Toward a Model of Knowledge Structure and a Comparative Analysis of Knowledge Structure Measurement Techniques

Richard Koubek

School of Industrial Engineering
Purdue University

September 1, 1991

Prepared for the Cognitive Science Research Program, Cognitive and Neural Sciences Division, Office of Naval Research, under grant number N0014-90-J-1256. Approved for public release, distribution unlimited. Reproduction in whole or in part is permitted for any use of the United States Government.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 September 1991	Interm 1 July 1990 to 1 June 1991

4. TITLE AND SUBTITLE
Toward a Model of Knowledge Representation and a Comparative Analysis of Knowledge Representation Measurement Techniques

5. FUNDING NUMBERS
N00014-90-J-1256
R&T 4421558

6. AUTHOR(S)
Richard J. Koubek and Daniel N. Mountjoy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wright State University
Dept. of Biomedical and Human Factors Engineering
Dayton, Ohio 45435

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences Program (Code 1142 CS)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This research attempts to develop and validate a proposed model of human knowledge representation.

Based on an extensive literature review, a battery of available knowledge representation measurement techniques was selected to detect the represenation differences between two experience level groups in the domain of clerical work. The techniques employed were card sorting, hierarchical clustering analysis, repertory grid, multidimensional scaling, Pathfinder, and pairwise similarity ratings.

Results validate the existence of all model dimensions. Two dimensions were determined to be affected by experience level. Post-hoc analysis revealed that an additional dimension, Representation Complexity, is a function of experience level differences, and should therefore be included in future model development. Furthermore, the capabilities of the various measurement techniques differed. Specifically, hierarchical clustering analysis was the most effective technique for detecting differences in representations between experience level groups.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Knowledge representation, skilled task performance, expertise	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

ABSTRACT

This research attempts to validate a proposed model of human knowledge structure. An operational definition of knowledge structure was derived which formed the basis for the construction of the proposed model. Until this time, understanding of knowledge structure has been influenced by the output of the various knowledge structure measurement techniques and their associated assumptions.

A battery of available knowledge structure measurement techniques was used in order to detect the structure differences between two experience level groups in the domain of clerical work. The techniques employed were card sorting, hierarchical clustering analysis, repertory grid, multidimensional scaling, Pathfinder, and pairwise similarity ratings. Subjects were required to perform the standard tasks associated with the use of each measurement technique.

Results validate the existence of all model dimensions. Two dimensions were determined to be affected by experience level. Post-hoc analysis revealed that an additional dimension, Structure Complexity, is a function of experience level differences, and should therefore be included in future model development. Furthermore, the capabilities of the various measurement techniques differed. Specifically, hierarchical clustering analysis was the most effective technique for detecting differences in structures between experience level groups.

Further research is needed to refine the proposed model. New knowledge structure measurement methodologies should also be developed in order to provide a more comprehensive examination of the various important parameters of knowledge structure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

As cognitive tasks proliferate in the workplace, it becomes increasingly important to maximize the performance of the workers' cognitive activities. Selection and training are two frequently avenues of achieving higher performance levels. However, in order to develop an efficient selection or training program, it is important to which cognitive functions are critical determinants of task performance. For example, Koubek and Mountjoy (1991) demonstrated that the manner in which an individual structured knowledge of word processing commands affected the performance of a complex text-editing task. In particular, those individuals with an abstract (Hierarchical) knowledge structure were more likely to complete the task than those with a concrete (Alphabetical) structure. Therefore, it appears that in order to maximize performance on any given cognitive task, it would be beneficial to "match" an individual possessing a particular knowledge structure with a task most suited for that structure type.

Work in understanding this phenomenon of knowledge structure may be traced as far back as the research of Tulving (1962) who determined that, given a list recall task, humans tend to order that list according to their structure of the relationships between concepts. Early work examining individual differences in the performance of master and novice chess players (de Groot, 1965; Chase and Simon, 1973) found that master chess players were able to interpret patterns or "clusters" in the meaningful arrangements of chess pieces which allowed them to streamline their memory processes.

These initial experiments spawned a flurry of research which attempted to discover the properties of individual differences in the way people organize domain related information. This work has encompassed a variety of domains such as physics and mathematics problem solving, computer programming, electronics, medical diagnosis, software design, and psychological diagnosis (e.g. Egan and Schwartz, 1979; Adelson, 1981; Shoenfeld and Herrmann, 1982; Murphy and Wright, 1984; Barfield, 1986; Hobus, Schmidt, Boshuizen, and Patel, 1987; Hardiman, Dufresne, and Mestre, 1989).

Researchers attempting to obtain information about the characteristics of knowledge structures have developed a diverse group of techniques which, to some extent, produce a pictorial representation of an individual's knowledge structure. These techniques vary greatly in the information obtained, method of elicitation, and the format with which the structure is presented.

The following section describes the essence of currently available techniques that have been used to measure knowledge structure. From this review, it may be seen that each of the techniques require assumptions regarding the nature of knowledge structure.

2.0 DISCUSSION OF EXISTING TECHNIQUES

All of the techniques discussed here are fundamentally introspective in nature. The techniques differ in terms of how systematic and formal the introspection is, but all techniques begin with individuals making some type of metacognitive judgements (i.e., judgements about what they know). However, many of the existing techniques then perform statistical analyses on these judgements. These analyses may give the investigator the impression that the source of the information is somehow more objective, such as with a physiological or performance-based measure. As a result, it is important for the investigator to keep a proper perspective regarding the origins of the resulting representations. The downfall of the introspectionist paradigm around the beginning of the twentieth century was in part due to the fact that individuals were unable to reliably introspect about their cognitive processes. Because of this, the use of introspection as a data source has developed a negative connotation. While it is acknowledged that the existing

techniques elicit information which is interpretive and high in variability, one should not ignore the utility of that information source. Metacognitive judgments can contain information which is difficult or impossible to obtain from objective sources.

Clearly, the dilemma of introspective sources of information is an issue which must be addressed. Therefore, one may argue that investigations into knowledge structure measurement techniques should not attempt to avoid introspective aspects of the measurement process. There are several common terms that describe, with some degree of accuracy, the type of knowledge elicited by these techniques. First, the knowledge elicited is static (Chi, Hutchinson, and Robin, 1989). Thus, this knowledge is relatively stable over time. Second, the type of knowledge elicited tends to be somewhat more declarative than procedural in nature. While some of the techniques can elicit more procedural types of knowledge, such as in terms of "if-then" rules (Anderson, 1983), the majority of research has tended toward elicitation of knowledge in terms of facts represented by concepts and relationships or common properties between concepts (Rips, Shoben, and Smith, 1973). Third, the existing knowledge measurement techniques provide representations of complex, cognitive domains. As a result of this, each technique may be more applicable to some domains than others.

The following sections (through section 2.3) provide a discussion of the procedures involved with the use of currently available knowledge structure measurement techniques. These techniques have been divided into three categories: verbal reports, clustering methodologies and scaling techniques.

2.1 Verbal Reports

Verbal reports are based on the intuitive notion that, if one wishes to obtain information concerning an individual's knowledge structure, then it is appropriate to have the individual verbally describe the characteristics of his or her domain knowledge. As will be discussed, there are several critical drawbacks associated with the use of verbal reports for knowledge structure measurement. Nevertheless, verbal reports are a popular means of eliciting cognitive information in a variety of domains.

The common elements among the verbal report techniques are as follows. First, all data is generated from verbal descriptions of the domain of interest. Second, the construction of a representation of the knowledge structure is based on the investigator's interpretation of the data. The investigator often applies a particular cognitive model (e.g. GOMS) to organize the verbal data.

The verbal reports section is divided into two stages: information elicitation and representation development. The information elicitation stage discusses verbal techniques which strictly involve the "information gathering" portion of the knowledge structure definition process. The representation development phase describes two models of cognitive task performance that investigators have used as a framework for the results of the verbal reports.

2.1.1 Information Elicitation

2.1.1.1 Interviewing

Interviewing has been the most widely used technique to elicit information from experts in the development of expert systems (Evans, 1988). Individuals are asked a series of questions concerned with the performance of the given problem or job, structured in such a way as to, first, elicit general concepts, and later, to determine relationships among those concepts (Meister, 1985; Olson and Rueter, 1987). Further, investigators use a *funnel* approach in which the line of questioning starts very broad and then becomes more specific as the interview proceeds. Interviewing is a retrospective technique in that the individual being interviewed elicits information subsequent to performance of domain related tasks.

An example of the use of an interview in an area other than expert system development is the work of McPherson and Thomas (1989). These authors used interviews to determine the relationship of declarative knowledge and performance of tennis. Both a situation interview and a point interview were performed in order to assess the current state of tennis knowledge and how this knowledge was employed during an actual tennis match. Means and Voss (1985) applied an interviewing method to determine the amount of depth, or hierarchy, in experts' and novices' knowledge of a movie theme. If the individuals answered the first level question correctly, they were probed for additional information about the particular scenario. Basically, the more information provided by the person, the deeper their understanding of the underlying themes of the film. Leinhardt and Smith (1985) incorporated an interview in their study of the expertise of mathematics instruction. In this case, the interview was used to confirm data collected in other manners.

2.1.1.2 Questionnaires

Questionnaires may also be an effective manner in which to elicit variables and their relationships important to performing a given task. The questionnaire is presented in a written format with open ended questions. Questions are posed in such a way as to allow the individual to identify critical variables, and in addition, the manner in which the variables interact (Olson and Rueter, 1987). Like interviewing, a questionnaire is a retrospective technique.

2.1.1.3 Observation

Perhaps the simplest method to obtain information about task performance is to observe the operator as he/she performs the task in question. In this situation, the observer must take notes and attempt to follow the person's thought process throughout task performance. Observation is called a concurrent technique because the information is elicited while task performance takes place. This technique allows the operator to perform the task without outside interference. The time constraints imposed upon the analyst during the task performance may not allow an accurate description of the task performance. However, if this process were video-recorded, the analyst would have the ability to stop and start the playback in order to provide additional time with which to make precise observations.

Observation methods are generally not well suited towards understanding complex cognitive domains because most of the process of performing the task is internal to the individual. Thus, observer bias is inherent, since judgements concerning the individual's thought processes must be inferred.

2.1.1.4 Interruption Analysis

Interruption analysis, also called probing, is similar in many ways to observation. However, the important difference between the two is related to the amount of inference demanded of the observer. If at any point during task performance the operator's actions or thought processes become unclear to the observer, the latter interrupts the task in order to probe the individual for information regarding the reasoning behind those unclear actions or thought processes (e.g. "Why did you do that?" or "What did you gain from that?").

Like simple observation, this technique allows the person to perform the task without outside interference -- that is, until the observer interrupts the task for additional information. As noted by Olson and Rueter (1987), once the task has been interrupted, it is difficult to resume.

2.1.1.5 Protocol Analysis

Protocol analysis is a further attempt to eliminate inferences drawn by outside sources. In the generation of a protocol, an individual is asked to "think out loud" as he/she performs the task to be analyzed. The individual should identity any goals and methods currently being utilized to reach the task solution. The entire process is videotaped so that a transcript may later be retrieved of the verbalizations and other physical processes which occurred during task performance.

A number of researchers have used protocol analysis to elicit knowledge in a variety of domains. Although the arenas of computer programming, software design, and word processing have been popular domains for the application of protocol analysis (Vessey, 1985; Soloway and Adelson, 1985; Koubek, Salvendy, Eberts and Dunsmore, 1987; Koubek and Salvendy, 1989; Zeitz and Spoehr, 1989; Koubek and Mountjoy, 1991), its application toward the general domain of education and problem solving is also prevalent in the literature (Chi, Feltovich and Glaser, 1981; Sweller, Mawer, and Ward, 1983; Leinhardt and Smith, 1985; Sweller and Owen, 1985; Lunderberg, 1987; Ploger, 1988). The use of protocol analysis in these domains is reasonable because the thought processes may be verbalized quite naturally. In other domains, in which the learning process is less verbal (such as playing golf), protocol analysis is much less effective (Cooke and McDonald, 1986).

This technique, according to Ericsson and Simon (1980), is a valid means of obtaining an individual's momentary cognitive processes since the individual's immediately preceding thoughts are stored in short term memory, and can be accessed by the individual directly. However, some investigators are concerned that task performance is altered during the verbalization process, and would, therefore, produce unnatural protocols (Ericsson and Simon, 1984; Musen, 1989; Nisbett and Wilson, 1977).

2.1.1.6 Discussion of Verbal Report Techniques

In past years, verbal reports have generally been labeled as introspective in their generation of data (Nisbett and Wilson, 1977), and, therefore, have been discarded by many as being valuable tools of knowledge elicitation. However, Ericsson and Simon (1980) argue that some of these techniques may be unfairly categorized as highly introspective (e.g. protocol analysis).

As is true with the application of any knowledge elicitation technique, strict guidelines must be followed in order for the output of verbal reports to be valid. Ericsson and Simon (1984) have identified three criteria that must be met to insure the proper use of verbal reports. The first of these is the "Relevance Criterion" which, basically, states that verbalization should be a normal part of the performance of the given task. The "Consistency Criterion" explains that consecutive verbalizations should be logically consistent. Finally, the "Memory Criterion" requires that the person remember a subset of the information that was attended to during task performance. Ericsson and Simon (1980, 1984) contend that, if these guidelines are followed, verbal reports can produce useful, reliable data for the study of cognitive processes.

2.1.2 Representation Generation

Once the verbal data has been collected, it must be transformed into an analyzable format. This process is neither easy, nor well defined, for most of the techniques previously described. Two techniques which do have guidelines to follow in the structuring of verbal data (GOMS models and problem behavior graphs) are described below.

2.1.2.1 GOMS Analysis

As described by Card, Moran and Newell (1983), one's "...cognitive structure consists of four components: (1) a set of Goals, (2) a set of Operators, (3) a set of Methods for achieving the goals, and (4) a set of Selection rules for choosing among competing methods for goals" (p. 140). When using the GOMS model, the transcript obtained from the knowledge elicitation stage must be transformed from natural language to a transcript which consists solely of the previously described components. At this point it is easy to see that, during the course of elicitation, it is imperative that the operator verbalize all goals and methods being utilized so that an accurate representation of knowledge is possible.

