Int. J. Man-Machine Studies (1985) **23**, 699-728

Measuring the structure of expertise

ROGER W. SCHVANEVELDT†, FRANCIS T. DURSO,‡ TIMOTHY E. GOLDSMITH,§
TIMOTHY J. BREEN AND NANCY M. COOKE

Computing Research Laboratory and Department of Psychology, New Mexico State University, Las Cruces, New Mexico 88003, U.S.A.

RICHARD G. TUCKER

Holloman Air Force Base, New Mexico, U.S.A.

JOSEPH C. DE MAIO

Honeywell Systems and Research Center, Minneapolis, Minnesota, U.S.A.

(*Received 25 June 1985*)

This report reviews work on defining and measuring conceptual structures of expert and novice fighter pilots. Individuals with widely varying expertise were tested. Cognitive structures were derived using multidimensional scaling (MDS) and link-weighted networks (Pathfinder). Experience differences among pilots were reflected in the conceptual structures. Detailed analyses of individual differences point to factors that distinguish experts and novices. Analysis of individual concepts identified areas of agreement and disagreement in the knowledge structures of experts and novices. Applications in selection, training and knowledge engineering are discussed.

1. Introduction

The term "expertise" refers to performance in a particular domain, such as chess, physics or medical diagnosis, that is superior to the performance of a number of other people within that same domain. Expert performance can consist of skilled motor behaviour, skill at rapidly recognizing complex patterns in the environment, skilled problem solving, decision-making skills, or a combination of these characteristics. Whereas the definition of expertise is quite straightforward, the explanation of expertise in terms of the cognitive factors that underly expert performance is not. For instance, what makes the chess master better at chess than the novice? Is it a superior short-term memory capacity, skill at perceptually analysing the chess board, or the use of different strategies? Researchers in cognitive psychology have attempted to address some of these issues concerning expertise (e.g. Chase & Simon, 1973).

The study of the cognitive factors underlying expertise has several applications. Those interested in education or training programs can benefit from this research. For instance, if it is determined that an expert organizes information in memory in a specific manner, then that organization could be explicitly conveyed to novices. Thus, an understanding of expertise can guide training programs.

† Author to whom to address correspondence and reprint requests.
‡ Visiting professor from the University of Oklahoma.
§ Now at the University of Mexico.

0020-7373/85/120699+30$03.00/0

Another timely application of expertise research is in the development of expert systems. Expert systems are knowledge-based computer systems that perform a variety of complex problem solving and decision making skills in very specific areas of expertise. The expertise embodied in an expert system consists of numerous facts and rules that are obtained from a human expert, typically through interviews with a knowledge engineer. This process of knowledge acquisition is one of the major bottlenecks in the development of an expert system (Hayes-Roth, Waterman & Lenat, 1983). Not only is the explicit listing of rules and facts by the expert tedious and time-consuming, but the verbal expression of knowledge is often very difficult. Cognitive research on human expertise provides methods for measuring knowledge structures. These methods could aid in the development of expert systems. Our goal in this project was to investigate expertise in fighter pilots using techniques from cognitive psychology.

In the past decade, cognitive psychologists have generated a considerable body of theory and data concerning the organization and retrieval of knowledge. Researchers have begun to gain an understanding of the representation of knowledge by investigating the exceptionally rich data bases of natural language and natural categories. This research has demonstrated that the organization of memory exerts important influences on the encoding and retrieval of information.

Much of the research in semantic memory has focused on the influence of semantic relatedness on the speed and accuracy with which task-relevant information can be retrieved from memory. Various terms, such as semantic similarity, semantic relatedness and semantic distance, have been used to refer to the degree to which concepts are related in meaning. The distance metaphor comes from an analogy to a multidimensional space where concepts are located according to values on various dimensions of meaning. Presumably, concepts near one another in multidimensional space are more closely related to one another than are concepts that are further apart in the space.

There have been several proposals concerning memory structures, but each one makes use of the idea that concepts in memory differ in their relatedness or psychological proximity. In network models, concepts are represented as nodes linked by labelled relations. Two concepts that are directly linked are viewed as more similar than are two concepts that are not linked or are indirectly linked (Collins & Quillian, 1969; Meyer & Schvaneveldt, 1976; Quillian, 1969). Similarly, in other network models, two concepts that share a number of links are viewed as more related than are two concepts that share fewer links (Collins & Loftus, 1975). In feature models (Rips, Shoben & Smith, 1973), where concepts are represented by vectors of features, two concepts that share a number of features are viewed as more similar than are two concepts that share few, if any, features. Both network and feature theories rely on psychological proximity to predict performance in a variety of tasks.

How theorists have determined the proximity of a particular set of concepts has varied widely. Most models applied to particular domains have relied solely, or primarily, on the intuitions of the theorists. There are, however, a number of notable exceptions. Smith, Shoben and Rips (1974) employed multidimensional scaling (MDS) procedures for a set of animal names to reveal important structural information. Similarly, Shepard (1963) and Kruskal (1977) have investigated the applicability of multidimensional spatial representations for a number of conceptual domains with some encouraging results. Often, theorists assuming a general network as an underlying

model have us
techniques, ho
data employed
Hutchinson, 19
1985). The pre
techniques.
This project
combat situatio
systematic struc
Meeting this goa
for assessing the
knowledge has,
facets of concep
not merely an a
of individuals th
of their concept
validating measu
experts should d
Several studie
structures of exp
expert–novice di
structure as a no
latter issue has in
area focus on how
in performance or
and validating str
in domains such a
1981; Chase & Si
McKeithen, Reitm
A number of co
the performance c
performance of no
in which meaning
recall the same ma
to recall positions
arranged as they v
(Chase & Simon, 1
chess novices in re
to be a greater deg
organization. Other
the experts as com
the items within a
Various explanat
suggested that the
able to encode or c
memory structure t
is not able to enco

model have used intuition or logical analysis to define network structures. There are techniques, however, that allow researchers to derive networks from the same proximity data employed by MDS (e.g. Durso, Schvaneveldt & Goldsmith, 1983; Friendly, 1977; Hutchinson, 1981; Schvaneveldt & Durso, 1981; Schvaneveldt, Durso & Dearholt, 1985). The present report makes extensive use of both MDS and network scaling techniques.

This project investigated the conceptual structures of Air Force fighter pilots for combat situations. The central goal was to demonstrate the existence and utility of a systematic structure of flight-related concepts in the memory systems of fighter pilots. Meeting this goal required applying structural analyses to data and developing methods for assessing the validity of the structural representations. The domain of fighter-pilot knowledge has, in principle, a rich cognitive component that should reflect important facets of conceptual structure in general. The domain is relatively self-contained and not merely an arbitrary subset of natural language. The domain allows identification of individuals that vary in their mastery of the domain and, presumably, in the nature of their conceptual structures. These variations in expertise provide one approach to validating measures of conceptual structure. Presumably, the conceptual structures of experts should differ systematically from the conceptual structures of novices.

Several studies in the cognitive psychology literature have investigated memory structures of experts and novices. Such studies have been concerned not only with expert-novice differences in memory structure but also with the development of this structure as a novice gains skill or experience and approaches the expert level. This latter issue has important applications in training and education. Other issues in this area focus on how experts organize information in memory, expert-novice differences in performance on recall and perceptual tasks, and methods for measuring, representing and validating structures of memory. Expert and novice studies have been conducted in domains such as chess, bridge, Go, physics and computer programming (Adelson, 1981; Chase & Simon, 1973; Chi, Feltovich & Glaser, 1981; Engle & Bukstel, 1978; McKeithen, Reitman, Rueter & Hirtle, 1981; Reitman, 1976).

A number of conclusions has emerged from expert-novice research. By definition, the performance of experts on the actual task in which they excel is superior to the performance of novices. Experts, however, show superior performance on recall tasks in which meaningful material is used, but are no better than novices when asked to recall the same material in a random arrangement. For instance, chess masters are able to recall positions of pieces on a chess board much more readily if the pieces are arranged as they would be in a game situation, rather than a random arrangement (Chase & Simon, 1973). However, meaningful arrangement of the pieces does not aid chess novices in recalling positions. Furthermore, as experience increases, there tends to be a greater degree of intragroup agreement, in relation to memory structure and organization. Other common findings include a larger chunk size and more chunks for the experts as compared with the novices. Chunks are units of information in which the items within a chunk are related to each other in a meaningful fashion.

Various explanations have been offered for these findings. Typically, it has been suggested that the expert is able to perceive a more global picture and, therefore, is able to encode or chunk items into larger units than is the novice. The novice has a memory structure that is not as highly organized as that of the expert and, therefore, is not able to encode as quickly or in as large units. It has also been suggested that experts have a memory structure that is hierarchically organized and, therefore, can recall a larger number of chunks. Thus, a high-level chunk may consist of a set of lower-level chunks, each containing additional chunks at a more detailed level. This hierarchical organization can structure large amounts of information effectively.

The central hypothesis underlying our work is that the organization of information in memory has a critical impact on flying performance. Understanding how critical information is organized in memory can be extremely useful in designing training programs and in increasing the effectiveness of the pilot–aircraft system. Knowledge of the systems individuals develop for organizing critical information can be used to tailor training systems to provide students the conceptual framework that will lead to expertise. It may also provide a useful evaluation and selection tool by allowing instructors to determine which individuals have mastered the prerequisite concepts for success in a particular training program. Finally, using scaling techniques to extract knowledge may provide a favorable alternative to the interview method in knowledge engineering.

2. General methodology

2.1. SUBJECTS

Three populations of fighter pilots were sampled for these studies. Ten instructor pilots (IPs) stationed at Holloman Air Force Base (AFB) and nine Air National guard pilots (GPs) from Buckley Air National Guard base served as the two groups of expert pilots. The IPs averaged 2583 h flying time and served as instructors for lead-in fighter training. The GPs averaged 6064 h flying time but were not classroom instructors. The third sample consisted of 17 undergraduate pilot trainees (UPs) stationed at Williams AFB. The UPs averaged 200 h of flying time and had recently completed Undergraduate Pilot Training which precedes advanced training with specialty aircraft. Thus, none of the UPs had undergone fighter lead-in training. This choice of subjects seemed to be appropriate because they were expected to exhibit some, but certainly not all, of the features characteristic of expert fighter pilots. In particular, the UPs should have a good command of general flying procedures (e.g. formation flying) but little or no command of air-to-air or air-to-ground combat situations.

