufsstellen eröffnen wollen und das notwendige Geld nicht mehr auf dem Konto haben, müssen Sie einen Kredit aufnehmen. Der Kredit ist begrenzt, er hängt ab von der Kreditwürdigkeit Ihres unternehmens. Wenn Sie keinen Kredit mehr bekommen, wird dies am Bildschirm angezeigt. Sie müssen dann versuchen, ohne Neuanschaffungen Gewinne zu erzielen. unbegrenzten Kredit erhalten Sie allerdings weiterhin für die laufenden Kosten, wie Löhne, Sozialausgaben und Werbungskosten, da dies ja keine Neuanschaffungen sind.

Alle Entscheidungen, die Sie in einem Monat treffen, werden zunächst nur als Planungen am Bildschirm angezeigt, ohne dass etwas berechnet wird. Erst wenn Sie der Meinung sind, alle notwendigen Entscheidungen getroffen zu haben und den Monat abschließen wollen, werden die Entscheidungen verwirklicht.

Auf der rechten Seite sehen Sie dann, wie sich die Eingriffe am Ende des Monats auf Ihre Firma ausgewirkt haben. Diese Variablen können Sie natürlich nicht direkt verändern.

Ganz oben sehen Sie das Gesamtvermögen, das am Ende des 12. Monats Ihrer Tätigkeit als Firmenchef möglichst hoch sein soll. Darunter werden weitere wichtige Angaben über Ihren Betrieb angezeigt.

In dem Kästchen unten rechts sehen Sie, welche Tasten Ihnen für Ihre Eingriffe zur Verfügung stehen. Mit den Pfeiltasten nach oben und unten können Sie einen Menüpunkt auswählen. Als Menüpunkt bezeichnen wir jede der besprochenen Eingriffsmöglichkeiten auf der linken Bildschirmhälfte, also Rohmaterial, Hemdenpreis, Werbung usw. Im Augenblick ist der Menüpunkt "Rohmaterial" ausgewählt. Sie erkennen dies daran, dass oben links das Wort "Rohmaterial" durch einen farblich veränderten Hintergrund hervorgehoben wird.

Bevor Sie mit der Simulation beginnen, erhalten Sie noch zusätzliches Wissen darüber, wie man die Hemdenfabrik erfolgreich steuert. Zudem ist in der Graphik oben der Anfangszustand der nächsten Hemdenfabrik dargestellt. Wenn es für Sie hilfreich ist, können Sie diese Graphik beim Lesen des Textes heranziehen, um ihn besser zu verstehen. Wenn Sie sie nicht als hilfreich empfinden, konzentrieren Sie sich ausschließlich auf den Text.

Nach 10 Minuten müssen Sie diese Informationen wieder abgeben. Versuchen Sie sich in dieser Zeit möglichst viel davon einzuprägen. Wenn Sie möchten, können Sie sich beim Lesen Notizen auf dem bereitliegenden Notizpapier machen. Ihre eigenen Notizen dürfen Sie auch während der Aufgabenbearbeitung behalten.

Wenn keine Fragen mehr bestehen blättern Sie bitte jetzt um und fangen an zu lernen.

Im Vergleich zu anderen computersimulierten Aufgaben ist die Hemdenfabrik ein relativ **stabiles System**. Kleine Eingriffe bewirken auch nur kleine Effekte. Man braucht also bei seinen Maßnahmen nicht zu vorsichtig vorzugehen, da die Fabrik nicht so leicht aus den Fugen gerät. Die Bank gewährt der Fabrik großzügigen Kredit. Wenn das Geld gewinnbringend investiert wird, ist es daher unproblematisch, Schulden zu machen. Verkaufsstellen oder Maschinen sollte man also nicht verkaufen, bloß weil man einen negativen Kontostand hat.

um das Gesamtvermögen zu maximieren, ist es einerseits sinnvoll, das System in einen **Gleichgewichtszustand** zu bekommen. Das erreicht man z. B., indem man dafür sorgt, dass sich Nachfrage und Angebot an Hemden bzw. Anzahl an Maschinen und Arbeitern entsprechen. Andererseits sollte man versuchen, den Betrieb zu vergrößern und den Verkauf zu steigern.

Der **Rohmaterialpreis** ist von sämtlichen Systemvariablen unabhängig. Er unterliegt allerdings deutlichen, marktbedingten Schwankungen. Man sollte darauf achten, dass immer ausreichend Rohmaterial vorhanden ist, um die Produktion sicherzustellen. Ein häufiger Fehler besteht darin, das Rohmateriallager nicht wieder aufzufüllen.

Die **Lagerhaltung** ist mit Kosten verbunden. Deshalb sollten nicht zu viele Hemden auf Halde produziert werden. Gleichzeitig sollte aber soviel produziert werden, dass die Nachfrage befriedigt werden kann.

Die Anzahl der fertigen Hemden im Lager und die Menge Rohmaterial im Lager müssen zusammen mindestens so hoch sein wie die Nachfrage. Wenn Angebot und Nachfrage auseinander klaffen, ist es sinnvoller, eines von beiden zu steigern als eines von beiden zu senken.

Die **Nachfrage** bestimmt, wie viele Hemden Sie pro Monat am Markt absetzen können. Die Ausgaben für Werbung erhöhen die Nachfrage, ebenso die Anzahl der Verkaufsstellen und die Anzahl der Handelsvertreter.

Ein höherer Hemdenpreis senkt die Nachfrage. Der Preis hat einen stärkeren Einfluss auf die Nachfrage als jeweils Werbung, Verkaufsstellen und Vertreter. Die Posten Werbung, Anzahl der Verkaufsstellen und die Anzahl der Handelsvertreter erhöhen die monatlichen Kosten. Die monatlichen Kosten werden auch durch Rohmaterial-Lagerkosten, Kosten pro Arbeiter (Lohn- und Sozialaufwendungen) und durch evtl. Sollzinsen erhöht.