Once the GOMS model has been completed, a comparison between two or more representations becomes easier to accomplish, since all representations are now in identical formats. However, the analysis process of representation differences has not, historically, been well defined. One common method has been pure, subjective opinion on the part of the analyst. This method, of course, does not permit a quantitative comparison of separate solutions, but will sometimes reveal aspects of the degree of hierarchy or linearity in a problem solving process. Others have attempted to quantify certain aspects of the GOMS model in order to perform a more empirical evaluation of the similarities between solution sets (Koubek et al., 1987; Koubek and Mountjoy, 1991).

2.1.2.2 Problem Behavior Graphs

The problem behavior graph is usually constructed based on the results of a verbal report (Newell and Simon, 1972). Problem behavior graphs allow the analyst to view individuals' thoughts as they progress through the solution set (see Figure 1). Each node (or box) in the graph represents the state of knowledge at that particular time. Arrows between the nodes represent the application of an operator to the previous state of knowledge, which results in a new state of knowledge. Time runs left to right, then top to bottom. When the individual returns to a previous state of knowledge (X), the new node is placed directly below X (since it occurred later in time) and is connected to X by a vertical line. If an operator is repeatedly applied to the same state of knowledge, a double arrow is placed between the two nodes.

The resultant graph provides a spatial representation of the individuals' task performance. This allows an easier (subjective) interpretation of the degree of hierarchy or linearity in the solution than does a strict GOMS model. In addition, the steps taken by the individual in order to reach a particular state of knowledge are uncomplicated to follow. Though, beyond these differences between a GOMS representation and a problem behavior graph, the analysis of the latter faces the same difficulties as the GOMS model: ill-defined analysis methods.

2.1.2.3 Discussion of Representation Generation Techniques

As can be seen from the previous descriptions of representation development techniques for verbal report data, much subjectivity is left to the analyst. Even if an attempt is made to quantify certain aspects of the representations, the ultimate decision as to which aspects are to be quantified is made by the analyst. This may tend to bias the results toward revealing the kind of information that the analyst wishes to obtain. However, if care is taken in the selection of the quantifiable aspects, as well as in the experimental design, interesting relationships between individual characteristics may be found.

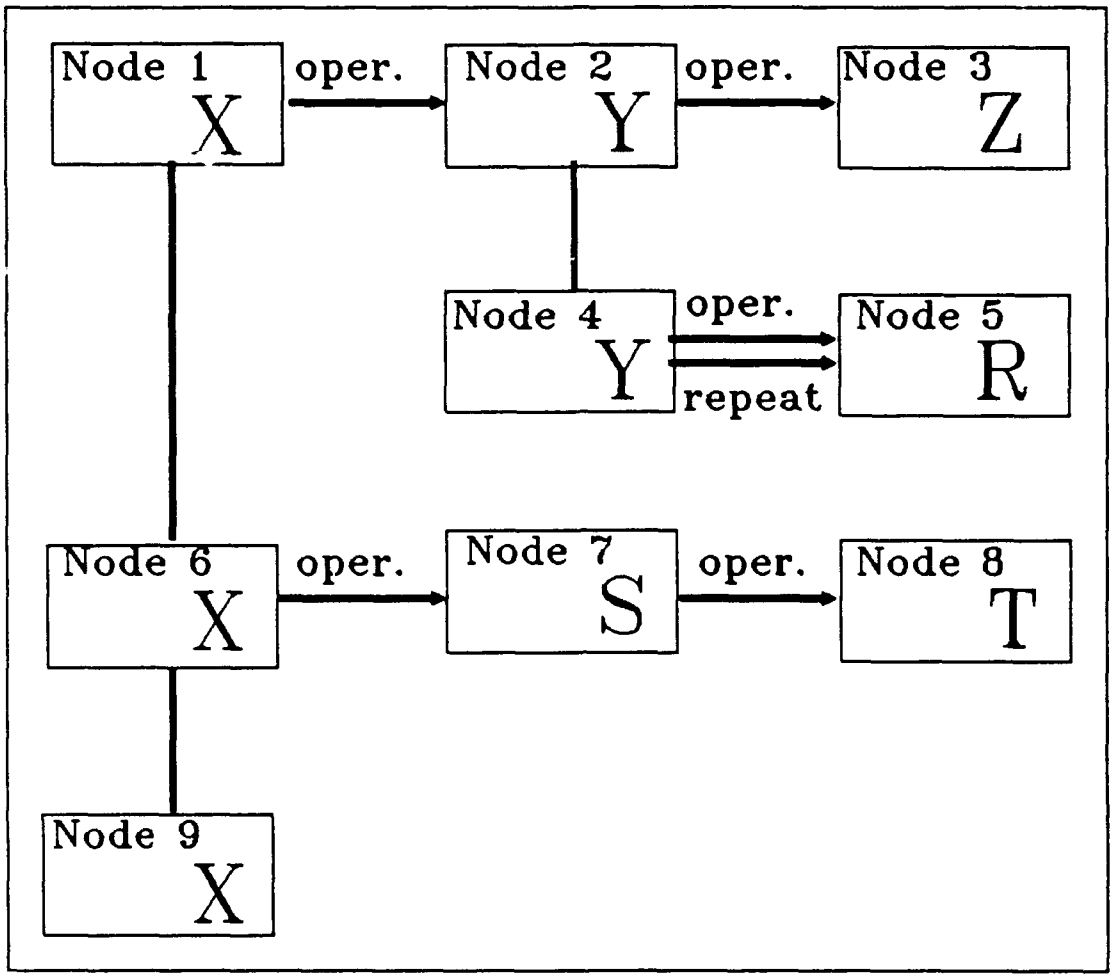

Figure 1. Example of a problem behavior graph.

2.2 Clustering Methodologies

The clustering methodologies can viewed as a midpoint between the scaling methodologies (see section 2.3) and the verbal reports that were previously discussed. The clustering techniques are more formal and systematic than the verbal reports and less so than the scaling methodologies. The important difference between the clustering methodologies and verbal reports is that, with verbal reports, the analyst produces the representation of the elicited knowledge based on their subjective interpretation of the data. With the clustering methodology, the representation is produced directly by the person (e.g. by having them perform a series of tasks) or is analytically derived from the results of a task. Thus, the investigator has substantially less influence on the characteristics of the structure obtained.

The foundations of the clustering methodologies began with the early work in expert-novice differences (de Groot, 1965; Chase and Simon, 1973; Reitman, 1976; Egan and Schwartz, 1979). This research suggested that domain-relevant concepts are grouped together in *clusters* by those individuals proficient in the domain of interest. Further, these clusters were viewed as the building blocks, from which, more elaborate knowledge structures might be constructed. De Groot (1965) and his colleagues examined the performance of master and novice chess players. The major finding of their work was that master chess players showed a superior ability to recall

chess board configurations only when the boards were arranged in actual game situations as opposed to random arrangements of the pieces. This result suggested that expert chess players perceived patterns (e.g. offensive and defensive groups) in the chess piece configurations. Indeed, the existence of a superior *knowledge structure* or problem representation seemed to increase the ability of a master chess player to efficiently recall large amounts of relevant information. Chase and Simon (1973) used a recall technique to establish boundaries of these cognitive clusters in chess board configurations.

The following techniques may counter some of the problems inherent in verbal report techniques and provide an alternative to the less formal and objective techniques mentioned above. However, these techniques still rely heavily on verbal report techniques to provide the initial sources of information. Indeed, all of the methods below essentially begin with techniques that are not far removed from the more traditional techniques of interviewing and protocol analysis.

Clustering methodologies are divided into three main stages: concept elicitation, cluster elicitation, and the analysis of representations. This division of concept list elicitation from the elicitation of the clustering relationships is one of characteristics that distinguishes clustering methodologies from verbal reports.

2.2.1 Concept Elicitation

The elicitation of a list of domain relevant concepts is an often underemphasized aspect in the generation of a representation of human knowledge structures. For example, if the derived list is missing several key concepts, the resulting representation may be significantly different from a representation which contained these concepts. Alternatively, if the list contains inappropriate concepts, these concepts may distort the desired representation. To date, few formal techniques have been created to systematically extract the important domain-related concepts (Cooke and McDonald, 1987).

One of the primary bottlenecks to the elicitation of domain relevant concepts is the lack of a well specified definition of a concept. In Cooke and McDonald's (1986) study of concept elicitation in the domain of automobile driving, they found that the concept list obtained was far from a homogenous sample of terms. Instead, they obtained a variety of phrases and words at different levels of abstraction and detail. The appropriate operational definition of a concept should include some specification of the depth of analysis that the investigator may obtain as well as domain dependent information. There are several terms used in the literature to describe the terms that are produced in concept elicitation: elements, concepts, objects, terms, nodes.

The techniques given below are, for the most part, borrowed from the verbal report techniques such as interviewing and protocol analysis. However, concept elicitation techniques differ from the verbal reports in two important ways. First, the format of the resulting information in concept elicitation is determined. Indeed, the output of these techniques is required to be a list of domain relevant concepts. This is in contrast to verbal reports, where the characteristics of the output are to a large degree uncertain. Also, verbal reports attempt to obtain all important knowledge concerning the domain simultaneously, whereas concept elicitation only generates domain relevant concepts.

2.2.1.1 Investigator Judgement

In the majority of studies examined in the literature, investigators obtained the concept list from a source other then the individuals under study: researcher's intuition or written technical information on the subject. This approach to concept elicitation would certainly seem to be counter productive because the information does not originate from the individuals being measured. Nevertheless, an investigator may be compelled to use this technique because it may

reduce costs substantially. Further, in domains such as computer programming or problem solving, a list of concepts is often previously available.

Cooke and Schvaneveldt (1988) selected programming concepts from the chapter headings of an introductory text on the subject. McKeithen, Reitman, Rueter, and Hirtle (1981) used single syllable ALGOL reserved words as the concept list. Chi, Feltovich, Glaser (1981) selected physics problems from a physics reference for use in their experiment. Adelson (1981, 1984) subjectively selected computer programs and flowcharts to present to novice and expert computer programmers. Hollands and Merikle (1987) selected psychological terms from several psychology texts and references. In these studies, the use of the investigator judgement technique apparently did not cause subsequent difficulties.

2.2.1.2 List Generation

This technique involves the manual listing of a set of domain relevant concepts by an individual or group. Cooke and McDonald (1986) describe three different ways in which domain critical concepts may be elicited in list form: critical concept listing, step listing and an outline generation method.

In critical concept listing, the individual is simply asked to list as many domain relevant concepts as possible. In step listing, the individual is prompted to list concepts related to a particular task that they have observed or performed. In the outline generation method, the individual is asked to list the headings and subheadings of a hypothetical book on the domain of interest. Rips, Shoben, and Smith (1973) used a technique similar to critical concept listing to elicit familiar bird and mammal terms. The 12 terms that were most frequently mentioned by students in a five minute period were selected.

2.2.1.3 Interview and Protocol Analysis

Cooke and McDonald (1986) also discussed and investigated an interviewing method. In their method, an individual observes an interviewing scenario in which one person asks another person open ended questions about the pertinent domain. Similarly, in protocol analysis, an individual would observe a recorded protocol of a domain relevant task and would be asked to record all significant words mentioned. These techniques, of course, are similar to the corresponding techniques in the verbal reports section. Schvaneveldt, Goldsmith, Durso, Maxwell, Acosta and Tucker (1982) used an iterative series of literature searches, task analyses, and interviews with fighter pilots to obtain a concept list for two tactical flight maneuvers.

2.2.1.4 Discussion of Concept Elicitation Techniques

The trend in the literature is for the list of domain relevant concepts to be much smaller than the actual size of the concept list that an expert possesses. To some extent, this occurs because of practical limitations on budgets and time, but also occurs because researchers often believe that a partial list of concepts is sufficient for the applications of the analysis.

Studies have found that different methods of concept elicitation tend to obtain certain types of information. Cooke and McDonald (1986) compared several concept elicitation techniques (concept lists, interview, task-based lists, chapter lists). For the domain of "driving a car", they found that concept listing and task-based listing generated mostly general rules ("Wear seat belts", p. 1428) while interviewing and chapter listing revealed mostly concepts ("brakes", p. 1428). They argued that different techniques are differentially suitable to obtain various types of information.

Typically, a person or group of people who are recognized as experts in the domain are asked to provide a set of domain relevant concepts. Less expert individuals are rarely asked to perform this task. The reasoning behind this strategy is as follows. Given a common set of

concepts, that were provided by experts, the individuals with greater expertise should be more able to appropriately indicate relationships between these concepts. The less expert individuals will, conversely, be less able to appropriately explain the relationships between the list of concepts.

To produce an adequate concept list, investigators must make a decision regarding the desired depth of analysis and closely examine the domain of interest to produce an operational definition of a concept which can guide the decision of an appropriate technique.

2.2.2 Cluster Elicitation

A cluster elicitation technique takes information in terms of a list of domain related concepts, establishes discrete or continuous relationships between the concepts, and outputs a spatial presentation of these relationships. Partial or similarity-based relationships are not produced explicitly by these techniques. For example, in cluster elicitation, the individual may make judgements about which concepts *go together*, but two concepts cannot be partially clustered.

The techniques described in this section can be divided into three subcategories based on the complexity of their resulting representations: single-level clustering, hierarchical (nested) clustering, and overlapping or complex clustering. In single level clustering, several clusters of the concepts are produced with no nested or overlapping clusters. Hierarchical clustering allows nested clusters but not overlapping clusters. Finally, complex clustering allows nested and/or overlapping clusters (see Figure 2).

At this point, it is also useful to mention that two types of clustering representations exist: the tree perspective and the closed curve perspective. Figure 2 provides an illustration of both clustering perspectives versus the levels of clustering complexity.

2.2.2.1 Card Sorting

Card sorting is a cluster-based technique. The term cluster-based refers to the idea that the technique tends to elicit information which determines how concepts are divided into groups rather than information about the psychological proximity of each concept to each other. While methods to derive similarity data from this technique exists, this information is obtained indirectly and may involve some significant assumptions. All studies identified which use this technique produced single-level representations.

The basic idea behind the card sorting technique is somewhat self explanatory. Individuals manually sort all of the concepts simultaneously. Each concept in the concept list is labeled on a card. The analyst presents all of the concepts to the person. Typically, the individual is instructed to divide the cards into groups based on which concepts "go together" (Gobbo and Chi, 1986, p. 224). Most frequently, there is no restriction placed on the number of groups which can be created. Further, duplicate concept cards are often encouraged (McDonald, Paap, and McDonald, 1990) so that a concept may appear in more than one group. This increases the variety of types of relationships which can be elicited. Often in card sorts, individuals are encouraged not to sort concepts that they are not familiar with. Weiser and Shertz (1983) had individuals sort programming problems and compared the labels individuals given to the clusters across groups.