2.2. MATERIALS

The development of the stimulus materials began with a task analysis of tactical flight maneuvers (Meyer, Laveson, Pape & Edwards, 1978). Based on the Meyer report and through interviews with four pilots from the 449TTW, two scenarios together with lists of assumptions, basic concepts, and related concepts were selected. One scenario involved split-plane maneuvers in air-to-air combat, and the other scenario focused on the low-angle strafe maneuver in air-to-ground combat. These two scenarios were chosen, in part, because they differ in inherent complexity. The split-plane scenario is considerably more complex, involving several possible configurations of aircraft, instruments, and possible actions. In contrast, the strafe scenario is inherently simple, involving a single aircraft and a limited number of actions. In addition, each of these scenarios involves some concepts that should be well understood by the UPs and some that should be relatively foreign to the UPs who have not received fighter lead-in training.

Each scenario consisted of a set of assumptions and 30 basic concepts. The two scenarios appear in Table 1. The basic concepts served as the critical stimuli for the

TABLE 1
Two scenarios with assumptions and basic terms

Assumptions	Basic concepts	
(*a*) *Split-plane maneuvers*		
OFFENSIVE	LOW YO YO	HIGH YO YO
KILL	LAG ROLL	BARREL ROLL
SINGLE BANDIT	GUNS	AIRSPEED
AGGRESSIVE	G LOADING	CUTOFF
COMMIT	6 O'CLOCK	SMASH
IR MISSILE PARAMETERS	SWITCHOLOGY	RADIAL G
TALLY HO	HEAT	SNAPSHOT
ENGAGED	EXTENSION	ANGLE OFF
SIMILAR AIRCRAFT	QUARTER PLANE	ASPECT ANGLE
DEFENSIVE TURN	OVERTAKE	PURE PURSUIT
	CORNER VELOCITY	LEAD PURSUIT
	RELATIVE ENERGY	LAG PURSUIT
	POWER SETTING	LIFT VECTOR
	ACCELERATION	3–9 LINE
	VERTICAL MANEUVERING	
	WEAPONS PARAMETERS	
(*b*) *Strafe maneuver*		
CLEARED	BULLET IMPACT	DIVE ANGLE
PANEL-TARGET	AIM OFF POINT	CLOSURE
SWITCHOLOGY	AIRSPEED	ALTITUDE
TARGET ACQUISITION	BANK	WALKING
CONTROLLED RANGE	PIPPER PLACEMENT	TRACKING
	RICOCHET	FINAL
	BURST	BUNT
	STABILIZE	FOUL
	GLIDEPATH	DRIFT
	FOUL LINE	GUNS
	RUN-IN LINE	AIM POINT
	PIPPER FIXATION	RANGE
	TRIGGER	PULL-UP
	YAW	FIRE
	RECOVERY	
	TRIM	

study. Thus, 30 concepts important to air-to-air combat and 30 concepts important to air-to-ground combat were examined in these studies.

2.3. PROCEDURE

Most scaling procedures for producing structural descriptions of a set of concepts require some measure of psychological distance between the concepts. Although two general methods have been used in the literature, inter-item distance in recall protocols (e.g. Adelson, 1981) and direct judgments of pairwise similarity or dissimilarity (e.g. Rips *et al.*, 1973), recent work by Cooke (1983) suggests that the latter measure provides a more sensitive and valid database for the delineation of conceptual structures. In the present project, measures of proximity were based on pairwise similarity or relatedness judgments.

Ratings

The subjects were told about similarity or relatedness ratings and the mechanical details of entering ratings on a TERAK microcomputer. The scenario was then described to provide a context for rating the basic terms, and the complete set of terms to be rated was shown to allow subjects to establish some criteria for rating the pairs of concepts.

The rating task itself consisted of presenting all possible pairs of the 30 basic concepts. Subjects rated the similarity or relatedness of 435 pairs of terms (i.e. 30 taken two at a time) during the session. For each pair of terms, the TERAK displayed the pair, a rating scale with the digits 0–9, and a bar marker to indicate the rating. Subjects were instructed that a number of factors might enter into a decision about similarity, including relatedness, co-occurrence, dependency and contingency. They were told that the purpose was to obtain their general impressions of the relatedness of the items and that they should not ponder their judgments. Subjects entered their rating by pressing a number key on the TERAK keyboard. The bar marker in the display was moved to the position corresponding to the number entered by the subject to indicate the rating given. The subject could change the rating by pressing another number key. When the subject was satisfied with the rating, pressing the space bar on the keyboard changed the display to show the next pair of items and reset the marker to the bottom of the scale. This procedure was followed until all 435 pairs had been presented. The order of the pairs was randomized for each subject, and the position of the two items in a pair was counterbalanced across subjects. A rating session required from 30 to 45 minutes to complete.

Familiarity ratings

In addition to the rating task, the UP subjects were asked to rate their familiarity with each of the terms. UPs rated each concept on a scale of 1–3, where 1 indicated no familiarity, 2 indicated familiarity, and 3 indicated the concept had been used in flying.

Data sets

Seven of the 10 IPs, each of the GPs, and each of the UPs supplied data in the split-plane scenario. For the strafe scenario, data were collected from six of the IPs and 16 of the UPs; no data were obtained from the GPs for the strafe scenario.

The obtained similarity measures were transformed into measures of psychological distance by subtracting the ratings from the maximum possible rating. The resulting numbers reflect distance, with the larger numbers representing greater psychological distance between concepts. For each scenario, for each subject, the data were placed in a 30×30 symmetrical matrix where all entries, other than the diagonal, represented the empirical judgment for a pair of concepts. Similar matrices of means were computed for each scenario and each group of subjects.

3. Measurement of conceptual structure

The goal of scaling procedures is to uncover latent structure in data. The structure is masked by the noise found in any set of empirical data. The scaling procedures assume that the latent structure obeys the assumptions of metric data, regardless of whether the empirical judgments meet these assumptions. For example, the latent structure is assumed to obey the triangle inequality assumption even though the empirical data contain violations of this assumption. In fact, the scaling procedures either manipulate the data to meet these assumptions, or they extract the parts of the data that meet these assumptions. Underlying these approaches is the belief that violations of the assumptions are due to noise and not to any meaningful psychological property (but see Tversky, 1977).

The differences among scaling procedures usually lie in the products of the procedures. There are procedures for generating hierarchical clusters, additive clusters, weighted free trees, multidimensional spaces, and link-weighted networks. Here we focus on the latter two procedures. MDS is a procedure that produces spatial configurations and has undergone conceptual, mathematical, and empirical scrutiny in a number of studies. Networks complement spatial representations in a number of ways and have distinct advantages over other non-spatial scaling procedures.

3.1. METHODS

Multidimensional scaling

MDS is a powerful technique for extracting the latent structure within the empirical similarity judgments. This is accomplished by arranging the concepts in n-dimensional space where the distances between points reflect the psychological proximity of the concepts.

The first step in obtaining such representations involves submitting the empirical ratings, along with the desired dimensionality, to an MDS algorithm. Optimal dimensionality was determined by a number of factors: stress and R^2 tended to elbow at three dimensions; the addition of a third dimension clarified the prior ones and was itself interpretable (Shepard, 1974); and the Isaac and Poor (1974) procedure suggested three dimensions. Once the data and desired dimensionality have been entered, the MDS algorithm returns a set of co-ordinates corresponding to the location of each concept in the space. The final step involves interpreting the resultant space along with the accompanying dimensions.

MDS supplies several useful pieces of information. First, it summarizes the data into a spatial configuration, which is complex at times, but is considerably more informative than are the numerous empirical judgments. Second, MDS captures the global relations among the concepts. That is, MDS considers the relationship of each concept to all other concepts and places the concepts along the dimensions of the space in a way that reflects these relations. Although such a procedure can distort local relationships (that is, the distance between any particular pair), the procedure is unsurpassed at revealing global structure. In particular, successful identification of the dimensions of the space supplies information about conceptual structure that cannot be gleaned from the original ratings nor from other scaling techniques. Finally, MDS supplies a metric (distance between concepts in multidimensional space) that has some useful applications.

Although a concept can be located in multidimensional space by a series of co-ordinates (one for each dimension), the co-ordinates are not as useful as are distances in comparing representations. The distance between a pair in MDS may be based on the Euclidean distance between the two points located by the co-ordinates corresponding to each concept. Euclidean distances in MDS preserve the structure of the representation, are independent of dimensional rotation, and meet the three standard metric assumptions: identity, non-negativity and triangle inequality.

To illustrate MDS, consider the two-dimensional spatial configuration in Fig. 1(a). The solution was based on the pairwise similarity judgments of undergraduate psychology majors for 16 naturally occurring objects. MDS has summarized the 16×16 matrix of distances into a representation that allows consideration of the global relations among concepts. In particular, the horizontal dimension reflects a non-living-living dimension; the vertical dimension is more difficult to label but seems to have captured a difference between plants and animals, although this applies to only the living members of the space. In order to fix the concepts in space, MDS has introduced some local distortions. Maple is closer in space to rose than it is to tree. These local distortions are due to the adjusting of the data that occurs in order to produce distances that obey the metric assumptions. It will be seen that network structures complement MDS nicely by supplying information about local relations among the concepts.

Link-weighted networks

A link-weighted network is a configuration in which concepts are depicted by nodes, and relationships are depicted by links connecting the nodes. The links are assigned a value or weight that reflects the strength of the relationship between the nodes. The value reflects the distance from one node to another along that link; the shorter the link, the closer the nodes. The network is general in that constraints are not placed on the possible relations that can be represented. For example, the hierarchical constraint found in cluster analysis is not placed on networks. Without this constraint, the representation is free to contain local relations other than hierarchical ones, although hierarchical relations may still be present (Christofides, 1975; Fillenbaum & Rapaport, 1971).

As has been noted, networks have formed the basis of research in a number of areas of cognitive science. Several psychological and artificial intelligence models of conceptual structure are based on such networks. The area of mathematics called graph theory is centrally concerned with properties of general networks. Although important theoretical and formal work has been conducted on these structures, scaling methods that yield general network structures have only recently appeared. Pathfinder (see Schvaneveldt, Durso & Dearholt, 1985), an algorithm that produces general weighted networks, was applied to the rating data.

Pathfinder produces a network with concepts represented as nodes and relations between concepts represented as links connecting some of the nodes. Links may be either directed (allowing traversal in only one direction) or undirected (allowing traversal in either direction). Thus, distances between concepts may be either symmetrical or asymmetrical. Of course, with symmetrical distance estimates, only undirected links can be included in the network representation.