Während der verbleibenden Lernzeit steht Ihnen auch eine Hemdenfabrik zur Verfügung, bei der Sie die ersten fünf Variablen Rohmaterial, Hemdenpreis, Werbung, Verkaufsstellen und Handelsvertreter probeweise variieren dürfen. Bitte verändern Sie die anderen Variablen nicht, wir werden dies überwachen.

Anfangszustand der nächsten Hemdenfabrik

Hier sehen Sie den Anfangszustand der Hemdenfabrik, die Sie am Ende der Simulation leiten sollen. Die Variablen, die auch in der Abbildung auftauchen, sind durchnummeriert. Wenn Sie möchten, können Sie sich bereits Maßnahmen für die ersten Monate Ihrer Leitung überlegen.

	Eingriffsvariablen			**Nicht (direkt) beeinflussbare Variablen**	
Nr.			Nr.		
9	Rohmaterial	583		Gesamtvermögen (DM)	112139
	Hemdpreis (DM)	84	11	Kontostand (DM)	30 119
	Werbung (DM)	1 850	14	Rohmaterialpreis (DM)	4.00
	Verkaufsstellen	1		Nachfrage	129
	Handelsvertreter	2	16	Hemden am Lager	417
3	50er-Maschinen	6	7	Verkaufte Hemden	129
3	100er-Maschinen	3	8	Produktion	465
5	Wartung (DM)	630		Produktionsausfall (%)	42
2	Arbeiter an 50er	3	7	Maschinenschäden (%)	8
2	Arbeiter an 100er	4		Maschinenauslastung (%)	78
	Lohn (DM)	1 530		Arbeitsauslastung (%)	85
	Sozialkosten (DM)	35	1	Arbeitsmotivation (%)	75

Appendix B

Additional Tables

Table B.1: Kolmogorov-Smirnov Test for normal distribution of AST measurement variables for Section 8.1.3 (pretest)

	# of Concepts	Edges	Cluster	Weighted Density	Diameter	RMPL
$\bar{x}$	10.79	27.48	4.68	1.08	2.50	.34
σ	4.71	21.04	2.14	.65	.95	.24
Z	.552	.649	.667	.763	1.613	1.511
$p(Z)$	.921	.798	.766	.606	.011	.021

Table B.2: Kolmogorov-Smirnov Test for normal distribution of AST measurement variables for Section 8.1.3 (posttest)

	# of Concepts	Edges	Cluster	Weighted Density	Diameter	RMPL
$\bar{x}$	15.71	53.09	6.38	.97	2.86	.22
σ	7.36	34.37	2.89	.58	1.062	.10
Z	.739	.728	.866	.518	1.220	.585
$p(Z)$	.646	.664	.442	.951	.102	.884

Table B.3: Kolmogorov-Smirnov Test for normal distribution of AST measurement variables for Section 8.1.4

	# of Concepts	Edges	Cluster	Weighted Density	Diameter	RMPL
$\bar{x}$	21.50	50.64	8.14	.60	3.21	.19
σ	13.38	30.55	4.622	.42	1.12	.07
Z	.812	.868	.900	.527	1.352	.36
$p(Z)$	.525	.439	.393	.944	.052	.999

Appendix C

Acknowledgements

I am indebted two organizations for their financial support: The German National Academic Foundation (Studiestiftung des deutschen Volkes) and the Japan International Science and Technology Exchange Center (JISTEC).

Furthermore, I am very indebted to my supervisor Prof. Dr. Wolfgang Scholl who provided help and assistance far beyond the regular scope of PhD supervision. The same applies to Dr. Sarah Spiekermann, who provided me with an optimal environment for my studies at my second supervisor's Institute of Information Systems, Prof. Oliver Günther, PhD. Prof. Dr. Heinz-Jürgen Rothe at Potsdam University has been a wise mentor in the complex field of knowledge elicitation and without the collaboration with his group (namely Petra Ceglarek and Maik Kermas), much of the work I conducted would not have been possible. I am also grateful to Prof. Kozo Sugiyama at the Japan Advanced Institute of Science and Technology (JAIST), where I had the unique opportunity to visit his Lab. I would also like to thank my new employer, Prof. Dr. Heinz Gutscher at University of Zurich, for the ability to finish the thesis at his institute. Finally, this work would have failed without the help from my family, my friends, and my partner Daniel. They provided emotional support during the difficult stages of this project.

I am also indebted to the following individuals for their efforts and support (in alphabetical order): Janina Blume, Petra Bulwahn, Arne Dekker, Prof. Dr. Peter Frensch, Hans-Florian Geerdes, Barbara Haller, Ludmila Herdt, Manuel Hertlein, Christina Karsten, Dr. Uli Klocke, Prof. Dr. Annette Kluge, Dr. Stephan Kolassa, Sebastian Kunert, Tobias Lattke, Conny Lochelfeld, Annemarie Mehle, Goran Milacevic, Anja Rettig, Siegfried Richter, Dr. Barbara Schaunburg, Emanuel Schmider, Tobias Schröder, Marion Schwenke, Michael Sengpiel, Nicole Steckloina, Doreen Stüve, Prof. Dr. Hans-Martin Süß, Sho Tsuji, Michaela Turß, Prof. Dr. Elke van der Meer, Xenia Weißbecher-Klaus, Thomas Wildermuth, and Zhisong Zhang.

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfasst und nur die angegebene Literatur und die angegebenen Hilfsmittel verwendet zu haben.

Zürich, den 04.12.2007

Bertolt Meyer