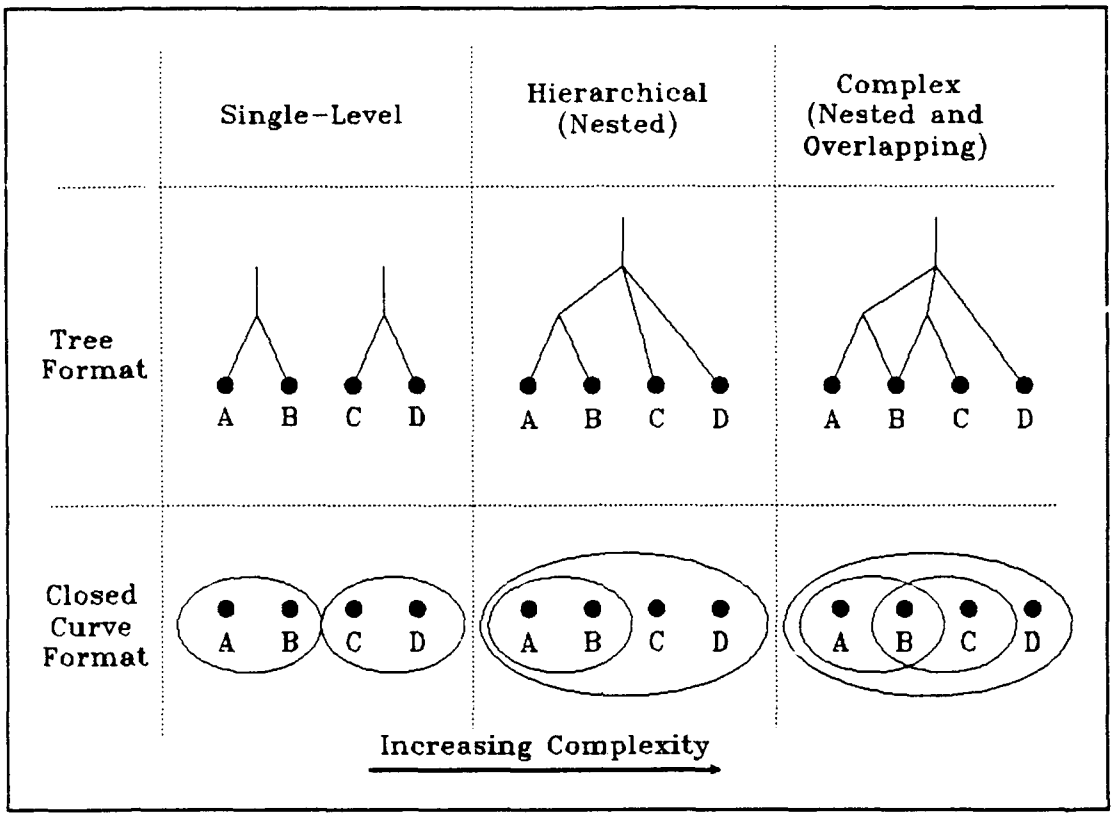

Figure 2. Examples of cluster types.

2.2.2.2 Ordered Trees

The basis of this technique, states that the way in which concepts are stored in an individual's long-term memory is reflected by the order they are recalled. Specifically, it is proposed that humans store and recall all items in a particular cluster before recalling items in another cluster (Olson and Rueter, 1987). Given the assumption that humans store information in chunks or "clusters" of concepts, attempts have been made to examine the recall process to identify the boundaries of these clusters.

Work examining the relationship between knowledge structure and the recall process was one of the first techniques used to try to measure knowledge structures. Chase and Simon (1973) had chess players reconstruct chess board configurations from memory to determine clusters of chess pieces. They argued that the response latencies between clusters of chess pieces should be longer than latencies within a cluster. Tulving (1962), McLean and Gregg (1967) and Bower and Springsteen (1970) also used inter-response latency information to derive cluster-type relationships for a set of concepts.

There were, however, several concerns with this technique. Reitman (1976) attempted to replicate Chase and Simon's study using another board game, Go. She found that the error variation in inter-response latencies was so large that she was unable to make useful conclusions about the boundaries of the clusters. Further, researchers had, at that time, begun to believe that clusters were often arranged hierarchically (Chase and Simon, 1973). This technique was not well suited for identifying nested clusters.

Reitman's dissatisfaction with the use of inter-response latencies led to a new technique which focuses on the sequence of concepts in the recall as opposed to temporal aspects (Reitman

and Rueter, 1980). In this technique, individuals were asked to repeatedly recall the concept list. Prior to recall trials, the individual completely memorized the concept list. On each trial, a different one of the concepts was presented as a cue word to initiate the recall. This encouraged the individual to base recall on their internal organization of the concepts rather than on the memorization of one sequence. An algorithm, developed by Reitman and Rueter, identified strings of concepts which were always adjacent to each other, regardless of order, and clustered them together. This technique is capable of identifying hierarchical relationships (but is also limited to the hierarchical format). The result of the algorithm is a spatial, hierarchical representation of the clusters (see Figure 3). McKeithen et al. (1981) used this technique to examine the knowledge structures of computer programmers.

A distinctive aspect of this procedure is that it can identify patterns of sequences in a particular cluster. This technique is often called *ordered trees* because of the sequence patterns developed in the hierarchical clusters. Figure 3 gives an example of an ordered tree for word processing commands, where an arrow with a single head indicates that the concepts in that cluster were always recalled in that sequence and in the direction that the arrow head points, while a two headed arrow indicates that concepts were recalled in either direction.

2.2.2.3 Closed Curve Analysis

The method of closed curves allows one to discover the relationships between objects which are represented spatially (Olson and Rueter, 1987). Traditionally, closed curves have been used to determine differences in the chunking ability of novices and experts in various spatial domains such as board games and circuit analysis (Reitman, 1976; Egan and Schwartz, 1979).

The application of the closed curves technique is straightforward (e.g. "Circle everything that goes together"). Thus, this task may be intuitively easy for individuals to perform. The subject matter for a closed curve technique may be any graphical representation of domain information, such as a plant layout or interface display panel. Figure 2 (closed curve perspective) gives several examples of how a person may encircle various elements.

In the case of Reitman (1976), the master Go player was shown a game board configuration which he had reproduced from memory at an earlier time. The master was then asked to circle the partitioning which he saw in those patterns. He was instructed that the game pieces should be grouped in small subpatterns, then enclosed in higher level patterns in order to indicate the functionality of the groups on several levels. Reitman discovered that the master Go player tended to chunk game patterns in overlapping clusters, not only as separate chunks or nested hierarchies. In addition, she found that the master Go player produced very reliable closed curve representations.

In another example, Egan and Schwartz (1979) asked a skilled electrician to circle functional units of a circuit diagram. In addition, the expert was asked to provide a verbal label for each functional group. The functional units were permitted to overlap or to be nested, dependent on how the expert understood the circuit to operate. Skilled electricians grouped schematics in functional units (e.g. amplifiers, feedback networks, filters and rectifiers), whereas novice electricians grouped identical diagrams in haphazard manners.

This method is unique in that it allows the individual to indicate chunks directly. In addition, closed curves is the only cluster elicitation technique (with the possible exception of card sorting using duplicate cards) which produces overlapping clusters representations as well as nested cluster representations.

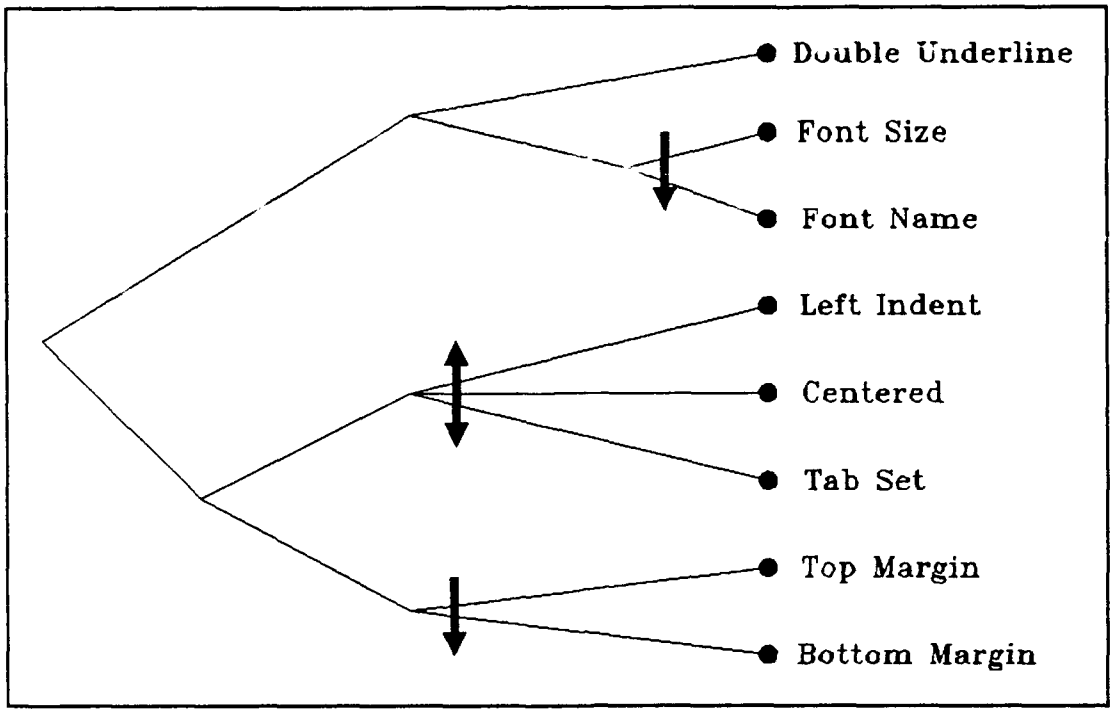

Figure 3. Hierarchical representation format of word processing commands.

2.2.2.4 Spatial Reconstruction

In this technique, individuals are asked to reconstruct an existing spatial system, such as a board game or circuit board. As the individual performs the reconstruction, the investigator collects information regarding the elements that are placed in the work space between glances back to the reference system. Elements placed between two glances are claimed to be chunked together.

Chase and Simon (1973) asked chess players of varying levels of experience to reconstruct mid- and end-game chess configurations on an adjacent chess board with the target board in plain view. The investigators recorded the order of reconstruction and the chess pieces placed between glances to the target board. They also recorded the between and within glance latencies for placement of each chess piece. Chase and Simon argued that the master chess player structures the chess configuration into patterns or "chunks" (p. 57). They hypothesized that the pieces placed between glances reflect the individual's interpretation of a structural relationship between those chess pieces. Spatial reconstruction produces a single level clustering representation. More complex relationships are not likely with this technique.

This technique is limited to spatial domains, such as board games, where the individual applies a structure to an existing configuration. There is no reasonable analogy of this technique to more symbolic domains. Because of the spatial domain restrictions, there is no concept elicitation step for this technique.

2.2.2.5 Hierarchical Clustering Schemes (HCS)

The final clustering technique to be discussed is hierarchical clustering schemes (HCS). This technique differs from those previously mentioned, in that the previous techniques allow the individual to directly produce the knowledge structure, while HCS generates the knowledge

structure through the use of an outside algorithm.

Hierarchical clustering analysis basically takes similarity data (to be more completely discussed in section 2.3.1.1) and converts that information into a cluster-type format. The cluster representation which results helps reveal to the researcher which concepts are most alike as well as most different.

Johnson (1967) is generally credited with the development of a particular set of hierarchical clustering schemes that are popular in the cognitive measurement literature.
It should be noted that the clustering process described by Johnson is not the only clustering scheme. In fact, Johnson's HCS is only one of many clustering methods currently available. Romesburg (1984) is one of several publications which describe many of the popular clustering techniques.

HCS has been used extensively in the literature. Once again, computer-oriented knowledge has been a popular domain for the application of clustering schemes. Adelson (1981) used HCS techniques to discover the differences in the way that expert and novice programmers organize programming concepts. The same type of study was also performed by McKeithen et al. (1981). Kay and Black (1984) employed HCS as a method to determine the changes in knowledge structure as novices developed into experts in text-editing. Other areas of application have included the work of Hopkins et al. (1987) which looked at the understanding of relationships between features of the human cardiovascular system, Schoenfeld and Herrmann's (1982) study of novice/expert perception of mathematical problems, and Schvaneveldt et al. (1982) who were concerned with the organization of critical flight information in memory.

The input to this technique is a half matrix of similarity judgements which are converted into measures of distance between each pair of concepts. The first step in the clustering process is to combine pairs of concepts that are closest in distance into a single, new concept. The value of that distance then becomes the point on the "tree" where the two concepts are joined. The new inter-concept distances are then computed by one of three methods: the minimum, maximum or average algorithm. A full description of the average method (more correctly called the *unweighted pair-group method using arithmetic averages*) may be found in Romesburg (1984). Olson and Rueter (1987) also describe the minimum and maximum methods. In any of the three cases, a new distance matrix is formed and the process is repeated until only one value exists in the matrix. This is the distance at which all concepts are assumed to be similar. The tree which results from the entire process is able to show the structure of concepts and their perceived similarities for a given individual or group.

2.2.2.6 Discussion of Cluster Elicitation Techniques

As a whole, the clustering methodologies give the individual being measured a maximum amount of control in deciding what the representation will look like. It is possible for all of the clustering elicitation techniques to elicit cluster labels from pertinent individuals after the clustering representation is obtained. Several investigators have asked individuals to suggest labels for the clusters in a representation. Egan and Schwartz (1979) obtained verbal labels for closed curve clusters of electronic symbols. McDonald et al. (1990) prompted individuals for cluster labels of a representation of a Unix communication commands. These cluster labels may produce additional concepts of higher levels of abstraction or relationships between concepts which can be used to further pursue analysis.

Clustering methodologies inherently have problems because of the discrete nature of clusters. There are often occasions when it is neither appropriate to totally group or separate a pair of concepts. Those concepts that are weakly related may cause some amount of uncertainty on the part of the domain expert regarding whether to cluster the pair or not. Unfortunately, the clustering methodologies do not provide for this possibility. The scaling methodologies (section

2.3) are better at representing relationships that are partial or uncertain. The relative importance of representing absolute or partial relationships is domain dependent.