The definition of Pathfinder networks is quite simple. We begin with a network in which each node is connected by a link to each other node for which we have distance

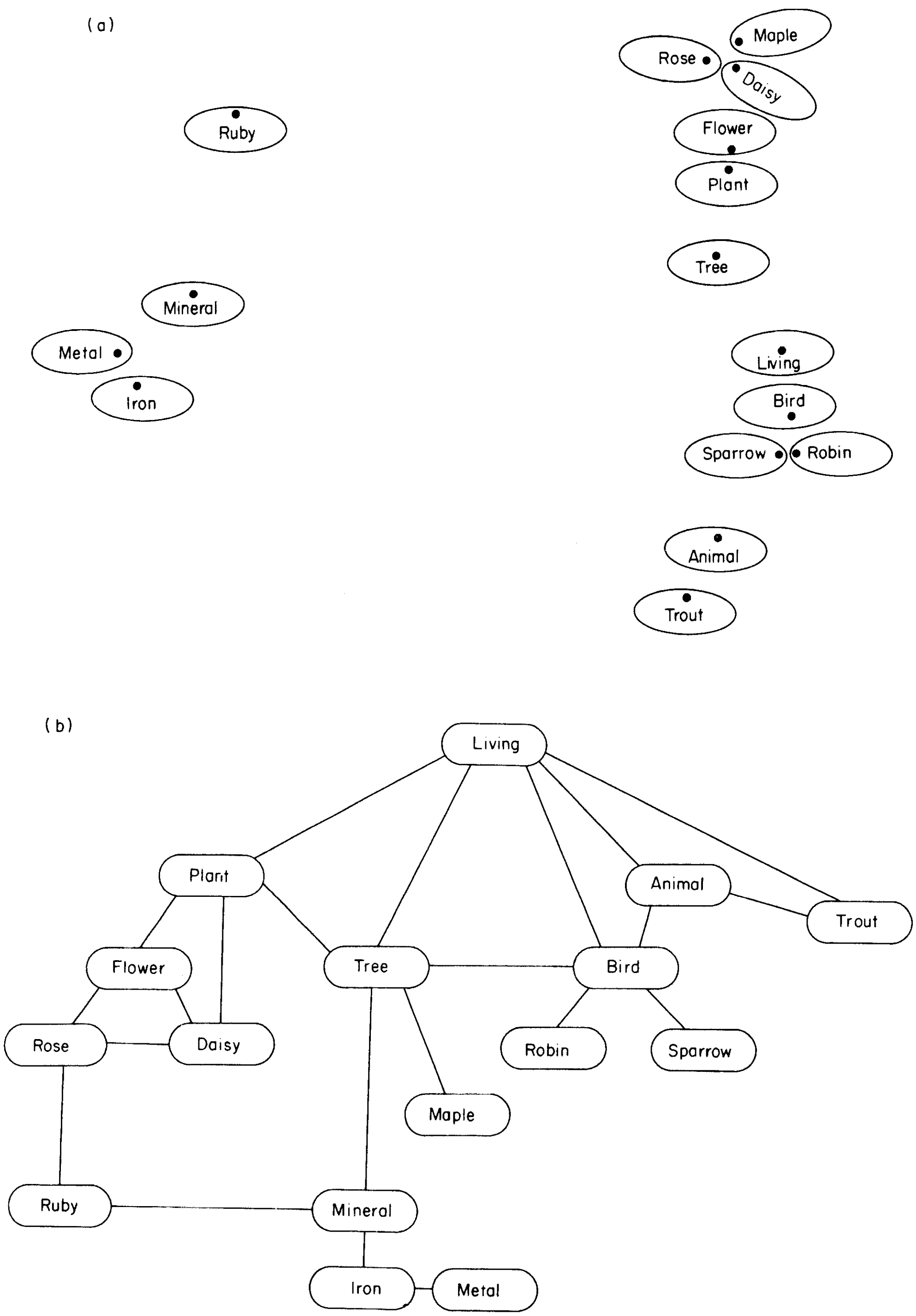


FIG. 1. (a) Two-dimensional MDS solution for 16 natural concepts. (b) Pathfinder network representation of the same concepts. Link weights have been omitted.

estimates. The distance estimate is associated with each link giving a weight or cost for the link. With a complete set of distance estimates, this network will correspond to a complete graph. After applying the Pathfinder algorithm, a link remains in the network if and only if that link is a minimum-length path between the two concepts. The length of a path is a function of the weights associated with the links in the path. Different functions for computing path length yield different networks. In particular, the number of links in the resulting network will decrease systematically with decreases in the computed lengths of multi-link paths in the network. Two methods have been used to define path length. One method, which subsumes several special cases, uses the Minkowski r-metric to compute path length. Although the r-metric was originally developed as a generalized distance measure in multidimensional space, it can also provide a general definition of the length of a path in a network. Let l_i be the weight associated with link i in a path. The set of all weights in a path with n links is given by l_i, $i = 1, 2, \ldots, n$. The length of the path, $L(P)$, is given by:

$$L(P) = \left(\sum_{i=1}^{n} l_i^r \right)^{1/r}, \quad 1 \leq r \leq \infty.$$

As the value of r varies over the allowable range, the number of links in the resulting networks varies systematically. In particular, as r decreases, additional links are added, but all links in networks with larger values of r are still included. With $r = \infty$, Pathfinder will produce a network which is the union of the minimal spanning trees for the network defined by the data (a complete graph if all pairwise distance estimates are available). The minimal spanning tree will be unique unless certain patterns of ties occur in the data. With $r = 1$, Pathfinder will simply use the sum of the link weights to determine the length of a path in the network. Intermediate values of r produce networks with intermediate numbers of links.

A second method for computing path lengths follows from the theory of spreading activation in network structures. This method computes the length of a path by summing the link weights in both directions, starting from the nodes at each end of the path. The path length is taken as the maximum sum to the node where the two summations intersect. This method is analogous to measuring the maximum distance to the intersecting node when the path is traversed simultaneously from both ends. This node would be the point where spreading activation from the two end nodes would meet, and the length would be the maximum distance travelled by the activation. This method for computing distance is called the parallel method because the distances are traversed in parallel from both ends of the path. The parallel method yields networks with an intermediate number of links compared to the Minkowski method with $r = \infty$ and $r = 1$.

A family of Pathfinder networks can be generated by varying both the function defining path length and the maximum number of links in paths. The numbers of links in a particular network varies systematically as a function of the values of these two parameters. Schvaneveldt *et al.* (1985) provide additional details. The networks presented here were all generated using the parallel definition of path length.

The Pathfinder algorithm can be illustrated by the natural concepts used to illustrate the results of MDS. The Pathfinder solution is shown in Fig. 1(b). Note that maple and daisy are not linked. A path existed in the complete graph (distance estimates) that was shorted than the direct distance between maple and daisy so the direct link was not included. Notice that maple and daisy were close according to MDS but were not linked by Pathfinder. Because Pathfinder extracts the latent structure rather than transforming the data, it is better able to reflect psychological proximity on a pairwise basis. On the other hand, Pathfinder does not produce global information of the kind supplied by MDS.

The use of general networks and the Pathfinder algorithm holds substantial promise in attempts to specify the local relations and structure present in a conceptual organization. In addition, it allows a detailed, concept-by-concept comparison across groups that differ in expertise.

3.2. RESULTS AND DISCUSSION

MDS

In this section, the properties of the specific MDS solutions for the pilot ratings are discussed. These solutions were based on the ordinal option in the ALSCAL collection of algorithms. Three dimensions supplied the optimal dimensionality for both the split-plane concepts and low-angle strafe concepts for all groups of pilots. The strafe scenario, because it is inherently less complex, had been expected to yield a smaller dimensionality. Perhaps when subjects are required to consider a single maneuver, nuances in the maneuver receive more attention than would be the case from a more global point of view. In addition, it was somewhat surprising that the novices (UPs) and experts (IPs and GPs) each resulted in a solution of equivalent dimensionality. In the discussion of Pathfinder, it will be suggested that this may be a limitation of the MDS scaling procedure used here, rather than a suggestion of equally complex solutions for novices and experts.

For illustration, the three-dimensional scaling solution of split-plane maneuvers for IPs is shown in Fig. 2. Figure 2(a) presents the position of each concept along the first two dimensions, Figure 2(b) presents the position of each concept along the first and third dimensions, and Fig. 2(c) presents the position of each concept along the second and third dimensions.

In order to identify the dimensions, assistance was obtained from experts at Holloman AFB and Williams AFB. Each of the dimensions has been identified for the split-plane solution, and one of the dimensions has been identified for the strafe solution. The split-plane concepts have one dimension associated with a temporal factor, one dimension which distinguishes particular maneuvers, and one dimension associated with factors distinguishing concepts that are related to distance from concepts related to orientation. The temporal dimension identifies the general time dimension within a scenario leading to split-plane maneuvers. In Figs 2(a) and (b), this dimension is the horizontal dimension ordered from left to right. The concepts on the extreme left (SWITCHOLOGY, HEAT and ANGLE-OFF) refer to events or considerations that occur early in the temporal sequence. To the right, the concepts refer to events and considerations occurring later in the sequence. Concepts occurring later in the sequence are actually consequences of actions performed early in the sequence. The second dimension is a contrast between lead pursuit and lag pursuit with LAG PURSUIT and the associated maneuvers near the top and LEAD PURSUIT and LOW YO YO near the bottom in Fig. 2(a). This dimension is again represented along the horizontal dimension in Fig. 2(c). The third dimension is the vertical dimension in Figs 2(b) and (c). This dimension has been tentatively identified as separating concepts that refer to actions and considerations related to the range or distance between aircraft from concepts related to the relative positions of the aircraft. This dimension separates concepts that concern distance from concepts that concern orientation.