2.2.3 Analysis of Representations

Upon review of the literature, a formal quantitative analysis methodology for these representations was not readily apparent for nested representations (Note: single level representations with overlapping may be converted to distance data using co-occurrence measures. See section 2.3.1.3). Due to this, the investigator or domain expert usually subjectively interprets the resulting representations.

2.3 Scaling Methodologies

The scaling methodologies can be logically subdivided into four stages: concept elicitation, relationship elicitation, representation development, and representation analysis. Concept elicitation involves the elicitation of a list of domain relevant concepts, as discussed above. This is followed by relationship elicitation in which the interrelationships between the concepts are determined.

This data is then submitted to a scaling algorithm in the representation development phase to produce a pictorial representation. Finally, this representation is examined in the representation analysis stage. Here, an investigator attempts to draw quantitative conclusions about the representation.

There are cases, such as the repertory grid technique, where the concept and relationship elicitation stages may be simultaneous. Indeed, one might argue that the separation of concept and relationship elicitation is an unnatural and constrictive way for humans to elicit structural information. However, the separation of these steps allows for a more orderly analysis of the system. The separation of these two stages is one of the important features that distinguishes the scaling and clustering methodologies from verbal reports. Of course, by applying a structure on the human knowledge system, inevitable biases and limitations are produced.

There are further ramifications of separating the elicitation of concepts and relationships which should be discussed at this juncture. Certain applications of this work (e.g. personnel selection and training) emphasize the consideration of individual differences in human knowledge structures. The approach that has been taken in the literature is to obtain a common set of concepts, then apply a particular relationship elicitation technique. This method reduces individual difference variance in the data analysis because each person does not have a unique list of concepts which must be assimilated into the analysis. This also greatly simplifies the combination of representations from different individuals in order to obtain a generalized representation of a particular knowledge structure.

2.3.1 Concept and Relationship Elicitation

The input for all of the following techniques (with the possible exception of list generation) is a list of domain relevant terms obtained from the concept elicitation stage. However, the format of presentation of the concepts differs among the techniques. The output of these techniques typically entails a set of numbers, in the form of a matrix, in which each number describes the psychological distance, or similarity, between a pair of concepts. At this point, the difference between a distance matrix and a similarity (or proximity) matrix should be made clear. All representation development techniques discussed here can accept a distance matrix as their input. On the other hand, it is more common in the literature to obtain estimates of relatedness, "similarity," between concepts. These similarity data, in the form of a proximity (or similarity) matrix, can be converted to distance information prior to its input to the representation generation procedures.

2.3.1.1 Pairwise Similarity Ratings

The technique of pairwise similarity ratings is based on the hypothesis that, given an individual's judgement or rating of the conceptual similarity of (or distance between) two concepts, this rating is related to the psychological distance between the two concepts in the person's memory. In this technique, individuals supply a similarity judgement for every possible pair of concepts.

Pairwise similarity ratings take the "microscopic" approach with regard to knowledge elicitation. In this technique, the individual makes judgements regarding the relationship between two concepts at a time. The rational behind this approach is that, by partitioning the elicitation process into manageable portions, the individual is not overwhelmed by trying to explain the entire domain simultaneously.

The pairwise similarity rating technique has been one of the most commonly used techniques for the scaling methodologies. Pairwise similarity ratings have been integrally connected to multivariate analysis techniques in domains other than cognitive measurement. Similarity ratings are the typical inputs to these quantitative techniques.

Schvaneveldt et. al. (1982) and Schvaneveldt, Durso, Goldsmith, Breen, Cooke and De Maio (1985) used pairwise similarity ratings in the context of flight maneuvers of pilots and pilot trainees. Cooke and Schvaneveldt (1988) implemented pairwise similarity ratings on abstract programming concepts. Koubek and Mountjoy (1991) used pairwise similarity ratings in the domain of word processing. Hopkins, Campbell, and Peterson (1987) obtained judgements of the relative predictability of values and properties in a heart vessel system. Many other instances of the use of pairwise similarity ratings are found in the literature (e.g. Enkawa and Salvendy, 1989; Esposito, 1990; Dayton, Durso, and Shepard, 1990).

In this technique, pairs of terms are presented, one at a time, to an individual. The individual is prompted to make a judgement regarding the conceptual "similarity" of the two concepts. The exact instructions given may vary from study to study. This judgement most often takes the form of a rating on a numeric scale (typically, 1-7 or 1-9). Thus, a high rating would describe two concepts that are psychologically "close together". Other formats include having the individual mark a position on a line. These ratings would be coded by measuring the position in millimeters or some other increment. Schiffman, Reynolds and Young (1981) argue that marking a continuous line is a more appropriate method because they feel many individuals are uncomfortable partitioning the scale into a integer scale. In spite of this, using a numeric scale lead to simpler data collection.

The resulting ratings may be converted into a distance matrix by subtracting the given rating from the maximum rating. Although this transformation is acceptable for ordinal or interval data types, research should be done to verify that this transformation maintains the psychological validity of the original similarity ratings.

There is another similarity rating technique that does not involve pairwise comparisons of concepts. In this technique, each concept is compared to all of the other concepts in the concept list. Thus, the other concepts may be sorted or clustered into groups of similar levels of similarity to aid the process (Schiffman et. al., 1981; Gammack, 1990). Hopkins et al. (1987) used this technique to obtain predictability judgements of the properties of the heart.

2.3.1.2 Repertory Grid

A central contribution of the repertory grid approach is the idea that concepts may be related on a variety of dimensions. As an alternative to pairwise similarity ratings where all dimensions of comparison are assumed to be lumped into one judgement, repertory grid allows the researcher to obtain a clear idea of the basis by which similarity judgements are made.

One of the fundamental ideas behind the repertory grid approach is the idea that the

similarity of two concepts can be compared by noting the relative placement of the concepts on a dimension selected by the individual. For instance, when concepts are judged by a person to be far apart or close together on a particular dimension, proponents of this technique might argue that we have obtained some information which partially determines the similarity of the two concepts.

The repertory grid technique is based on personal construct theory proposed by Kelly (1955) for clinical psychology applications. Boose (1985, 1986) provides a thorough explanation of the development of personal construct theory for applications in measuring knowledge structures. Boose developed an automated knowledge acquisition tool, the Expertise Transfer System (ETS), which uses the repertory grid approach. Colthart and Evans (1981) used the repertory grid technique to elicit knowledge structures concerning bird taxonomies.

Several authors who discuss repertory grid technique (Boose, 1985; Olson and Rueter, 1987) discuss that the technique includes an initial session in which a list of domain-related concepts are produced (i.e., concept elicitation). Incremental interviewing techniques have been used to enrich the concept elicitation process in conjunction with repertory grid (Keen and Bell, 1981). Shaw (1980) used feedback techniques in an automated format to further enhance elicitation.

After the concept list is derived, three domain-relevant concepts are randomly selected from the list. These three concepts are presented to the individual. The individual is then asked to state a dimension (or construct) which distinguishes any two of the concepts from the third. For example, if the person were given the three concepts "hail", "rain", and "snow", the individual might conclude that the concepts can be discriminated based on the physical state of the water in each instance. Using this dimension, rain would be distinguished from hail and snow because rain is water in liquid form, and hail and snow can be thought of as water in a solid state. Thus, the person might select the bipolar dimension "liquid versus solid state".

The above process is repeated for different combinations of three concepts until the analyst determines that a representative set of dimensions has been collected. At this point, the set of elicited concepts and dimensions are assembled in a grid format with concepts listed along the horizontal and bipolar dimensions listed along the vertical. Figure 4 provides an example of such a grid comparing several industrial air filtering systems on a variety of dimensions. In recent applications (Boose, 1985), each bipolar dimension is treated as a scale (e.g. 1-5 or 1-7). In this case, the analyst must select one end of the dimension to be the low rating and the other to be the high rating. This is often not an obvious task (e.g. wet versus dry). Individuals are then asked to rate the concepts on each dimension.

There are several techniques that are used to convert rating grids into similarity matrices. The most common technique involves using the difference with which two concepts fall on each dimension as a measure of their similarity. Often, the average difference of two concepts across all dimensions is used as a general measure of the similarity of (or distance between) two concepts.

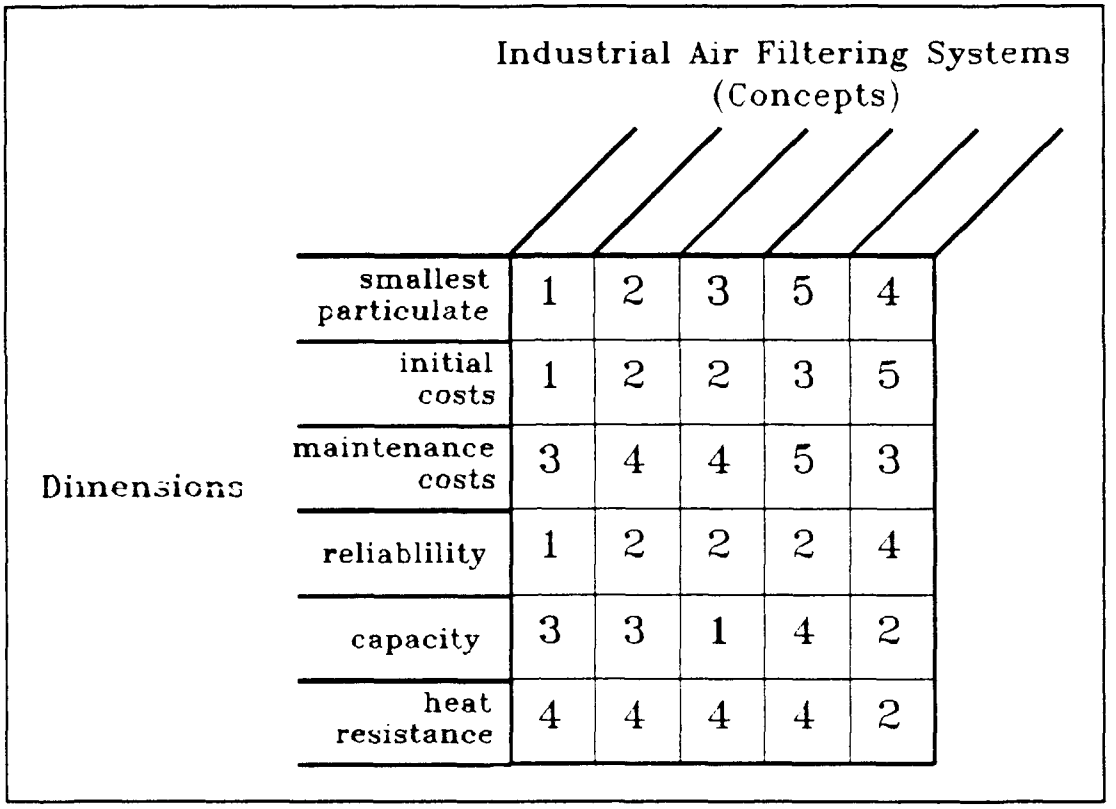

Figure 4. Example of a repertory grid.

2.3.1.3 Co-occurrence Analysis

Co-occurrence is a statistical term that is a measure of the likelihood that two concepts will appear together. This section describes the use of co-occurrence measures of concepts in repeated card sorts as a measure of concept similarity. The most commonly used measure of co-occurrence is conditional probability.

The actual task involved here is identical to the card sorting task mentioned in the clustering methodology section. Here, however, several methods are described which convert single iteration card sorts into a proximity matrix.

There are two variations on how the card sorts may be converted into a proximity matrix. These variations are based on whether a group or an individual are being measured. Using a group of people, one can measure the conditional probability that two concepts will be clustered together across all of the individuals' card sorts. Alternatively, one may have a single person complete many card sorts, using the same list of concepts, then compute conditional probabilities across trials. Other statistical measures of concept co-occurrence have also been investigated (McDonald, Plate, and Schvaneveldt, 1990).

Regardless of the exact technique, co-occurrence data serves as measures of similarity between concepts. This measure of similarity would have a range of zero to one (zero denoting complete dissimilarity and one denoting complete similarity). The higher the conditional probability, the more similar the concept pair. These measures of similarity can be converted to distance data by subtracting the probability from one (see Figure 5).

McDonald and Schvaneveldt (1988) and McDonald et al. (1990) used conditional probability card sorting methodologies for UNIX functions. Hollands and Merikle (1987) used

conditional probabilities as a measure of concept similarity to model cognitive representations of database menu interfaces. Hirtle and Mascolo (1986) presented a card sorting task of city landmarks to forty undergraduates. Each person sorted cards three times. Therefore, Hirtle and Mascolo obtained conditional probability information both across and within individuals by effectively combining the two techniques mentioned above. Chi et al. (1981) used a sorting methodology in the categorization of physics problems across skill levels. Shoenfeld and Herrmann (1982) use a card sort to cluster concepts in the domain of mathematical problem solving ability.

2.3.1.4 Sequential Proximity Measures

There are recall techniques that can produce proximity data. One of these techniques is very similar to the card sorting methodology. In this procedure, the person recalls the concept list several times and the analyst computes the conditional probability that two items are adjacent. This procedure suffers from the fact that many of the concepts will never be adjacent (Cooke and McDonald, 1987).

Another way that proximity information can be obtained is by the measurement of inter-item distance between concepts in the sequential recall list. Thus, if a pair of concepts have five items between them in a particular recall list, the psychological distance between the items is defined to be five. Friendly (1977) also discusses this method. Repeated measures can be used, along with the use of rotating "seeds" or cue concepts (Reitman and Rueter, 1980; Gammack, 1990), to obtain measures of the consistency of the inter-item distance.

List generation is nearly identical to the recall-type techniques. As such, the distinction between the two may be insignificant. In list generation, the person is asked only to produce a list of domain concepts in a list type format and the sequential information of that list is analyzed. The major difference between list generation and the other techniques is that the list of concepts are produced at the same time as the relationships. As a result of this, list generation is not as systematic and efficient as the other techniques in which concepts are elicited prior to list analysis. Murphy and Wright (1984) have used a list generation technique to examine clinical psychological categories.

Event Record Analysis is similar to list generation techniques in the sense that the relationships between concepts are obtained by processing sequential records of concepts. However, instead of the list being obtained from a standard listing of concepts, the list is derived from a protocol or observation of a domain related task (Cooke and McDonald, 1987). The methods of obtaining similarity information from this data are identical to that of the inter-item distance measures for recall measures.

2.3.1.5 Twenty Question Technique

The twenty question technique is based on the idea that if the individual tries to guess the identity of a hidden target item, they will tend to ask questions that will narrow the possible alternatives. Gammack (1990) used this technique to classify locomotives.

In this technique, each concept is used as a target item that the individual must guess. The person is provided with the complete list of concepts. For each target item, the individual asks the experimenter questions regarding the characteristics of the target concept. For each response to the individual's probing question, the individual indicates which concepts were eliminated or retained by the previous question. Each question is also recorded. The number of the question at which a particular concept is discriminated from the target concept is defined as the similarity between the two items. The questioning process is repeated with each of the concepts as the target item. These data can be converted to distance information by subtracting the resulting ratings from twenty (the maximum number of questions).

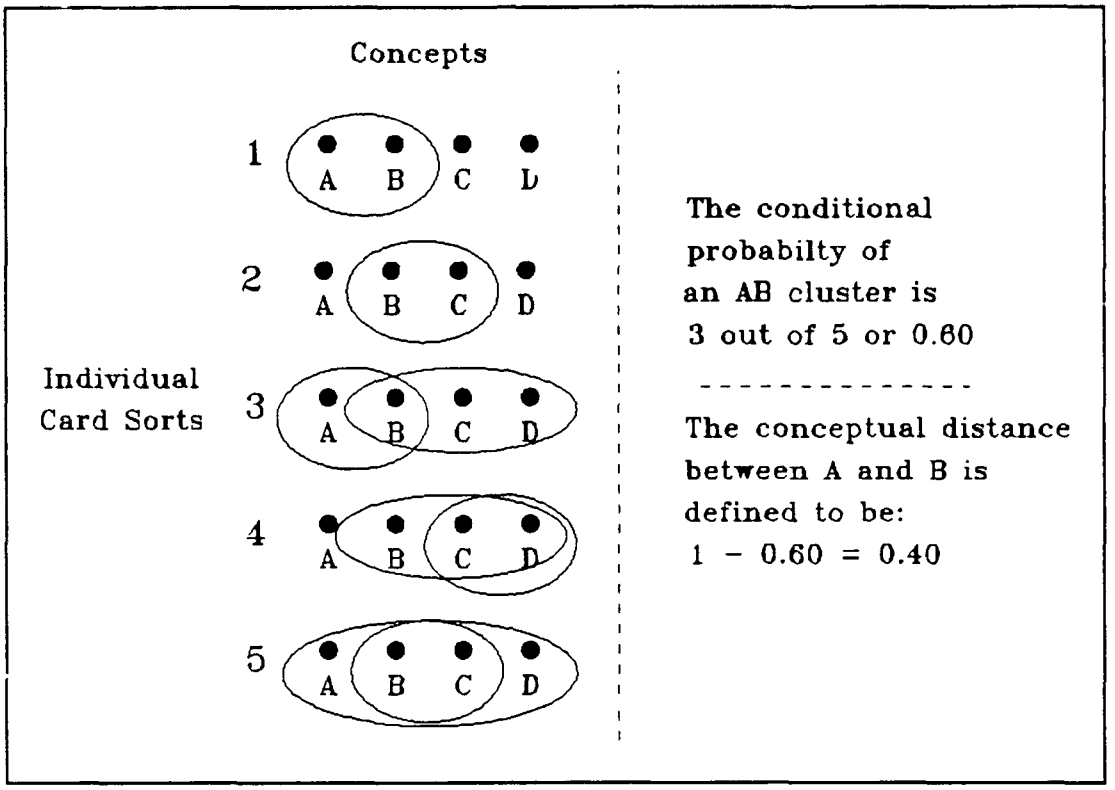

Figure 5. Conversion of co-occurrence data into conceptual distance.

2.3.2 Representation Development

Inputs to these representation generation techniques generally are similarity matrices as described in the previous section. The outputs of these techniques are spatial, symbolic representations of the knowledge structures.

2.3.2.1 General Weighted Networks (GWN)

General weighted networks are structures based on graph theory (Harary, 1969). Concepts are represented as points called "nodes" in the pictorial representation. The relationship between concepts are represented as "links" which are lines that connect certain pairs of nodes. Each link has a weight which is equivalent to the distance measure obtained for the two concepts that the link connects. A series of links (in which all nodes are distinct) is called a path. Figure 6 depicts a GWN representation of word processing commands.

General weighted networks reduce a set of inter-concept distance measurements by only including the most salient connections in the network. The basic idea behind the construction of a GWN is as follows: A link between any two nodes is included in the network if and only if the weight of that length is at least as small as the combined weight of any other possible path that connects the two nodes (Dearholt and Schvaneveldt, 1990).

One distinctive aspect of GWN's is that they can accommodate asymmetric distance data. Asymmetric data are produced when the similarity of each pair of concepts is judged to be different depending on the order of presentation of the concepts. Asymmetric data are represented as directional arrows on the links in the network.

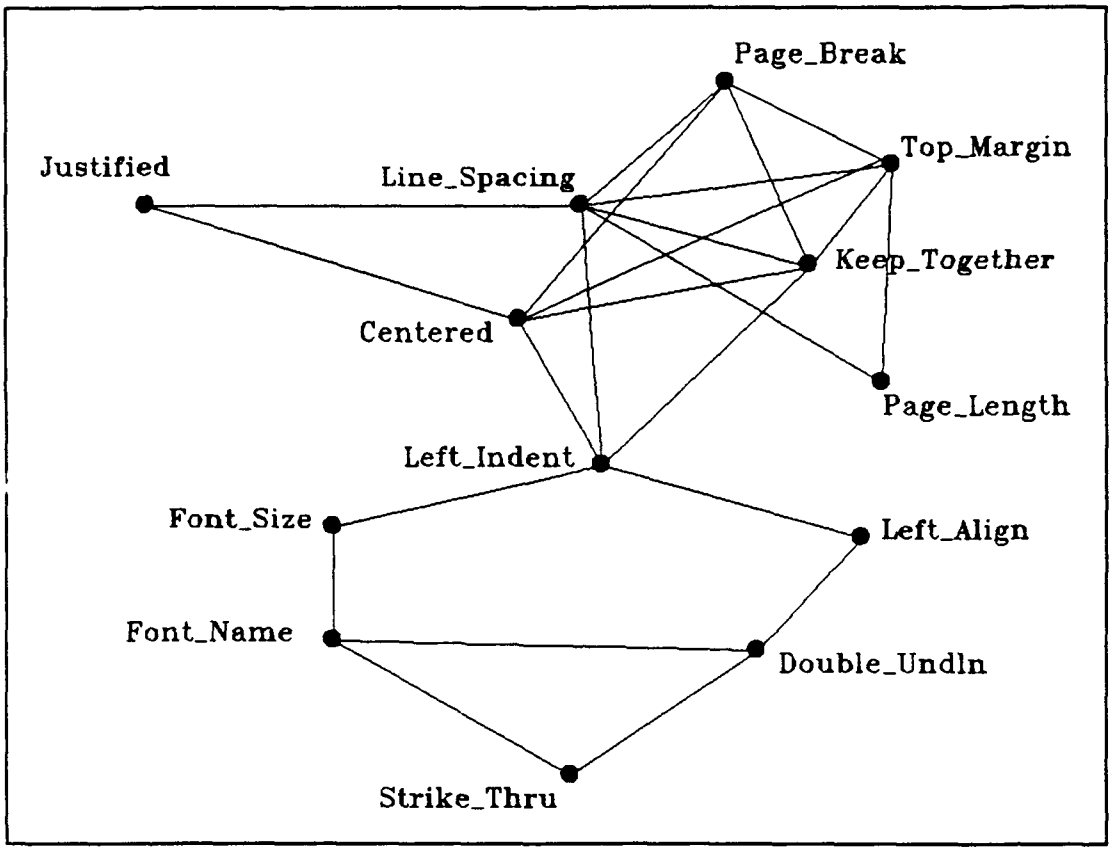

Figure 6. Example general weighted network of word processing commands.

There are several different applications in the literature where a GWN has been used as a technique to represent knowledge structures. Schvaneveldt, Durso, and Dearholt (1987) developed an algorithm called Pathfinder which produces a class of networks known as PFnets. Two parameters control the complexity and characteristics of the network. The r parameter determines how the paths between any two nodes are calculated. The selection of r also determines the measurement assumptions in the data (e.g. when r is set to infinity, ordinal data is assumed) The q parameter is the maximum number of links between two concepts that will be considered for inclusion in the network. PFnets have been widely used as a scaling methodology for knowledge structure measurement (Schvaneveldt et al., 1985; Cooke, Durso, and Schvaneveldt, 1986; Cooke and Schvaneveldt, 1988).

Hopkins et al. (1987) use another type of GWN called a directed graph (Miyamoto and Nakayama, 1984). All GWN's function basically the same way. They only differ in the way that path lengths are computed and in the provisions for link inclusion.

General weighted networks are one of the most flexible types of representations. In some ways, these representations are comparable to the overlapping or complex clustering representations discussed in Section 3.2. While their equivalency has not been examined mathematically, there are some important commonalities between these two representations.

2.3.2.2 Multidimensional Scaling (MDS)

Multidimensional scaling is a mathematical procedure with its roots found as early as 1928 (Schiffman et al., 1981). MDS represents the similarity of concepts spatially on an n-

dimensional map. In particular, the more similar two concepts are, the closer together they appear on the map. According to Schiffman et al. (1981), MDS models are algebraic equations which summarize certain information in the data. These algebraic models have geometric complements which yield the characteristic resultant spatial map of concepts.

MDS has been used in a variety of applications over its lengthy history, and in addition, has been implemented into many software packages (e.g. ALSCAL, MINISSA, POLYCON, INDSCAL/SINDSCAL, KYST, and MULTISCALE). Polzella and Reid (1989) employed MDS techniques to discover differences in performance characteristics of expert and novice combat pilots. Polzella and Reid argue that MDS allows deeper insight than simpler metrics "...into the underlying representation associated with performance of a complex task" (p.141). Similarly, MDS has been used to evaluate cognitive representations of fighter pilots (Schvaneveldt et al., 1982; Schvaneveldt et al., 1985). Other areas of application have included memory organization (Cooke, Durso and Schvaneveldt, 1986), the analysis of spatial aptitude (Pellegrino and Kail, 1982), the representation of perceived relations among properties of the human heart/blood vessel system (Hopkins et al., 1987), the changes in representation of text-editing commands with experience (Kay and Black, 1984) and the determination of the interaction between learning processes and the cognitive representation of problem solving (Enkawa and Salvendy, 1989).

Once the similarity data is input to a MDS procedure, the analyst must decide on the number of dimensions to be represented (it is obvious that the interpretation of results becomes more difficult as the number of dimensions increases). The program then assigns a set of coordinates to each of the concepts. This is known as the starting configuration. Distances are then calculated between the concepts and are compared to the actual data. If the distances are not nearly correct, the concepts are moved to new coordinates, and the distances are computed once again. The goodness of fit is given by a "stress" value, i.e., the lower the stress, the better the fit of the data. A number of iterations of the adjustment process may be performed until the analyst is comfortable with the fit of the data. It is generally noted that stress values decrease and correlations increase as the number of dimensions increase (Schiffman et al., 1981).

The analyst, when satisfied with the representation of the data, must then begin the task of deciding the placement and names of the axes. That is, on which dimensions are the data separated, and where does that separation occur? In some cases there may be clear distinctions between qualitative aspects of the data, however, it is possible that no distinction between dimensions can be made. Herein lies a difficulty with MDS representations. However, Schiffman et al. (1981) maintain that the data is represented in a sensible manner -- the distinctions are simply not clear.

2.3.2.3 Discussion of Representation Development Techniques

Several studies have been completed which quantitatively compare the performance of various representation development techniques. Schvaneveldt et. al. (1985) compared the ability of MDS, GWN, and raw proximity ratings to predict membership of an individual into groups of instructor pilots, guard pilots and novice pilots. They found that both MDS and GWN outperformed the raw ratings in predictive success. Further, they found that the MDS representation marginally outperformed the GWN. Goldsmith and Johnson (1990) obtained somewhat different findings in the domain of classroom learning. They found that GWN's generally outperformed both MDS representations and raw ratings in the prediction of students' final grades based on the match between the students' and the instructor's representations.

As a whole, the representation generation methodologies provide a way for the researcher to obtain a quick, pictorial summary of the information in the data.

2.3.3 Analysis of Representations

One of the criticisms with the use of the scaling methodologies is that once the pictorial representation is obtained, it is often unclear how this information can be applied to solve problems. Certainly, the representation may be subjectively evaluated by the analyst or domain experts. To some degree, this path of analysis results in the same difficulties as the verbal report methodologies discussed earlier. Granted, these representations certainly can take the role of supporting the traditional knowledge elicitation process (Olson and Rueter, 1987). However, for the variety of applications mentioned (e.g. selection, training) it is important to find ways to quantitatively evaluate and compare representations across individuals.

In general, one is interested in comparing the representations of two or more individuals in some particular domain. There are several forms that this comparison may take. First, in some instances, one may be interested in classifying individuals in groups based on their representations. Second, it may be appropriate to obtain a measure of structural similarity between a pair of representations. Third, one may wish to obtain information about common or distinguishing features between representations. The particular type of comparison selected depends to a large extent on the objective of the comparison. Validation of a particular cognitive measurement technique is a common goal. More frequently, one is interested in using these techniques for a particular application.

For example, if one is interested in selecting an employee for a particular position, the interest would be to determine the knowledge structure which is most associated with high levels of expertise in the domain and determine the candidate whose knowledge structure is most similar to the high expertise model knowledge structure. Alternatively, one may be interested in designing an interface in which the characteristics of the interface match (or adapt to) the knowledge structure of the potential user.

2.3.3.1 Analysis of General Weighted Networks

Graph Similarity Measures: There are a variety of methods available that attempt to assess the similarity or relatedness of two networks. These techniques employ different assumptions about what makes two representations structurally similar. Goldsmith and Davenport (1990) emphasize that "structural relatedness" is a subjective, perceptually based measure. Thus, there may be several ways in which structural similarity can be derived (structural similarity should not be confused with concept similarity, such as in pairwise similarity ratings). Below is a brief review of several approaches to structural similarity. See Goldsmith and Davenport (1990) for a formal mathematical treatment of these techniques.