The low-angle strafe maneuver also yielded a temporal order dimension in the solution. Again, this dimension occurred as the first dimension in the solution, and it reflects the order in which the concepts would occur to pilots in executing the low-angle strafe. Interestingly, these temporal dimensions appear to reflect the psychological

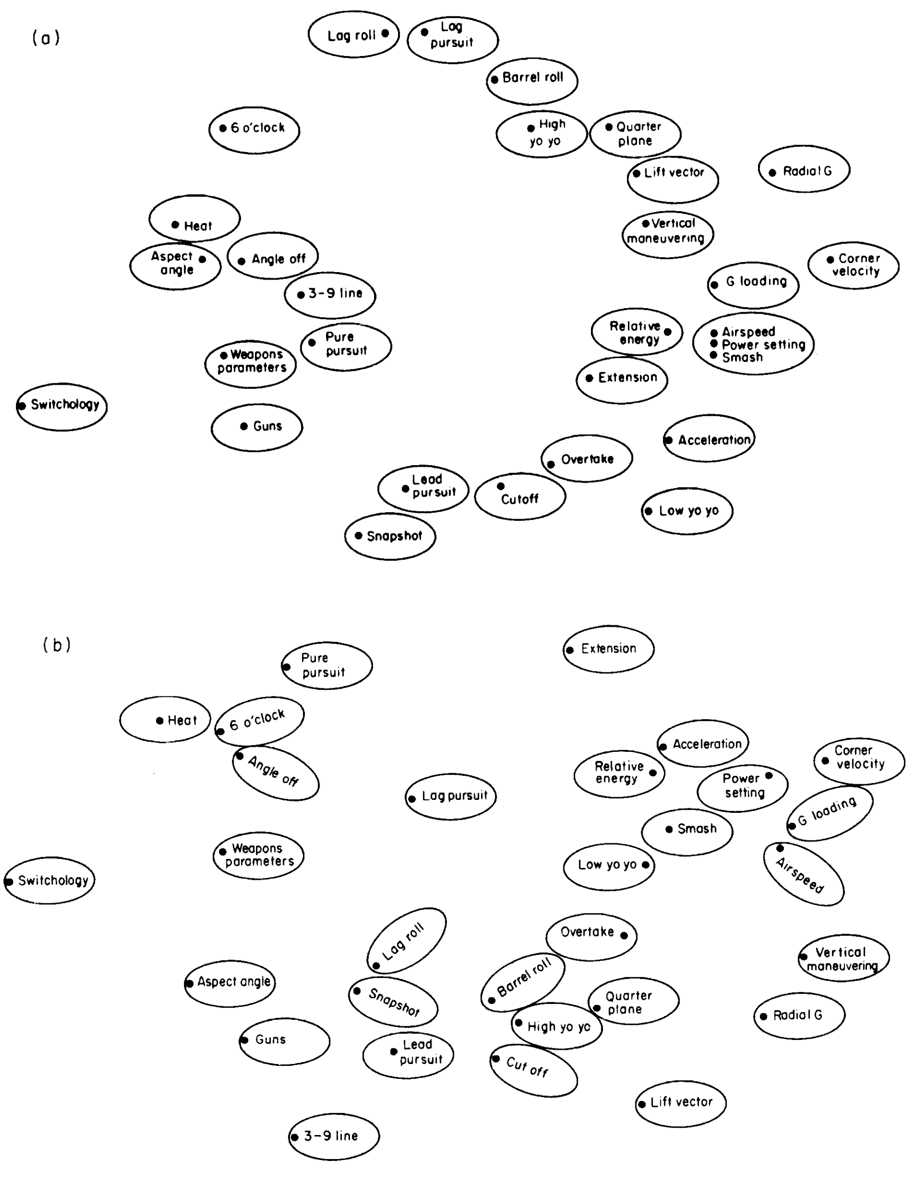
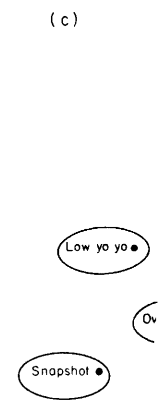


FIG. 2. Two dimensional projections of a three-dimensional MDS solution for the IP ratings of the split-plane concepts. (a) dimensions 1 (horizontal) and 2 (vertical). (b) dimensions 1 (horizontal) and 3 (vertical). (c) dimensions 2 (horizontal) and 3 (vertical).



ordering of the
time. Apparent
occur, in order
The temporal
for pilots.

The dimen
tentative sup
solutions. M
clusions that
section of th

Pathfinder
An analysis
maneuvers
networks, s
nodes is arl
of the pos
dimension

Consist
complexit
confirmin
However,
do differ
is that th
the IP n
network
a struct

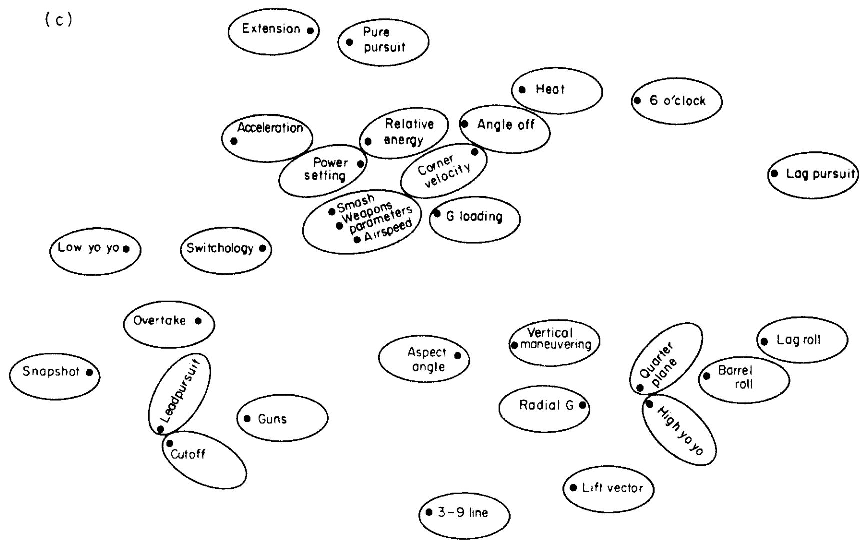


FIG. 2 (*cont.*)

ordering of the concepts rather than the order in which the events occur in physical time. Apparently, pilots must consider several factors early in time, before they actually occur, in order to be able to concentrate on critical factors such as aiming and firing. The temporal order dimension is a powerful one in the organization of these concepts for pilots.

The dimensional organization of the concepts is interesting, and it lends some tentative support to the validity of the analytic procedures underlying the MDS solutions. More fine-grained analyses of the structures are required to lead to conclusions that may be usefully applied. The metric-based analyses considered in the section of this report on validation represent a step in that direction.

Pathfinder

An analysis was performed on the data from UPs, IPs and GPs for split-plane maneuvers and the low-angle strafe maneuver using Pathfinder. To illustrate the networks, solutions for IPs and UPs appear in Figs 3 and 4. Since the location of nodes is arbitrary (the information in the network is contained in the links), the layout of the positions of nodes is the same for the two networks. Incidentally, the two-dimensional MDS solution for the IPs was used to locate the nodes on the page.

Consistent with the MDS analyses, Pathfinder supplied networks of comparable complexity (i.e. number of links) for the split-plane and the strafe scenarios, disconfirming the expectation that the strafe maneuver would be viewed as less complex. However, Pathfinder does suggest that conceptual structures of experts and novices do differ in complexity. The most striking difference between the IP and UP networks is that the network derived from the student data is considerably more complex than the IP network. The UP network has 51 links compared with 40 for the instructor network. This pattern is even more extreme for the strafe concepts, with IPs producing a structure of 39 links compared with 65 links for the UPs. This result can be contrasted

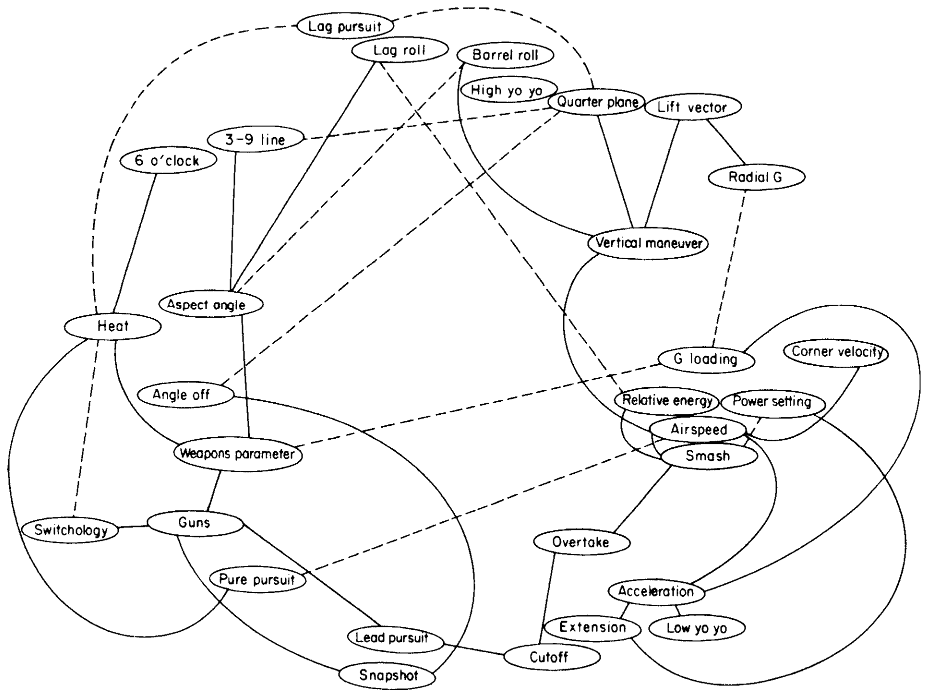


FIG. 3. Pathfinder network representation of the IP ratings of the split-plane concepts. Link weights have been omitted.

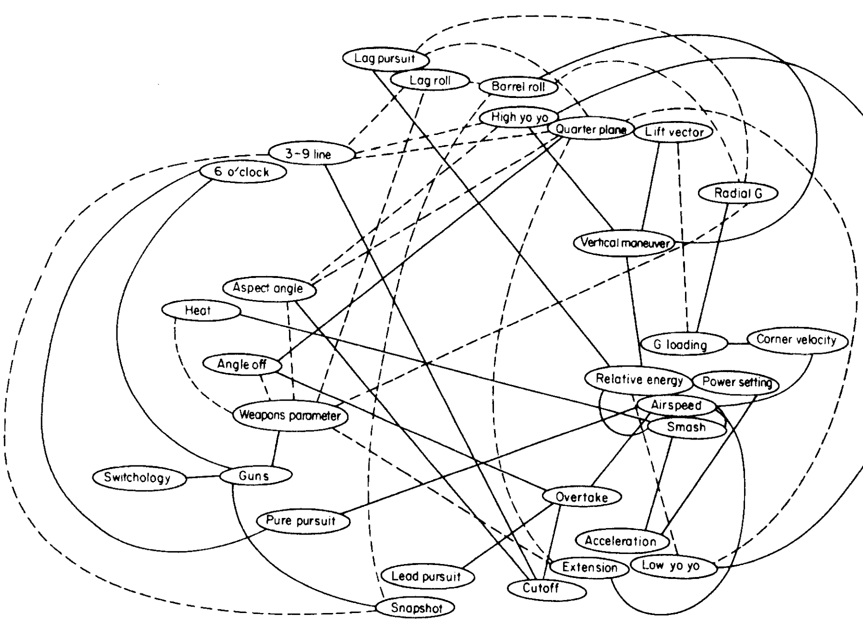


FIG. 4. Pathfinder network representation of the UP ratings of the split-plane concepts. Link weights have been omitted.

STRUCTURE OF

with MDS whic
a characteristic
to identify the
network.
Pathfinder re
For example, tl
ASPECT ANC
AIRSPEED), a
hierarchical. In
number of loca
scaling technic

3.3. SUMMARY

Both MDS an
matrix to a m
the underlyin
whereas MDS
to find the be
information.
related and s
the rating da
to meet the
distance inf
information
space, as dic

4. Validati

The previou
MDS and I
concepts w
comparing
among gro
difference i
their conce
of pilots s
conceptua
identified
support th
ing differe
cedures p
Thus, the
structures
within gr

4.1. DISCI

In additi
are avail

with MDS which yielded three-dimensional solutions for all pilot groups. Apparently a characteristic of expertise is not a more complex structure, but rather, experts tend to identify the important, critical information and associations, yielding a simpler network.

Pathfinder reveals interesting structural facets of the conceptual structure of the IPs. For example, there are several concepts that link to multiple concepts (GUNS, HEAT, ASPECT ANGLE, VERTICAL MANEUVERING, ACCELERATION, SMASH, AIRSPEED), and the presence of cycles suggests that the representation is not strictly hierarchical. In addition, the network of split-plane concepts for the IPs highlights a number of local relationships (to be discussed later) that are not apparent in the MDS scaling techniques that were used here.

3.3. SUMMARY

Both MDS and Pathfinder reduce a large amount of data in the form of a dissimilarity matrix to a much smaller set of data, but they tend to highlight different aspects of the underlying structure. Pathfinder focuses on the local relationships among concepts, whereas MDS provides more global information about the concept space. In the effort to find the best Euclidean fit to the data, MDS sacrificed some of the pairwise distance information. MDS placed concepts near each other in space that were not viewed as related and separated related concepts. Because the network extracts information from the rating data that follows the metric assumptions rather than altering the rating data to meet the metric assumptions, the links present in the network highlight pairwise distance information. On the other hand, Pathfinder does not supply the global information that led to the identification of underlying dimensions of the conceptual space, as did MDS.

4. Validation

The previous sections have described techniques for representing conceptual structures. MDS and Pathfinder both produce relational and organizational information about concepts within a particular domain of knowledge. One means of validating and comparing MDS and Pathfinder is to use these conceptual structures to discriminate among groups and to predict group membership. It is reasonable to assume that difference in experience among the pilots should be mirrored in differences among their conceptual structures. It is also reasonable to assume that members of a group of pilots share certain characteristics in their conceptual structures. Given an IP's conceptual structure for the split-plane maneuvers, can this individual be correctly identified as an IP? Accurate classification based on conceptual structures would support the validity of the structure. Classification also provides one means for comparing different ways of defining conceptual structures. Furthermore, classification procedures provide a means for assessing the nature of group and individual differences. Thus, the interest in this phase of the project is in evaluating the validity of conceptual structures and in assessing similarities and differences of structures both across and within groups of pilots.

4.1. DISCRIMINATION

In addition to arranging a set of concepts in multidimensional space, scaling techniques are available for placing individuals or groups of individuals in multidimensional space. Carroll and Chang's (1970) individual differences MDS can be used to locate individuals along the same dimensions in which concepts are placed. Thus, the output of this procedure illustrates which dimensions are most highly weighted for specific groups or individuals. For instance, a point located close to the zero co-ordinate for a particular dimension indicates this dimension was not critical for that particular group. The ALSCAL version of Carroll and Chang's INDSCAL program is used here.

The location of groups of pilots along the three dimensions used previously in this work is not only useful in further distinguishing groups from each other, but is also a valuable technique for validating the dimensions. If the resulting three dimensions are meaningful, it is expected that these dimensions would be more critical to expert pilots than to the novices who lack the understanding and organization of the concepts found in more experienced pilots.

Method and results

In order to plot the three groups of pilots in multidimensional space, the distance matrices for each of the IPs and each of the GPs for each scenario were submitted to an individual differences scaling procedure. This yielded an expert space. The dimensions found earlier with classical MDS were mirrored in the INDSCAL solution. The distance matrix from each UP, one at a time, was added to the distance matrices of the experts, and then the space was recomputed. The dimension weights for each UP were recorded and compared with the dimension weights for the experts as derived from the expert space. Results shown in Table 2 indicated that UPs weight each

TABLE 2

Weighting of dimensions for each group of pilots for split-plane and strafe scenarios

Dimension	IPs	GPs	UPs	Mean
(*a*) *Split-plane manoeuvres*				
1. Temporal	0·3475	0·3023	0·1923	0·2807
2. Distance-orientation	0·3142	0·2856	0·1779	0·2592
3. Lead-lag	0·3270	0·2717	0·1761	0·2583
Mean	0·3296	0·2865	0·1821	
(*b*) *Strafe manoeuvre*				
1. Temporal	0·3819		0·1844	0·2832
2. Unknown	0·3327		0·1833	0·2580
3. Unknown	0·3020		0·1687	0·2354
Mean	0·3389		0·1788	

dimension less than do experts. Planned comparisons setting experiment-wise alpha to 0·05 (test-wise alpha = 0·006) confirmed these conclusions in each case. For the split-plane scenario, experts relied more heavily on the temporal dimension, $t(32) = 7{\cdot}84$, the orientation-distance dimension, $t(32) = 7{\cdot}73$, and the lead-lag dimension, $t(32) = 6{\cdot}98$, than did the novices. Similarly, in the strafe scenario, the temporal dimension was weighted more by experts, $t(21) = 10{\cdot}33$, as were the two unidentified dimensions, $t(21) = 6{\cdot}32$ and $t(21) = 8{\cdot}52$, for dimensions 2 and 3, respectively.

As mentioned prev
expected if the dimen
of the dimensions ar
appears that none of
experienced groups.
Secondly, results of
discriminable: IPs ten
than did GPs. This ten
$t(14) = 2{\cdot}25$, $p < 0{\cdot}05$. S
that the groups have
these dimensions recei
tendency for IPs to use
experience.

4.2. CLASSIFICATION

In the previous section,
of their conceptual struct
as a member of a partic
then pattern classificatio
Classification procedu
object to one of two or n
variables. The groups ar
Objects are represented b
of classification involves
in order to locate the clos
There are many differe
on discriminant analysis. I
be met including: (a) mea
functions are computed w
objects must exceed the n
linear combination of oth
group are equal; and (e) g
Because these assumptio
cedure based simply on dis
discussion of this technique
and correlations do not ent
objects to be categorized ar
pattern vector. The *i*th elen
Because feature values are in
as points in a multidimensi
of the object. The goal is t
pattern space into regions
particular class of patterns.
Linear discriminant func
linear combination of the f
function has the form:
g

As mentioned previously, heavier weighting of the dimensions by experts would be expected if the dimensions had a psychological validity; UPs lack the understanding of the dimensions and organizational structure of the more experienced pilots. It appears that none of the dimensions are as salient for the UPs as for the more experienced groups.

Secondly, results of this scaling indicated that the two groups of experts are also discriminable: IPs tended to weight each dimension in the split-plane solution more than did GPs. This tendency was, however, only significant for the lead-lag dimension, $t(14) = 2{\cdot}25$, $p < 0{\cdot}05$. Some differences between IPs and GPs would be expected given that the groups have dissimilar backgrounds. The initial qualified interpretation of these dimensions receives some support from the INDSCAL findings. The overall tendency for IPs to use the dimensions more than do GPs may reflect their classroom experience.

4.2. CLASSIFICATION

In the previous section, it was shown that the groups were discriminable on the basis of their conceptual structures. Here, methods are developed for classifying an individual as a member of a particular group based on the individual's conceptual structure and then pattern classification techniques are used to analyse conceptual structures.

Classification procedures are generally concerned with the problem of assigning an object to one of two or more groups. The groups may vary along several attributes or variables. The groups are defined such that each object belongs to only one group. Objects are represented by a list of numerically described attributes. The general notion of classification involves comparing each object's position to each group's prototype in order to locate the closest group.

There are many different classification techniques. One common method is based on discriminant analysis. Discriminant analysis requires that a number of assumptions be met including: (a) measurement at the interval or ratio level because discriminant functions are computed with means, variances and correlations; (b) the number of objects must exceed the number of attributes defining an object; (c) no attribute is a linear combination of other attributes; (d) population covariance matrices for each group are equal; and (e) group populations have multivariate normal distributions.

Because these assumptions are excessively restrictive, we used a classification procedure based simply on distances in feature space. Nilsson (1965) provides a general discussion of this technique which differs from discriminant analysis in that variances and correlations do not enter into the classification procedure. With this technique, objects to be categorized are represented by a list of feature values in the form of a pattern vector. The *i*th element of the vector represents the value of the *i*th feature. Because feature values are in the form of real numbers, pattern vectors can be considered as points in a multidimensional space where each dimension represents an attribute of the object. The goal is to develop discriminant functions that will partition the pattern space into regions containing only those points or patterns belonging to a particular class of patterns.

Linear discriminant functions (the only type used here) assume that a weighted linear combination of the feature values can classify patterns. A linear discriminant function has the form:

$$g(x) = W_0 + W_1X_1 + \cdots + W_dX_d.$$

where W_i represent the weights and the X_i represent the feature values of the d features. Classes that can be properly separated with linear discriminant functions are known as linearly separable. The first analysis consisted of determining a discriminant function to classify all but one person from each of two groups. We then attempted to classify the remaining two individuals. This procedure was repeated for all possible combinations of $n-1$ people from one group and $n-1$ people from a second group.

The method for generating linear discriminant functions began with minimum distance classification. In this case, a prototype point representing the central tendency of a class of patterns is constructed for each group of $n-1$ individuals. The prototype is simply the average of the feature values of all patterns belonging to a group. The initial linear discriminant function, or the decision surface separating the patterns, was the perpendicular bisector of a line connecting the two prototype points.