Subentities: In this approach, a knowledge structure is viewed as an "entity" or as a group of subentities (e.g. cycles, cliques). Using this perspective, Graham (1987) reports a technique by Ulam in which structural similarity is defined as the inverse of the number of subentities required to construct both knowledge structures. The number of subgraphs range from one, for identical graphs, to a number equivalent to the number of edges (for two graphs with an equal number of edges).

Path Lengths: In this measure, the distance between concept pairs is used as a reference for comparing two knowledge structures. For two concepts that are not directly connected, the distance between them is defined to be the shortest path through the network. A correlation coefficient is computed by comparing the distances between each concept pair across networks. To do this, both networks are displayed in matrix form, where each cell represents the distance between two nodes. Next, the correlation coefficient is computed for the two networks on a cell by cell basis, which gives an indication of the similarity of the two networks. There are several variations of the path length method described by Goldsmith and Davenport (1990).

Cooke and Schvaneveldt (1988) used the path length method to verify that the network representations of groups of computer programmers were more similar for groups of individuals that had similar levels of expertise. Goldsmith and Johnson (1990) used the path length method to predict a student's performance by comparing the students' network representations with that of the instructor of the course. They found that the path length method did not predict course grades as well as the neighborhood method described below. Goldsmith and Davenport (1990) commented that the path length measure will be more effective in situations where distance information is more critical (such as a classification task), but less effective in situations where specific relationships between concepts is more important (e.g. the description of a corporate structure).

Neighborhoods: Another way to analyze network representations is to examine the "neighborhood" around each concept. A neighborhood can be defined as the set of nodes that are within one link of the node of interest (Goldsmith and Davenport, 1990). For each node in the networks being compared, the ratio of the number of links common to the node in both networks to the number of links connected to the node in either network is computed. The number which results from this measure ranges from 0 to 1, where 0 is least similar and 1 is most similar. Several variations of the neighborhood measure can be obtained by changing the way that the number of like connections are normalized.

Dayton et al. (1990) used a neighborhood type measure in the analysis of insight problems. They were interested in determining key or focal nodes in a network. They developed several measures of the focal node based on the degree of the node (number of incident links) and the node which is closest to all other nodes. A related issue involves the search for hierarchical type relationships in networks. In a network representation, a hierarchical relationship tends to have a star-like appearance. For instance, the superordinate concept will be in the middle and all the subordinate concepts will be linked to the superordinate. One measure of this property in a network representation is called the Starness Index (Durso and Coggins, 1990).

Feature Analysis: Although correlation coefficients may be computed to determine the similarity of two networks, this does not provide any distinct information as to *how* they are similar. Feature analysis is one method of providing such information.

In order to determine differences in the understanding of tactical flight maneuvers between undergraduate pilots, Air National Guard pilots and instructor pilots (Schvaneveldt et al., 1985), the network solutions of the instructors were defined as the standard to which all less experienced pilots' networks would be compared. The understanding of each concept was analyzed by determining the number of extra or missing associations to other concepts within the less experienced pilots' solutions. In doing this, four categories of concepts became apparent: well defined, under-defined, over-defined and misdefined. The proportion of critical links and the proportion of extra links in the undergraduate networks were computed, and a median split along each of the variables provided the basis for the classification of the concepts into one of the four previously mentioned categories. A similar methodology was followed by Cooke and Schvaneveldt (1988) in their exploration of the concept representations of naive, novice, intermediate and expert computer programmers.

Schvaneveldt et al. (1985) claim that this analysis would benefit the construction of a training curriculum. This claim would seem valid since the subsequent breakdown of concepts from a feature analysis provides information as to which concepts of a particular domain may need additional instruction, and which are well understood.

2.3.3.2 Analysis of Multidimensional Scaling Solutions

Structural Similarity Measures: The similarity of two MDS solutions can be obtained by calculating the Pearson correlation coefficient of the corresponding euclidean distances between

each concept pair for the two solutions. Although this method does not result in knowledge of *how* the individuals are similar, it at least provides enough information to determine which individuals' representations are most alike. Goldsmith and Johnson (1990) used this technique to compare MDS solutions of students with the instructor's representation.

INDSCAL: INDSCAL (individual differences MDS) is a computer program, based on a model of the same name, developed by Carroll and Chang (1970). Several modified versions of the INDSCAL program have since been developed. According to Schvaneveldt et al. (1985), INDSCAL may be used in order to see where certain groups, or individuals, are located with respect to the dimensions derived from the original MDS solution. In explanation, the relevance of a particular dimension to an individual or group can be determined from the distance of that individual (or group) to the zero coordinate for that dimension; therefore, the closer an individual lies to the zero coordinate of a dimension, the less important that dimension is to that individual.

Person Space: According to Schvaneveldt et al. (1985), a personal space representation is a means to represent individuals in dimensions relevant to themselves as opposed to relevant to the concepts. This method may be useful in the selection of an individual based upon his/her proximity to other individuals within the person space solution. For instance, if the dimensions of a person space were expert/novice, then an individual located close to experts within the solution would also be classified as an expert, whereas an individual located close to novices in the solution would be classified as a novice.

The person space is created by first, deriving an inter-subject distance matrix. The values for this matrix originate from distances in the MDS solutions of each individual. The differences between all distances for each possible pair of individuals are computed, and are then input to a MDS procedure. The number of dimensions to be represented is again left to the discrepancy of the analyst. A one dimensional solution will account for most of the variance in the data. In addition, a single dimension representation will most likely yield an ordered representation along a given dimension (e.g. novices to the left and experts to the right).

Linear Discriminant Functions: Schvaneveldt et al. (1985) also describe the use of linear discriminant functions in order to classify individuals into two or more groups based upon their cognitive representations. Basically, this method utilizes vectors in order to divide the solution space into regions which are comprised only of particular types of patterns (e.g. expert/novice).

This technique has been used by Schvaneveldt et al. (1982,1985) to classify experienced and unexperienced pilots based upon the representations of their understanding of flight concepts. Schvaneveldt et al. (1982) also analyzed PFnets in this manner.

The first step in the creation of the discriminant functions is to determine the minimum distance classification. This is done by first representing each class of individuals by a single point (the prototype point) whose location is derived from the central tendencies of n-1 values in each group. A line is then drawn to connect these points. In the case of only two prototype points, the linear discriminant function is the equation of the perpendicular bisector of the line between the prototype points. This function is also called the decision surface, for this is the line which separates one classification from the other. If all n-1 individuals are correctly classified by this line, the remaining individuals are then input to the procedure for classification. As indicated by Schvaneveldt et al. (1982), this method works well if the individuals are clustered tightly around the prototype points.

A training algorithm may be employed if the first linear discriminant function fails to correctly classify the n-1 individuals. This is done by altering the weight vector by multiplying it by a constant until all combinations of n-1 individuals are correctly classified by the decision surface. It is indeed possible that this process may continue for several iterations until the correct decision surface is found.

2.4 Summary

In summary, it appears a wide variety of KR measurement techniques have been used in the literature, ranging from verbal reports to discriminant functions

It is clear that each KR measurement methodology implicitly makes assumptions regarding the contents, or components, of the human knowledge structure. Rarely, however, are these assumptions made explicit. In fact, few publications go so far as to provide an operational definition of knowledge structure: the object of measurement. According to Goldsmith and Johnson (1990), "Palmer (1978) noted that the field is 'obtuse, poorly defined, and embarrassingly disorganized'(p.259). Although more than a decade has passed since Palmer made these observations, there is ample evidence to suggest that his observations remain valid (p.243)." A potential explanation of this disarray may rest in the fact that specific techniques are designed to measure particular components of the KR, and conflicting outcomes are a result of measuring different elements of the overall construct labeled knowledge structure.

Present understanding of KR has been driven primarily by the individual measurement techniques available and associated mathematical assumptions. While this work has provided a rich empirical base, a comprehensive model is now necessary for further systematic advancement in understanding human knowledge structure. Such a model can yield parsimony, serve as a foundation and impetus for model-based KR measurement techniques and identify KR parameters for predicting operator performance in cognitive-oriented tasks.

2.5 Model of Knowledge Structure

To begin, a working definition of knowledge structure was defined as follows: "Human Knowledge structure is defined as the structure of interrelationships between concepts and procedures (elements) in a particular domain, organized into a unified body of knowledge". In addition, a preliminary effort at identifying parameters, or features, which define the KR model is undertaken. The above definition suggests two main components of knowledge structure: elements and their interrelationships. These elements can be further described through two additional dimensions, or parameters. First, the knowledge structure elements may either be declarative or procedural concepts, and second, these elements can exist at various levels of abstraction. A cursory examination of concept lists used in the previously described research indicates both declarative and procedural concepts are used for relationship elicitation. Also, existing work, (Adelson, 1984; Koubek and Salvendy, 1989; etc.) has shown elements at varying degrees of abstraction. For example, critical elements in the domain of computer programming may be "FOR-NEXT" loop and "CONTROL STRUCTURE". The first element is procedural in nature, while CONTROL STRUCTURE is considered more declarative. Also, FOR-NEXT LOOP is a component of CONTROL STRUCTURE and therefore is at a lower level of abstraction.

In describing the second major component, relationships, two additional parameters can also be identified. First, the degree of relationship between elements can vary in degree, and second, more than one type of relationship can simultaneously exist between two or more elements. Obviously, the primary input to many knowledge structure measurement techniques is the proximity matrix, for which varying degrees of relatedness is the basis. Also, techniques such as overlapping closed curves, multidimensional scaling and repertory grid have indicated the existence of multiple relationships between elements. For example, three potential elements in a process control task, FLOW RATE, TEMPERATURE, and VOLUME, may be related to one another in varying degrees and in more than one way. For example, FLOW RATE and VOLUME may be related based on the physical proximity of their displays and/or their interaction with PRESSURE. The proposed model is outlined in table 1.

2.6 Derivation of Hypotheses

The purpose of this study is three-fold. The first issue is to show that the KR dimensions in the model proposed hereto indeed exist. Next, it is determined whether an individual's experience level in a given cognitive-oriented domain affects the characteristics of the proposed model dimensions, and finally, differences are identified in existing KR measurement techniques' capabilities (reviewed above) to reveal these differences in knowledge structure. Therefore, the following primary hypotheses are proposed:

Hypothesis One: Each proposed dimension of knowledge structure exists.

Hypothesis Two: The characteristics of the proposed KR model dimensions are affected by domain experience.

Hypothesis Three: Differences exist between the measurement capabilities of current KR measurement methodologies across the proposed model dimensions.

The purpose of the first hypothesis is to provide support for the proposed KR model, and in effect, give cause to conduct the remainder of the study. The second hypothesis attempts to reveal which of the proposed model dimensions are a function of experience level in a cognitive domain. Finally, the third hypothesis attempts to expose the effectiveness of available KR measurement techniques, which will point towards areas of those techniques that are in need of future improvement.

Table 1. Proposed model of knowledge structure.

COMPONENT	FEATURE
Element	Unique concepts, or knowledge units, of a domain exist as elements in the KR.
	Elements can exist as declarative or procedural concepts.
	Multiple levels of abstraction among elements can exist in a single KR.
Interrelationship	Relationships exist between the elements.
	The relationship between elements can vary in degree.
	More than one type of relationship can simultaneously exist between two or more elements.

3.0 METHOD

This experiment required experienced and naive individuals in a given domain to perform a series of tasks associated with selected KR measurement techniques. Representational differences between the two subject groups were quantitatively assessed based on the output from the knowledge structure measurement techniques.

3.1 Task

In order to test the above hypotheses, the domain used must be cognitive-oriented, and must be broad enough to encompass each of the proposed KR model dimensions. The domain chosen which meets these criteria is that of clerical work. The particular contents of this domain are outlined in the Dictionary of Occupational Titles (DOT 201.362-030).

The particular tasks performed were those necessary to carry out the procedures dictated by the individual KR measurement techniques. Because, at this point in time, our understanding of knowledge structure is dependent upon the measurement techniques currently available, a battery of techniques was used in an attempt to gain a more complete perception of the presence of the proposed model dimensions. The KR measurement techniques selected to be tested are representative of the previously discussed prominent categories. In particular, card sorting, hierarchical clustering analysis (Proc Cluster, SAS Institute, 1988), multidimensional scaling (ALSCAL) and Pathfinder were incorporated. Furthermore, an analysis of the repertory grid technique and pairwise similarity ratings was performed. Verbal report techniques, such as GOMS analysis and problem behavior graphs, were not selected for inclusion in the study because of the subjectivity involved in the derivation of the knowledge structures and their lack of well defined analysis methods.

3.2 Stimulus

The experimental stimulus used in each of the following tasks consisted of 30 clerical domain-relevant concepts which were elicited in pilot study interviews. Six subjects participated in the pilot study, of which three individuals were experienced secretaries and three were naive. Each subject was asked to list as many concepts as possible which they considered to be important to the general secretarial job. These concepts, they were told, may be anything from equipment used to procedures followed throughout the day.

Upon completion of the interviews, one master list of concepts was generated for each subject group. To insure that the test stimulus would consist of concepts representative of both subject groups, the intersection of the master lists was recorded. This process yielded 24 concepts. Since the purpose of the study was to draw out differences between experience levels, an additional six concepts were extracted from the master list of the experienced secretaries. These six were concepts that at least two of the three secretaries held in common. Therefore, the stimulus list was comprised of 30 domain-relevant concepts, 24 in common to both subject groups, and six in common only to the experienced group.

3.3 Subjects

Thirty subjects participated in the experiment. Fifteen were secretaries with at least five years of experience, and the remaining fifteen were considered naive in the field (having a maximum of one year of secretarial experience). The mean experience of the secretaries was 15.20 years with a standard deviation of 7.92 years, while the mean experience of the naive group was 0.20 years with a standard deviation of 0.41 years. Each of the subjects were informed as to the purpose of the experiment and were paid for their participation.

3.4 Experimental Design

The experimental design is a straightforward comparison of experienced and naive group means on each of the dependent variables tested throughout the experiment. However, because of the nature of two of the dependent variables, tests between proportions were performed.

The independent variable in this study is experience level: Experienced and Naive. Because of the nature of the specific KR measurement methodologies used, the dependent variables differ from technique to technique across the model dimensions. Each are quantitative in order to reduce the subjectivity in the analysis, and have been derived from the assumptions of the various techniques or previous literature. The dependent variables are presented in table 2, and are described here for each technique.