If this initial function successfully classifies all the n_1-1 and n_2-1 "known" individuals, it then attempts to classify the two "unknown" individuals. If, however, the starting function fails to classify correctly the training set of known individuals, a training procedure alters the function by successive adjustments to the weight vector W by adding a fraction of the pattern vector (X) that was incorrectly classified to the weight vector. This produces a new weight vector $W' = W + cX$, where c is a positive number that controls the extent of the adjustment. The procedure is terminated as soon as the weight vector correctly classifies all patterns in the training set. The remaining two individuals are then classified.

These pattern classification techniques can provide one measure against which to test the validity of representations of conceptual structure. For example, assuming that there are differences in the way novices and experts organize knowledge about the area in question, empirical methods of representing memory organization should reflect this difference in expertise. Pattern classification techniques can then be used on information derived from empirically generated representations of conceptual structure, in an effort to classify individuals as novices or experts.

Method

Pattern classification analysis was performed on data obtained from GPs, IPs and UPs for split-plane maneuvers, and on IPs and UPs for low-angle strafe maneuvers. Three types of patterns were generated for each individual, based on MDS, networks and raw ratings. Network patterns were formed for each individual by taking the presence or absence of links in the network for each pair of concepts. The network pattern for each subject consisted of a vector of ones and zeros representing the presence or absence of a link, respectively, between each pair of concepts. An MDS pattern consisted of a vector of the distances between the members of each pair of concepts in multidimensional space. Patterns based upon the original ratings were formed by considering the similarity rating for each pair of concepts as a feature of a pattern. All three methods resulted in patterns with 435 features corresponding to all the possible pairs of 30 concepts.

If Pathfinder and MDS successfully capture the latent structure within the ratings, classification based on these techniques should be superior to classification based on the ratings. Further, classification of the experts (i.e. IPs vs GPs) should be more difficult than any classification involving UPs, if these techniques have produced psychologically valid measures of conceptual structure.

Results and discussion

The results of the pattern classification analysis are shown in Fig. 5. These percentages are based on different numbers of classifications, depending on the number of members of each group. In general, there are $2 \times n_1 \times n_2$ individual classifications for each comparison. Thus, the number of possible correct classifications ranged from 306 for UP–GP down to 126 for IP–GP. Overall, classification of individuals into groups was better than chance for patterns derived using all three methods, especially in discriminating novices (UPs) from experts (IPs and GPs).

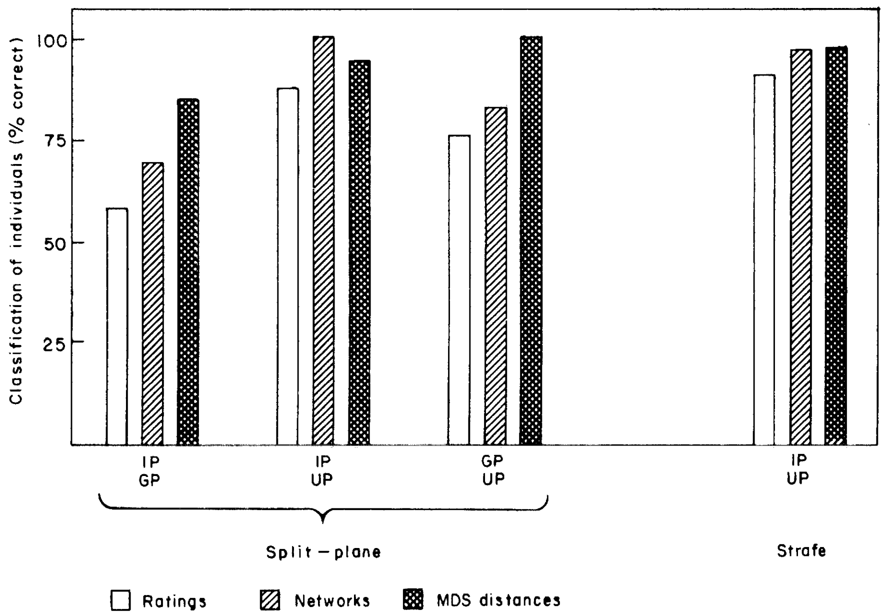


FIG. 5. Results of pattern classification procedure applied to each pair of pilot groups for concepts from both scenarios.

Classification was least successful using the original ratings and most successful using distances derived from MDS solutions. For each pair of groups, classification based on MDS distances was superior to classification based on the original ratings. MDS was very successful at revealing the underlying structure in the rating data. Classification accuracy using patterns derived from networks was also better than the original ratings in each case. Thus, networks also reveal structure that is not directly available in the original ratings.

Several conclusions follow from these results. First, pattern classification techniques can discriminate novices from experts with a high degree of success based upon both structural descriptions of their conceptual structure, and raw similarity ratings. Second, both MDS and Pathfinder showed superior classification compared to the empirical ratings, with MDS outperforming Pathfinder by 4% on expert-novice classifications and 16% on expert-expert classifications. Finally, all procedures found it more difficult to make expert-expert classifications based on conceptual structure than expert-novice classifications.

These classification results suggest that the structures imposed by MDS and networks on raw data capture some valid structural information about the differences in the way distinct groups of pilots organize conceptual information. In comparing MDS with network representations, it would appear that MDS captures somewhat more of the structural information that is useful in discriminating between groups of individuals. The MDS superiority was especially noticeable in discriminating between the groups of experts.

4.3. SUMMARY

In this section, an attempt was made to validate the conceptual structures represented by networks and MDS. Validation took two forms. First, differences among the pilots, especially between novices and experts, are revealed by these structures. Second, conceptual structures based on MDS or Pathfinder contained more useful information for classifying individuals than did the original ratings.

5. Applications

5.1. PREDICTION AND SELECTION

Prediction of pilot performance and selection of pilot trainees for job placement are two potential applications of this work. Knowledge concerning the differences between each individual pilot trainee and a group of expert pilots enables one to select the single trainee who most resembles the experts. Several techniques provide information about individual differences. The methods described in this section have each been used previously in this work to achieve other goals. For instance, pattern classification has been used as a means of validating various types of cognitive representations (networks, MDS). However, this technique can also be used to examine the similarities or differences of an individual in relation to his group or to other groups. The assumption that individuals and groups differ in their cognitive organization underlies each of these techniques.

The individual difference analysis is exemplified with the split-plane scenario. In each case, an attempt was made to rank the UPs in relation to the experts. Following a discussion of each technique, the different rankings of the UPs generated from the three techniques are compared.

Pattern classification

Pattern classification techniques are useful in reflecting how similar an individual is to his group and how similar an individual is to some other group. Two measures are of particular interest, the distance from an individual to the other group's prototype (how like the other group is the individual) and the distance from an individual to the belonging group prototype point (how like the group is the individual). The prototype point is that point representing the central tendency of a class of patterns. The distance between an individual and a class prototype reveals how strongly that individual represents the average features of that class. Identical ranks for individual UPs were obtained using MDS and network patterns.

Table 3 gives the distances from an individual to the group prototype for experts (IPs and GPs) and the distance from each UP to the group prototype for novices

Individual difference data for split
the UPs. Low rank

	To expert prototype
IP_1	110
IP_2	145
IP_3	106
IP_4	130
IP_5	122
IP_6	142
IP_7	103
GP_1	125
GP_2	133
GP_3	111
GP_4	132
GP_5	129
GP_6	140
GP_7	122
GP_8	173
GP_9	116
UP_1	169 (6)
UP_2	180 (14)
UP_3	166 (5)
UP_4	193 (17)
UP_5	189 (16)
UP_6	159 (2)
UP_7	166 (4)
UP_8	173 (10)
UP_9	162 (3)
UP_{10}	174 (11)
UP_{11}	170 (7)
UP_{12}	188 (15)
UP_{13}	171 (8)
UP_{14}	172 (9)
UP_{15}	177 (12)
UP_{16}	155 (1)
UP_{17}	179 (13)

* Prototypical UP.

(UPs). All individuals were clas
to each UP score is a rank inc
with a rank of 1 indicating the
This information can be use
their own group or most like

TABLE 3

Individual difference data for split-plane concepts (numbers in parentheses are ranks for the UPs. Low ranks, 1, suggest high similarity to experts)

	Distances			
	To expert prototype		From INDSCAL origin	Along person dimension
IP_1	110		530	208
IP_2	145		437	190
IP_3	106		592	214
IP_4	130		545	207
IP_5	122		595	208
IP_6	142		458	170
IP_7	103		631	227
GP_1	125		458	224
GP_2	133		481	167
GP_3	111		575	209
GP_4	132		515	241
GP_5	129		448	238
GP_6	140		429	189
GP_7	122		525	196
GP_8	173		354	0
GP_9	116		565	166
		To novice prototype		
UP_1	169 (6)	132	367 (8)	346 (6)
UP_2	180 (14)	149	268 (16)	389 (13)
UP_3	166 (5)	144	482 (1)	263 (3)
UP_4	193 (17)	162	332 (10)	473 (17)
UP_5	189 (16)	149	267 (17)	441 (15)
UP_6	159 (2)	141	353 (9)	258 (1)
UP_7	166 (4)	148	471 (2)	318 (5)
UP_8	173 (10)	145	323 (11)	285 (4)
UP_9	162 (3)	122*	383 (5)	327 (6)
UP_{10}	174 (11)	147	310 (12)	353 (10)
UP_{11}	170 (7)	135	461 (3)	347 (8)
UP_{12}	188 (15)	159	298 (13)	447 (16)
UP_{13}	171 (8)	128	372 (7)	349 (9)
UP_{14}	172 (9)	131	447 (4)	356 (11)
UP_{15}	177 (12)	156	284 (14)	379 (12)
UP_{16}	155 (1)	145	380 (6)	259 (2)
UP_{17}	179 (13)	144	276 (15)	401 (14)

* Prototypical UP.

(UPs). All individuals were classified on the correct side of the decision surface. Next to each UP score is a rank indicating how close the UP is to the expert prototype, with a rank of 1 indicating the greatest similarity to the experts.

This information can be used to select individuals who are most like members of their own group or most like members of another group. For instance, UP16 is the

UP closest to the expert prototype (rank of 1). This suggests that the conceptual structure of UP16 is most similar to that of the experts. Thus, one might predict that UP16 would perform more like an expert on a particular flight-related task than would other UPs. On the other hand, UP4 is furthest from the expert prototype, suggesting a larger distinction between this individual and experts than between UP16 and experts. Thus, these measures provide a means of detecting within and between group differences and, consequently, are prediction and selection aids.