3.5 Dependent Variables

3.5.1 Declarative and Procedural Concepts

This particular dimension is independent of the measurement techniques, and is determined by the percentage of procedural concepts listed during a concept elicitation process.

3.5.2 Multiple Levels of Abstraction

Card Sorting: The average number of cards in a pile. A large number of concepts in a pile represents a higher order of abstraction in one's knowledge structure. This is based on the assumption that more concepts present in a pile require a more abstract classification label.

HCS: The average number of concepts in a cluster at the median joining distance. A large number of concepts in a cluster represents a higher order of abstraction in one's knowledge structure. This is based on the same assumption as above.

Pathfinder: The number of stars present in a representation (where a star is defined as a concept with at least five incident links). A larger number of stars represents a higher order of abstraction in one's knowledge structure. This is based on the same assumption as above.

3.5.3 Multiple Relations

Card Sorting: The number of repeated cards. A larger number of repeated cards provides evidence for more types of simultaneous relations existing between concepts.

Repertory Grid: The number of dimensions elicited. A greater number of dimensions elicited provides a basis for more types of simultaneous relations between concepts.

MDS: The number of distinct dimensions labeled by the subject. Subjects are shown a plot of one MDS dimension at a time, and are asked to label the dimensions on which the concepts are arranged. The more distinct number of dimensions labeled, the more types of simultaneous relations exist between concepts.

3.5.4 Varying Degree of Relatedness

Card Sorting: The co-occurrence of a randomly selected concept pair in a pile. Two concepts occurring together in the same pile are more strongly related than two in separate piles.

HCS: The co-occurrence of a randomly selected concept pair in a cluster. Two concepts occurring together in the same cluster are more strongly related than two in separate clusters.

Repertory Grid: The distance between a randomly selected concept pair. All dimensions are collapsed, and the distance is taken between the average rating for each selected concept. The *more related* two concepts are, the smaller the distance between them.

MDS: The distance between a randomly selected concept pair. The euclidean distance is calculated between the concepts in the two-dimensional solution. The closer the two concepts are in space, the more related they are.

Pathfinder: The number of links between a randomly selected concept pair. The number of links is determined by the shortest possible path between the two concepts. A smaller number of links between concepts denotes a higher relation between those concepts

Pairwise Similarity Ratings: The similarity of a randomly selected concept pair. Therefore, a higher similarity rating denotes more similarity between given concepts.

3.6 Procedure

Card Sorting: Each subject was presented with a pile of 30 index cards. Each card contained one of the domain concepts elicited during the pilot study interviews. The subjects were asked to sort the cards into piles based upon which concepts they felt "go together". They were told that, when they finished, there may be as few as one pile (if all concepts were seen as going together) or as many as 30 piles (if all concepts were seen as being separate). Additional cards could be requested if the subject determined that a particular concept belonged in more than one pile. It was stressed that there were no right or wrong answers and there was no time constraint. When the sorting was complete, the subjects were asked to provide a label for each pile (that is, to provide a reasoning for their placement of the concepts). The total number of piles and their respective labels were recorded.

Repertory Grid: Following the card sorting procedure, subjects were presented with three randomly selected concept cards. They were asked to determine a dimension that would separate two of the concepts from the other one. Random triads were presented until the experimenter determined that a representative list of dimensions had been elicited. To construct the rating grid, the domain concepts were listed down the left-hand side of the grid while the elicited dimensions were listed across the top. Because of the dependency of the dimensions on an individual subject, each grid was tailored specifically for that subject. Subjects were presented with the grid the following day, and were allowed to complete it at home. They were asked to rate each concept on an ordinal scale from one to seven on each dimension, and to return the rating sheet as soon as they had finished.

Concept Elicitation: On a second day, subjects were asked to list as many concepts as possible which they felt to be important in the secretarial field of work. They were told that these concepts could be anything from equipment used to procedures followed throughout the day (identical to the pilot study interviews mentioned earlier). The percentage of procedural and declarative concepts were determined for each subject (i.e., "files" is considered to be a declarative concept, whereas "filing" is considered procedural).

Table 2. Dependent variables for Hypothesis Two.

KR MODEL DIMENSION	TECHNIQUE					
	Card Sorting	HCS	Repertory Grid	MDS	Pathfinder	Similarity Ratings
Declarative & Procedural Concepts	Percentage of Procedural Concepts Elicited					
Multiple Levels of Abstraction	Average Number of Cards Per Pile	Average Number of Concepts Per Cluster	NA	NA	Number of Stars	NA
Multiple Relations	Number of Repeated Concepts	NA	Number of Dimensions Elicited	Number of Dimensions Labeled	NA	NA
Varying Degree of Relatedness	Co-Occurrence of Pairs in Piles	Co-Occurrence of Pairs in Clusters	Distance Between Concept Pairs	Distance Between Concept Pairs (2-D)	Number of Links Between Concept Pairs	Similarity of Concept Pairs

NA = No quantitative measure available

Pairwise Similarity Ratings: A computer program was developed which presented all possible pairs of the domain concepts to the subjects (resulting in 435 concept pairs). The subjects were shown one pair at a time, and were asked to rate their perceived similarity of the two concepts on a Likert-type scale from one to seven (one meaning extremely dissimilar and seven meaning extremely similar). The subjects were asked to use the entire range of values as necessary. The subjects were also told that a number of factors may influence the similarity of the concepts, and that all factors should be considered when a rating was assigned. This data was later submitted to Proc Cluster, ALSCAL and Pathfinder for further analysis. The default settings were used for each of these techniques.

4.0 RESULTS

4.1 Hypothesis One

As stated earlier, the purpose of this hypothesis is to show that the proposed dimensions within the model of knowledge structure do indeed exist. To do this, it was necessary to determine if the values collected within each dimension for at least one measurement technique were either different from zero (as in the case of Static and Procedural Concepts and Multiple Relations) or different between concepts. The specific processes involved for each dimension are discussed below.

Static and Procedural Concepts

To test for the existence of this dimension, a t-test was performed on the average percentage of procedural concepts elicited from the Experienced group in order to determine that this percentage was greater than zero. The results of this test were significant, $t(14) = 12.07$; $p < 0.005$.

Multiple Levels of Abstraction

The starness values of two randomly selected concepts (Communication and Liaison) were compared to determine if differences in levels of abstraction exist within the Experienced group. The t-test was significant, $t(28) = 3.90$; $p < 0.01$, indicating a difference in the level of abstraction between the two concepts.

Multiple Relations

The number of repeated cards in the card sorting procedure (Experienced group) was tested to determine if this number was greater than zero. The t-test was significant, $t(14) = 2.43$; $p < 0.025$, providing evidence for the simultaneous existence of multiple relations between concepts.

Varying Degree of Relatedness

To determine the existence of this dimension, three concepts were selected by an independent observer: Accounting, Payroll and Shorthand. A t-test was then performed on the similarity ratings of the two concept pairs Accounting-Shorthand and Accounting-Payroll for the Experienced group. The test was significant, $t(28) = 2.82$; $p < 0.01$, indicating that varying degrees of relatedness exist between concepts within the knowledge structure.

4.2 Hypothesis Two

Static and Procedural Concepts

This variable was not significantly different between experience level groups, $t(23) = 0.90$; $p < 0.38$ (see table 3).

Multiple Levels of Abstraction

Card Sorting: The test between experience group means did not yield significant results for this variable, $t(28) = 0.57$; $p < 0.57$.

HCS: The results of this test were statistically significant, $t(28) = 2.53$; $p < 0.02$. The clusters of the Experienced group contained more concepts than the clusters of the Naive group.

Pathfinder: While not statistically significant, $t(28) = 1.91$; $p < 0.07$, there appears to be a trend for the Experienced group's representations to contain more stars than the Naive group's representations.

Multiple Relations

Card Sorting: This test did not yield statistically significant results, $t(28) = 0.51$; $p < 0.62$.

Repertory Grid: The number of dimensions elicited were not significantly different for the two experience level groups, $t(28) = 0.74$; $p < 0.46$.

MDS: This test did not yield statistically significant results, $t(27) = 0.47$; $p < 0.64$.

Table 3. Summary of results for Hypotheses Two and Three.

KR MODEL DIMENSION	TECHNIQUE						SUMMARY
	Card Sorting	HCS	Repertory Grid	MDS	Pathfinder	Similarity Ratings	Significant Differences
Declarative & Procedural Concepts	$X_E = 0.82$ $SD_E = 0.13$ $X_N = 0.74$ $SD_N = 0.24$ $p < 0.39$						NO
Multiple Levels of Abstraction	$X_E = 6.28$ $SD_E = 2.54$ $X_N = 5.77$ $SD_N = 2.34$ $p < 0.57$	$X_E = 3.19$ $SD_E = 0.64$ $X_N = 2.71$ $SD_N = 0.39$ **$p < 0.02$**	NA	NA	$X_E = 13.60$ $SD_E = 5.64$ $X_N = 10.33$ $SD_N = 3.46$ $p < 0.07$	NA	YES
Multiple Relations	$X_E = 1.93$ $SD_E = 3.08$ $X_N = 1.47$ $SD_N = 1.77$ $p < 0.62$	NA	$X_E = 9.47$ $SD_E = 2.17$ $X_N = 8.87$ $SD_N = 2.26$ $p < 0.46$	$X_E = 3.2$ $SD_E = 0.94$ $X_N = 3.36$ $SD_N = 0.84$ $p < 0.64$	NA	NA	NO
Varying Degree of Relatedness	$p_E = 0.13$ $p_N = 0.07$ $p < 0.32$	$p_E = 0.33$ $p_N = 0.00$ **$p < 0.007$**	$X_E = 1.01$ $SD_E = 0.54$ $X_N = 0.67$ $SD_N = 0.44$ $p < 0.07$	$X_E = 2.19$ $SD_E = 0.51$ $X_N = 2.06$ $SD_N = 0.75$ $p < 0.58$	$X_E = 3.00$ $SD_E = 1.60$ $X_N = 3.27$ $SD_N = 1.39$ $p < 0.63$	$X_E = 1.73$ $SD_E = 1.28$ $X_N = 2.60$ $SD_N = 1.50$ $p < 0.10$	YES
Significant Differences	NO	YES	NO	NO	NO	NO	

NA = No quantitative measure available

Varying Degree of Relatedness

The measurement of this dependent variable required that a pair of concepts be randomly selected by an outside party. The restriction placed on this selection was that the pair must have come from the six domain concepts that were only in common to the experienced secretaries in the pilot study. This would allow an observation of the differences in the degree of relatedness between the two experience groups for a presumably higher level pair of concepts. The pair chosen was "Accounting" and "Supervising Employees".

Card Sorting: The test between group proportions did not yield statistically significant results, $z = 0.46$; $p < 0.32$.

HCS: Statistically significant results were found with this technique. A test between group proportions yielded $z = 2.46$; $p < 0.007$.

Repertory Grid: The result of this test was not statistically significant, $t(28) = 1.88$; $p < 0.08$.

MDS: This technique did not yield statistically significant results, $t(28) = 0.56$; $p < 0.58$.

Pathfinder: The result of this test was not statistically significant, $t(28) = 0.49$; $p < 0.63$.

Pairwise Similarity Ratings: The result of this test was not statistically significant, $t(28) = 1.7$; $p < 0.10$.

4.3 Hypothesis Three

Based on the results of the individual techniques used within Hypothesis Two, it is readily seen that differences do exist in the KR measurement capabilities between techniques on the model dimensions. In particular, HCS was the only technique that extracted significant differences ($P < .05$) in the knowledge structures of the two experience level groups.

5.0 DISCUSSION

5.1 Existence of Model Dimensions

The first hypothesis, that each proposed dimension of knowledge structure exists, was supported. This was shown by testing that the values obtained during experimentation for each of the dimensions were either different from zero (in the case of Procedural and Declarative Concepts and Multiple Relations) or different between concept pairs (in Multiple Levels of Abstraction and Varying Degree of Relatedness). Therefore, it has been shown that the proposed dimensions are a good *starting point* in the development of a comprehensive model of knowledge structure. Further, the support of this hypothesis provided the necessary foundation on which to perform the testing of Hypotheses Two and Three.

5.2 Affect of Experience Level on Model Dimensions

In regard to the second hypothesis, that the characteristics of the proposed knowledge structure model dimensions are affected by domain experience, three outcomes were possible for each of the proposed model dimensions: (1) None of the techniques administered found differences between experience levels, (2) a subset of the techniques administered found differences between experience levels, and (3) each of the techniques administered found differences between experience levels. As long as at least one of the techniques administered revealed significant differences between experience level groups for a given model dimension, there is evidence to suggest that the characteristic of the particular dimension is affected by differences in experience level. Therefore, it is suggested that each dimension in the model differs with experience level and may influence an individual's performance on a cognitive task. Such evidence was found for the dimensions of Multiple Levels of Abstraction and Varying Degree of Relatedness. However, the dimensions that did not yield significant differences

between subject groups should not be discarded as possible determinants of performance. Perhaps differences would have been found had the techniques used been more sensitive to those dimensions of knowledge structure. Furthermore, one cannot rule out the possibility that the domain chosen for this study was not one in which experience level necessarily produces differences on all knowledge structure dimensions.

In addition to the proposed model dimensions, a post-hoc analysis indicated that another dimension measured with Pathfinder, Representation Complexity (the total number of links present in the representation), is an indicator of representation differences between experience levels, $t(28) = 2.43$; $p < 0.03$. Therefore, any future model of knowledge structure should include a complexity parameter.

5.3 Differences in Measurement Techniques

The third hypothesis, that differences exist in knowledge structure measurement capability among techniques on the model dimensions, was supported. This was shown by the different results obtained by the various techniques across the model dimensions. The results of this study indicate that hierarchical clustering analysis was the only technique to extract significant differences between the experience groups on any of the model dimensions. It appears then, that the remaining techniques were not sensitive enough in this scenario to yield significant differences between experience groups. It is interesting to note that the one technique that extracted representation differences between skill levels used statistical methods to transform the data prior to producing the representation. This provides support for the latent trait techniques as discussed in the literature review (section 2.0), and suggests that, indeed, these transformation techniques are useful in eliciting the contents of knowledge structure. Although they did not yield significant results at the 0.05 alpha level, in comparison to the other techniques, Pathfinder (Multiple Levels of Abstraction) and the repertory grid technique (Varying Degree of Relatedness) appear to have revealed differences between experience levels and should, therefore, be considered as useful techniques for measuring these respective dimensions. Given these results, it appears that Pathfinder and hierarchical clustering analysis are useful techniques in revealing differences in the levels of abstraction between different experience groups, while the repertory grid technique provides an effective indication of the varying degree of relatedness between pairs of domain-related concepts. As mentioned earlier, none of the measurement techniques employed in this study were able to detect differences between experience levels for the dimension of Multiple Relations. Although differences were not found for this dimension, the author believes this may be due to a weakness of the available measurement techniques. More research is needed to determine the useability of this dimension to determine knowledge structure differences.

5.4 Recommendations

Three specific areas of future work are discussed below: Refinement of the model of knowledge structure, the development of new measurement methodologies, and the application of this research in the work place.

From the results of this study, it may be seen that there is a need to refine the proposed model of human knowledge structure. In particular, the inclusion of an additional dimension that represents the complexity of the knowledge structure is necessary. Furthermore, it is imperative that work continue in the development of knowledge structure measurement techniques. For, none of the techniques used here have the ability to measure each of the parameters of the proposed model of knowledge structure.

The focus of future work should be on the development of a single, comprehensive measurement methodology capable of eliciting reliable information from each of the parameters of knowledge structure. The results of this experiment necessitate that any new technique be able

to determine differences in the dimensions of Multiple Levels of Abstraction, Varying Degree of Relatedness, and Representation Complexity. In addition, the technique should elicit quantitative data to reduce subjectivity in the interpretation of the knowledge structure dimensions, and should be applicable for use in a variety of work place domains.

When these goals have been accomplished, it may be possible to predict an individual's performance on a given cognitive-oriented task based upon the characteristics of that individual's knowledge structure. Such information would be applicable in personnel selection, training and job design. For example, certain jobs may be better suited for individuals with a higher level of abstraction in their knowledge structure. Training could be geared toward the development of certain knowledge structure attributes needed for an operator to efficiently perform a given job. Finally, jobs themselves could be designed in such a way as to be "matched" with the knowledge structure of the operator, for there has not been much success with training individuals to adopt a particular knowledge structure (Adelson, 1984).

In summary, this experiment has validated the existence of four proposed dimensions of knowledge structure, identified two as being affected by differences in domain experience levels and identified an additional dimension to be included in a future model of knowledge structure. Furthermore, differences were found in the measurement capabilities of available knowledge structure measurement techniques. Therefore, the primary objectives of this effort were satisfied. These results support the importance of validating a formally defined model of knowledge structure, and the need for the development of new measurement methodologies in order to advance our understanding of the field.

6.0 CONCLUSION

The impetus for this research was the historical lack of an operational definition of knowledge structure, and, likewise, the non-existence of a formally defined model of knowledge structure. Such information is necessary in order to further the advancement of work in understanding knowledge structure by bringing parsimony to the field.

The model of human knowledge structure proposed by here appears to be a reasonable first attempt at identifying the actual parameters of the human knowledge structure. Furthermore, it was shown that the parameters Multiple Levels of Abstraction and Varying Degree of Relatedness were affected by representation differences between experience levels. Therefore, since experience is often equated with skill, it may be possible to predict cognitive-oriented task performance based upon the characteristics of these two model dimensions. This has implications for personnel selection, training and job design.

Finally, differences between the measurement capabilities of the various knowledge structure measurement techniques were revealed. None of the techniques utilized were able to determine representation differences for each proposed dimension of knowledge structure. This points to the need to develop new measurement methodologies capable of eliciting information regarding all knowledge structure parameters.

Refinements of the model of knowledge structure are needed. However, through this experiment, the first steps have been taken to provide an operational definition, and validate a formally defined model of knowledge structure in order to provide organization to a previously disorganized field.

REFERENCES

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & Cognition*, 9, 422-433.

Anderson, J. R. (1983). *The Architecture of Cognition*, London, England: Harvard University Press.

Barfield, W. (1986). Expert-novice differences for software: Implications for problem solving and knowledge acquisition. *Behaviour and Information Technology*, 5, 15-29.

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. (1986). *Expert Transfer for Expert Systems Design*, New York: Elsevier.

Bower, G. H., and Springsteen, F. (1970). Pauses as recording points in letter series. *Journal of Experimental Psychology*, 83, 421-430.

Card, S. K., Moran, T. P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Carroll, J. D., and Chang, J. (1970). Analysis of individual differences in multidimensional scaling via and n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.

Chase, W. G., and Simon H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.

Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.

Chi, M. Hutchinson, J., and Robin, A. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly*, 35, 27-63.

Colthart, V., and Evans J. St. B. T. (1981). An investigation of semantic memory in individuals. *Memory & Cognition*, 9, 524-532.

Cooke, N. M., Durso, F. T. and Schvaneveldt, R. W.(1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12, 538-549.

Cooke, N., and McDonald, J. (1986). A formal methodology for acquiring and representing expert knowledge. *Proceedings of the IEEE*, 74, 1423-1431.

Cooke, N., and McDonald, J. (1987). The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems. *International Journal of Man-Machine Studies*, 26, 533-550.

Cooke, N. J., and Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man-Machine Studies*, 29, 407-427.

Dayton, T. Durso, F. T., and Shepard, J. D. (1990). A measure of the knowledge reorganization underlying insight. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organizations*, New Jersey: Ablex Publishing Corporation. 267-277.

De Groot A. D. (1965). *Thought and choice in chess*, The Hague: Mouton.

Dearholt, D. W., and Schvaneveldt, R. W. (1990). Properties of Pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organization*, New Jersey: Ablex Publishing Corporation. 1-30.

Durso, F. T. and Coggins, K. A. (1990). Graphs in the social and psychological sciences: Empirical contributions of Pathfinder. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organizations*, New Jersey: Ablex Publishing Corporation. 31-51.

Egan, D. E., and Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, 7, 149-158.

Enkawa, T., and Salvendy, G. (1989). Underlying dimensions of human problem solving and learning: Implications for personnel selection, training, task design and expert systems. *International Journal of Man-Machine Studies*, 30, 235-254.

Ericsson, A. K., and Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.

Ericsson, K. A., and Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*, Cambridge Massachusetts: MIT Press.

Esposito, C. (1990). A graph-theoretic approach to concept clustering. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organizations*, New Jersey: Ablex Publishing Corporation. 89-100.

Evans, B. T. (1988). The knowledge elicitation problem: A psychological perspective. *Behaviour and Information Technology*, 7, 111-130.

Friendly, M. L. (1977). In search of the M-gram: The structure of organization in free recall. *Cognitive Psychology*, 9, 188-249.

Gammack, J. G. (1990). *Expert conceptual structure: The stability of Pathfinder representations*. In R. W. Schvaneveldt (Ed.), Pathfinder Associative Networks, New Jersey: Ablex Publishing Corporation. 213-226.

Gobbo, C., and Chi, M. (1986). How knowledge is structured and used be expert and novice children. *Cognitive Development*, 1, 221-237.

Goldsmith, T. E., and Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organizations*, New Jersey: Ablex Publishing Corporation. 75-88.

Goldsmith, T. E., and Johnson P. J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organizations*, New Jersey: Ablex Publishing Corporation. 241-254.

Graham, R. L. (1987). A similarity measure for graphs--reflections on a theme of Ulam. *Los Alamos Science*, 114-212.

Harary, F. (1969). *Graph Theory*, Reading, Massachusetts: Addison-Wesley.

Hardiman, P. T., Dufresne, R., and Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novice. *Memory & Cognition*, 17, 627-638.

Hirtle S. C., and Mascolo, M. F. (1986). Effect of semantic clustering on the memory of spatial locations. *Journal of Experimental Psychology*: Learning, Memory and Cognition, 12, 182-189.

Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., and Patel, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses: Expert-novice differences. *Medical Education*, 21, 471-476.

Hollands J. G., and Merikle, P. M. (1987). Menu organization and user expertise in information search tasks. *Human Factors*, 29, 577-586.

Hopkins, R. H., Campbell, K. B., and Peterson, N. S. (1987). Representations of perceived relations among the properties and variables of a complex system. *IEEE Transactions on Systems, Man, and Cybernetics*, smc-17, 52-60.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241-255.

Kay, D. S., and Black, J. B. (1984). Changes in knowledge representations of computer systems with experience. *Proceedings of the Human Factors Society*, 28, 963-967.

Keen, T. R. and Bell, R. C. (1981). One thing leads to another: a new approach to elicitation in the repertory grid technique. In Shaw, M. Ed., *Recent Advances in Personal Construct Technology*. New York: Academic Press.

Kelly, G. A. (1955). *The Psychology of Personal Constructs*, Norton: New York.

Koubek, R. J., Salvendy, G., Eberts, R. E., and Dunsmore, H. (1987). Eliciting knowledge for software development. *Behaviour and Information Technology*, 6, 427-440.

Koubek, R. J., and Mountjoy, D. N. (1991). The impact of knowledge representation on cognitive-oriented task performance. *International Journal of Human-Computer Interaction*, 3, 31-48.

Koubek, R. J., and Salvendy, G. (In press). *Cognitive performance of super-experts on computer program modification tasks. Ergonomics.*

Leinhardt, G. and Smith, D.A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, 77, 247-271.

Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis. *Reading Research Quarterly*, 22, 407-433.

McDonald, J. E., Paap, K. R., and McDonald, D. R. (1990). *Hypertext perspectives: Using Pathfinder to build Hypertext Systems.* In Schvaneveldt, R. W. (Ed.), Pathfinder Associative Networks: Studies in Knowledge , New Jersey: Ablex Publishing
Corporation. 197-212.

McDonald, J. E., Plate, T. A., and Schvaneveldt, R. W. (1990). Using Pathfinder to extract semantic information from text. In R. W. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organizations*, New Jersey: Ablex Publishing
Corporation. 149-164.

McDonald, J. E. and Schvaneveldt, R. W. (1988). The application of user knowledge to interface design. In R. Guindon (Ed.), *Cognitive Science and its Applications for Human-Computer Interaction*, Hillsdale, NJ: Erlbaum. 289-338.

McKeithen, K. H., Reitman, J. S., Rueter, H., and Hirtle, S. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.

McLean, R. S., and Gregg, L. W. (1967). Effects of induced chunking on temporal aspects of serial recitation. Journal of *Experimental Psychology*, 74, 455-459.

McPherson, S. L., and Thomas, J. R. (1989). Relation of knowledge and performance in boys' tennis: Age and expertise. *Journal of Experimental Child Psychology*, 48, 190-211.

Means, M. L., and Voss, J. F. (1985). Star wars: A developmental study of expert and novice knowledge structure. *Journal of Memory and Language*, 24, 746-757.

Meister, D. (1985). *Behavioral Analysis and Measurement Methods*, New York: John Wiley & Sons.

Miyamoto, S., and Nakayama, K. (1984). A directed graph representation based on a statistical hypothesis testing and application to citation and association structures. *IEEE*, smc-14, 203-212.

Murphy, G. L., and Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 144-155.

Musen, M. A. (1989). *Automated Generation of Model-Based Knowledge-Acquisition Tools*, Pitman, London: Morgan Kaufmann Publishers, Inc.

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*, Englewood Cliffs, New Jersey: Prentice-Hall.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.

Olson, J. R., and Rueter, H. H. (1987). Extracting expertise from experts: Methods for knowledge acquisition. *Expert Systems*, 4, 152-168.

Palmer, E. S. (1978). Fundamental aspects of cognitive representation. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and Categorization*, 259-303. Hillsdale, NJ: Earlbaum.

Pellegrino, and Kail (1982). Process of spatial aptitude. In R. S. Sternberg (Ed.), *Advances in the psychology of Human Intelligence*, New Jersey: Lawrence Erlbaum Associates. 1, 311-365.

Ploger, D. (1988). Reasoning and the structure of knowledge in biochemistry. *Instructional Science*, 17, 57-76.

Polzella, D. J., and Reid, G. B. (1989). Multidimensional scaling analysis of simulated air combat maneuvering performance data. *Aviation, Space, and Environment Medicine*, 60, 141-144.

Reitman, J. S. (1976). Skilled perception in go: Deducing memory structures from inter-response times. *Cognitive Psychology*, 8, 136-156.

Reitman, J. S., and Rueter, H. R. (1980). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology*, 12, 554-581.

Rips, L., Shoben, E., and Smith, E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.

Romesburg, H. C. (1984). *Cluster Analysis for Researchers*, California: Lifetime Learning Publications.

SAS Institute (1988). *SAS/STAT User's Guide*. Cary, NC: SAS Institute.

Schiffman, S. S., Reynolds, M. L., and Young, F. W. (1981). *Introduction to Multidimensional Scaling*, New York: Academic Press.

Schoenfeld, A. H., and Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 8, 484-494.

Schvaneveldt, R. Durso, F., Goldsmith, T., Breen, T., Cooke, N., Tucker, R., and De Maio, J. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, 12, 699-728.

Schvaneveldt, R. W., Durso, F. T., and Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, New York: Academic Press. 249-284.

Schvaneveldt, R. W., Goldsmith, T. E., Durso, F. T., Maxwell, K., Acosta, H. M., and Tucker, R. G. (1982). *Structures of Memory of Critical Flight Information*, (Report AFHRL-TP-81-46). Las Cruces, New Mexico: New Mexico State University, Department of Psychology.

Shaw, M. L. G. (1980). *On Becoming a Personal Scientist*. London: Academic Press.

Soloway, E., and Adelson, B. (1985). The role of domain experience in software design. *Transactions on Software Engineering*, SE-11, 1351-1360.

Sweller, J., Mawer, R. F., and Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, 112, 639-661.

Sweller, J., and Owen, E. (1985). What do students learn while solving mathematics problems. *Journal of Educational Psychology*, 77, 272-284.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, 69, 344-354.

Vessey, Iris (1985). Expert in debugging computer programs: A process analysis. *International Journal of Man-Machine Studies*, 23, 459-494.

Weiser, M., and Shertz, J. (1983). Programming problem representation in novice and expert programmers. *International Journal of Man-Machine Studies*, 19, 391-398.

Zeitz, C. M., and Spoehr, K. T. (1989). Knowledge organization and the acquisition of procedural expertise. *Applied Cognitive Psychology*, 3, 313-336.