UP9 is the closest to the prototype of the undergraduates. In fact, if pattern classification is conducted beween experts and UPs, using only UP9 as a single "known" UP, classification is successful for 91% of the cases. Having this information about members of a class could have pedagogical value. Instructors are often concerned about the level at which to aim a lecture. One possibility is to make certain the prototypical student has understood the material.

Individual differences scaling

Another method of ordering individuals is the INDSCAL MDS procedure. Earlier, this technique was discussed to show that the pilot groups were separable based on their cognitive structures. It is also possible to determine how each individual pilot weights the dimensions and then to compare individuals in the resultant space. INDSCAL locates individuals along the same dimensions on which concepts are located. Thus, this scaling procedure reveals how much a particular individual relies on a particular dimension. Earlier in this report, it was shown that UPs as a group tended to weight the dimensions less than did experts. Here, the extent to which each individual UP considers the dimensions can be determined. Individuals can be ranked by their distance from the origin (0, 0, 0). For example, a UP that does not weight any of the dimensions would be 0 units from the origin and quite unlike any expert (who relies on the dimensions). A UP that weights the dimensions heavily has in some sense a conceptual structure more like that of an expert.

Table 3 (column 2) gives, for each individual, the distance from the origin of the three-dimensional MDS solution reported earlier. UP3 is the furthest from the origin of all the UPs, suggesting that this individual relied heavily on the same dimensions as did the experts, in making similarity judgments. In contrast, UP5 is very near the origin, suggesting that this UP made very little use of the expert dimensions. As with the pattern classification discussed previously, the INDSCAL procedure allows comparisons of UPs with experts. It has the advantage of restricting comparison to the same conceptual space for experts and novices, but does not supply the information about the prototypical student that the pattern classification technique yields.

Person space

The final individual difference analysis to be considered is a hybrid of the previous two techniques. It is an attempt to represent the individuals in a multidimensional space, but within a space that has dimensions relevant to the subjects, not to the concepts. Thus, the plan is to position subjects in a space where the dimensions reflect differences among the subjects. If this is successful, one of the dimensions should reflect expertise.

To represent individuals in multidimensional space, an intersubject distance matrix was derived, similar to the interitem distance matrix derived for the concepts. Distances for this matrix were derived from distances from each individual three-dimensional MDS solution. In this case, the individual can be thought of as a point in "n-dimensional" space ($n = 435$ dimensions based on the 435 distances for all pairs of 30 terms). The distance between two individuals would take into account the difference in distance for each of the 435 pairs of points for the two individuals. These distances resulted in a matrix of distances with individuals as rows and columns. The entries in this matrix were simply the distance from one individual to another. These distance values were then scaled in multidimensional space using one and two dimensions.

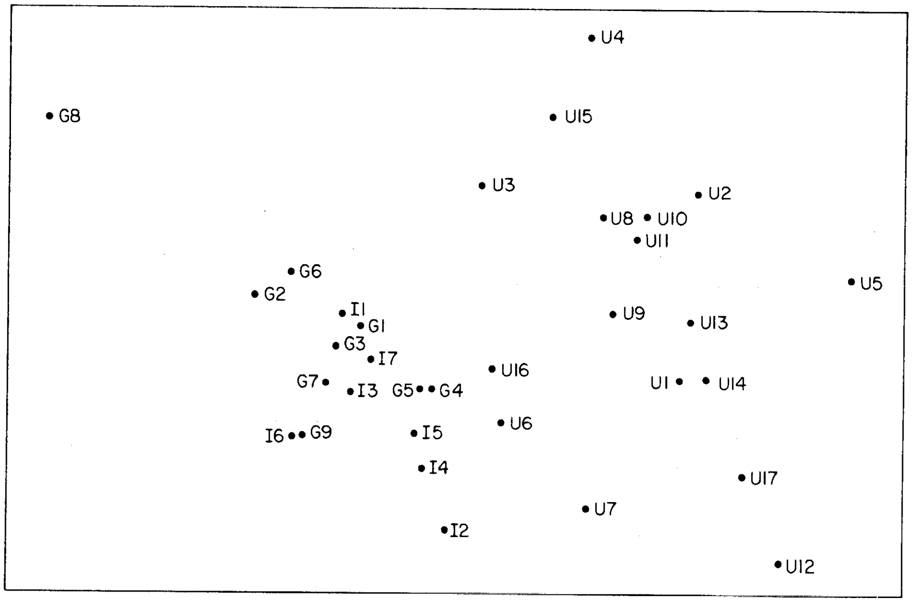


FIG. 6. Two dimensional person space showing the location of each IP (I), each GP (G), and each UP (U).

Figure 6 shows the two-dimensional MDS solution. The experts and UPs are clearly linearly separable. One of the dimensions can be readily identified as a dimension of expertise, suggesting the technique has been successful at establishing a multi-dimensional space with dimensions that characterize the subjects, rather than the concepts. The second dimension is suggestive of a pedagogical-operational dimension, with the IPs and GPs tending to occupy different ends of that dimension.

A one-dimensional MDS solution should extract the single dimension that accounts for the most variance. The values along that dimension were transformed to set the first pilot at a co-ordinate of 0; these values appear in Table 3. Comparison of these values indicates the relative distance from one individual to another. Thus, it seems that the one-dimensional solution ordered individuals along an expertise dimension. The creation of the "person space" clearly helps to define the separate groups of subjects and the locations of individuals in relation to the groups.

Again, UPs that are close to the expert end of the continuum should have organizations of flight-related information similar to those of experts. For this technique, UP6 appears closest to the experts, with UP4 being most distant. In general, representation of individual pilots in multidimensional space provides a measure of distance that can be used in prediction and selection.

Comparison of techniques

Each of the three techniques just discussed provides information that orders UPs in relation to the experts. The techniques point to a different "best" student. However, inspection of Table 3 also suggests that there is substantial agreement among the three techniques in their ordering of the UPs. All agree, for example, that UP3 is superior to UP4. Spearman correlations were performed on the ranks of the UPs in order to more rigorously compare the techniques. The intercorrelation matrix appears in Table 4. All correlations were significant. As can be seen, there is a good deal of agreement

TABLE 4
Matrix of Spearman correlations for three individual difference measures

	Person space	INDSCAL	Pattern classification
Person space	1·00	0·63	0·92
INDSCAL	0·63	1·00	0·74
Pattern classification	0·92	0·74	1·00

among the techniques. Pattern classification, INDSCAL, and the person space developed here supply converging validation that may be useful in the selection of students. Further, the techniques supply additional, different details about the populations that may further facilitate selection or aid in training.

5.2. TRAINING

In addition to implications for personnel and selection, an understanding of the cognitive structures of experts and novices should have implications for training. Although the underlying structures of some students are closer than are others to those of experts, it should be possible to facilitate acquisition of the expert structure for the students in general. Accomplishing this requires that it first be determined which concepts are not well understood by the UPs relative to expert fighter pilots. This requires that the information critical to expertise be determined. This is not a trivial problem because any individual expert will tend to have associations that, while perhaps useful, are not necessary for expertise. Thus, the problem is to determine which associations in an expert knowledge base are necessary or essential for expertise and which are not. The scaling procedures considered in this report may help to select the critical associations.

We defined critical information as those components of the cognitive structures that tend to be present in all experts. Any information in the knowledge structure of one group of experts, but not another, cannot be prerequisite component of expertise. After the information critical to expertise is established, one can compare the UPs to the experts, and isolate the concepts that have been mastered by the students. In addition, those concepts which are the most disparate from those of the experts can be determined and thus provide some information concerning which deficits should be addressed first in any pedagogical intervention.

The right stuff

The split-plane data of the two groups of experts, IPs and GPs, were considered first. Defining expertise based on these two groups has several advantages. Most importantly, comparing two relatively different groups that are both expert reduces the likelihood that idiosyncratic components of the cognitive structures will manifest themselves as critical components of expertise. As we showed earlier, IPs could be distinguished from GPs; although, of course, the distinction was not as great as between UPs and either of our expert groups. The information common to IPs and GPs should be the minimal structure necessary for expertise.

The Pathfinder analyses were used to determine those features of networks that tended to be characteristic of expertise. The networks for IPs and for GPs considered earlier were compared and any common links were extracted. The resulting network consisted of one major network, three isolated concept pairs, and five isolated single concepts. Thus, IPs and GPs agreed on a way to interconnect 19 of the split-plane concepts and agreed on an additional three pairwise associations; IPs and GPs did not agree on any particular link for five concepts. Then the isolated pairs were linked to the main network by allowing either member of the pair to connect to a particular concept in the main network. For example, IPs and GPs agreed that AIRSPEED is related to the pair ACCELERATION-EXTENSION; however, IPs had AIRSPEED and ACCELERATION linked, whereas GPs linked AIRSPEED with EXTENSION. A similar procedure linked the isolated single concepts to the main network. The resulting structure of "expertise" is shown in Fig. 7.

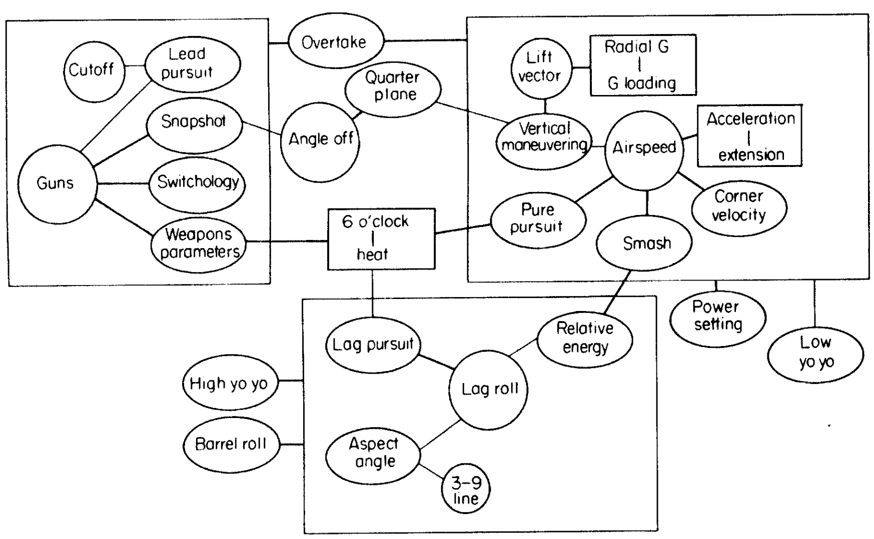


FIG. 7. Network overlap for IPs and GPs. Bold links are also found in the UP network.

Novices and experts

Next, the network for UPs was compared with the expert structure. A bold line in Fig. 7 represents a link present in the UP network. As can be seen, a number of critical links are also held by UPs. These links centre around the concept of airspeed and, to a lesser extent, the concept of guns. Several links are not present in the UP network. For example, the links near the concept lag roll are almost totally absent from the UP network.

In order to quantify UPs' understanding of the concepts, each of the 30 split-plane concepts was considered individually. For each concept, students might differ from experts in two general ways. Students may not have some of the critical associations that experts have (as illustrated in Fig. 7). Alternatively, students may have associations between concepts that neither group of experts has (e.g. the network of UPs has a link between HIGH YO YO and LOW YO YO, although for any expert these concepts are relatively unrelated). These two general dimensions can combine to produce four different types of concepts: (a) a concept can be well-defined in that the critical links are present and the student does not see many spurious additional relations; (b) a concept can be overdefined in that the critical links are present, but the student also has a number of inappropriate links; (c) a concept can be underdefined in that the critical links are absent as are idiosyncratic links; and (d) a concept can be misdefined in that the critical links are absent and the student has many idiosyncratic, nonexpert connections.

The following values were computed for each concept: (a) the proportion of critical links found in the UP network, and (b) the proportion of the UP links that were found only in the UP networks (i.e. the "extra" links that occurred in neither the IP nor the GP network). A median split along each variable led to the classification of concepts appearing in Table 5. The well-defined concepts tend to be those involved in flying the aircraft, with some consideration of the aircraft's tactical functions, but no terms showing an understanding of split-plane maneuvers or an understanding of air-to-air combat scenarios. Interestingly, these well-defined concepts probably evoke a reasonably accurate meaning from a reader naive of tactical flight procedures.

Students' judgments of familiarity are reasonably consistent with the classification in Table 5. Well-defined concepts are the most familiar with misdefined concepts the least familiar. However, familiarity does not always agree with the Pathfinder analyses. For example, some underdefined concepts (CUTOFF and VERTICAL MANEUVERING) and some misdefined concepts (ACCELERATION, 6 O'CLOCK, RELATIVE ENERGY) are judged as very familiar. For each of these "familiar" concepts, it seems the student is not aware of the true scope of the concept. The student seems to have an understanding of the concept in a narrower sense than does the expert. Inspection of Fig. 7 reveals that each of these concepts has a critical connection in the expert structure to a concept that the students have not experienced. For example, whereas students are familiar with ACCELERATION and, in fact, have a global understanding of it, they are missing the critical connection with EXTENSION, a concept with which they have had little or no experience. Thus, UPs have an understanding of ACCELERATION in the same sense that a psychology undergraduate, who knows nothing of analysis of variance, might have an understanding of Student's *t*.

If UPs do have much of the cognitive structure of the experts for the well-defined concepts, and little of the expert structure for the misdefined concepts, then classification based on these subsets of concepts should reflect the difference in understanding. Classification of UPs and experts based only on the well-defined concepts should be relatively poor because there would be little information in these

Table 5
Thirty split-plane concepts grouped by UP understanding

	Familiarity
Well-defined concepts	
AIRSPEED	3·00
OVERTAKE	3·00
G-LOADING	2·94
SMASH	2·81
LIFT VECTOR	2·75
CORNER VELOCITY	2·38
GUNS	2·00
SWITCHOLOGY	1·81
Mean	2·59
Underdefined concepts	
CUTOFF	3·00
VERTICAL MANEUVERING	2·88
ANGLE-OFF	2·00
WEAPONS PARAMETERS	1·75
PURE PURSUIT	1·56
LEAD PURSUIT	1·50
ASPECT ANGLE	1·19
QUARTER PLANE	1·13
Mean	1·88
Over-defined concepts	
POWER SETTING	3·00
BARREL ROLL	2·94
HI YO YO	1·50
RADIAL G	1·31
Mean	2·19
Misdefined concepts	
ACCELERATION	2·94
6 O'CLOCK	2·81
RELATIVE ENERGY	2·75
LAG PURSUIT	1·50
LOW YO YO	1·38
3–9 LINE	1·25
SNAPSHOT	1·19
LAG ROLL	1·19
HEAT	1·19
EXTENSION	1·19
Mean	1·74

concepts that would allow classification. On the other hand, classification based on the misdefined concepts should be relatively successful because the UPs differ a great deal from the experts.

Pattern classification of UPs and experts was performed using each of the four subsets of concepts. Correct classifications were 82%, 85%, 85% and 100% for well-defined, overdefined, underdefined and misdefined sets of concepts, respectively. Again this classification is particularly easy, even when using only the minimum distance classification (i.e. the initial step in classification) and only a subset of the concepts. However, consistent with expectations, there were differences at this high level of success: the well-defined concepts yielded poorest classification and misdefined concepts yielded the best classification. In fact, with only the subset of misdefined concepts the classification was perfect. In addition to supporting Table 5, these results have the practical advantage of allowing classification with only a small portion of the total data set, thus reducing the amount of data that needs to be collected and the run time of the algorithm.

The information provided by our analysis of individual concepts could be of use in organizing a training curriculum. The analysis shows which concepts should be stressed and which seem to be well understood. In short, these methods give detailed information about the the structure of knowledge in addition to an indication of what students need to learn.

5.3. EXTRACTING EXPERTISE

With the advent of expert systems technology, the problem of obtaining expert knowledge in a form appropriate for coding into assertions and rules has become a primary concern (Hayes-Roth, Waterman & Lenat, 1983). The techniques we have discussed in this paper may prove to be useful in the effort to develop more formal procedures for obtaining and representing expert knowledge. Gammack and Young (1985) have discussed possible uses of multidimensional scaling in defining the relations among key concepts. The networks produced by Pathfinder may also be of use to the knowledge engineer in determining which pairs of concepts need further analysis to determine specific relations.

As an illustration, we presented the links obtained in the IP networks for both the split-plane and strafe concept sets to an expert fighter pilot. The expert identified seven types of links in each set of concepts. For the split-plane concepts, the types of links were: AFFECTS (15 links), IS A (11 links), LEADS TO (five links), DESIRABLE (four links), ACCEPTABLE (two links), SELECTS (two links), and INSTRUMENT OF (one link). The link types in the strafe concepts were: AFFECTS (12 links), DETERMINES (11 links), IS POINT OF REFERENCE FOR (five links), DESIRABLE (five links), IS (three links), AVOIDS (two links), and INSTRUMENT OF (one link).

The network solution reduced the 870 pairs (435 in each set) to 79 links (40 in split-plane and 39 in strafe) for detailed analysis by our expert. Evaluating the 79 linked pairs was relatively simple compared with working with all of the original pairs. The use of the network scaling technique along with the identification of the type of each link results in seven types of relations for each set of concepts or a total of about ten distinct relations. This reduction could be a useful step in the process of organizing and codifying a large number of potential relations. The Pathfinder scaling algorithm can help pinpoint the important pairs of concepts for additional analysis.

5.4. SUMMARY

Possible applications of the scaling procedures to selection and training have been suggested. Information about conceptual structure might be profitably applied to aid in decisions about assignments to fighter aircraft. Three techniques developed to assess individual differences showed considerable agreement in their suggestions of which undergraduate pilots should be directed to lead-in fighter training. In addition, the attempts here to apply information about conceptual structure to training suggested particular points in the UPs' understanding that could benefit from intervention. The particular weak points in the knowledge structure were suggested by Pathfinder and MDS, sometimes in the face of the students' self-perception of their familiarity with the concepts. The subset of misdefined concepts contained enough information about the pilots to classify them perfectly.

The success of the methods at discriminating among pilots of varying expertise based on measures of conceptual structure suggests that scaling methods may provide some empirical techniques for measuring the structure of expertise. These techniques should have application in training and selection as well as in artificial intelligence systems that attempt to represent knowledge structures. The application of these techniques in obtaining and representing expert knowledge is a promising direction for future work.

This research was supported by contract F33615-80-C-0004 from the Air Force Human Resources Laboratory awarded to R.W.S. while F.T.D. was a visiting professor.

References

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

CARROLL, J. D. & CHANG, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometric*, **35**, 283–319.

CHASE, W. G. & SIMON, H. A. (1973). Perception in chess. *Cognitive Psychology*, **4**, 55–81.

CHI, M. T. H., FELTOVICH, P. J. & GLASER, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121–152.

CHRISTOFIDES, N. (1975). *Graph Theory: An Algorithm Approach*. New York: Academic Press.

COLLINS, A. M. & LOFTUS, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407–428.

COLLINS, A. M., & QUILLIAN, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 240–247.

COOKE, N. M. (1983). Memory structures of expert and novice computer programmers: Recall order vs. similarity ratings. *Unpublished Masters Thesis*, New Mexico State Univeristy.

DURSO, F. T., SCHVANEVELDT, R. W. & GOLDSMITH, T. E. (1983). *Empirical networks of natural categories*. Paper presented at the meeting of the Southwestern Psychological Association, San Antonio.

ENGLE, R. W. & BUKSTEL, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology*, **91**, 673–689.

FILLENBAUM, S., & RAPAPORT, A. (1971). *Structures in the Subjective Lexicon*. New York: Academic Press.

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. & YOUNG, R. M. (1985). Psychological techniques for eliciting expert knowledge. In BRAMER, M. A., Ed., *Research and Development in Expert Systems*, Cambridge University Press, pp. 105–112.

HAYES-ROTH, F., WATERMAN, D. A. & LENAT, D. B., Eds. (1983). *Building Expert Systems*. Reading, Mass: Addison-Wesley.

HUTCHINSON, J. W. (1981). Network representations of psychological relations. *Unpublished Doctoral dissertation*, Stanford University.

ISAAC, P. D. & POOR, D. S. (1974). On the determination of appropriate dimensionality in data with error. *Psychometrica*, **39**, 91–109.

KRUSKAL, J. B. (1977). Multidimensional scaling and other methods for discovering structure. In ENSLEIN, RALSTON and WILF, Eds., *Statistical Methods for Digital Computers.* New York: Wiley.

MCKEITHEN, K. B., REITMAN, J. S., RUETER, H. H. & HIRTLE, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, **13**, 307-325.

MEYER, D. E. & SCHVANEVELDT, R. W. (1976). Meaning, memory structure, and mental processes. *Science*, **192**, 27–33.

MEYER, R. P., LAVESON, J. I., PAPE, G. L. & EDWARDS, B. J. (1978). *Development and Application of a Task Taxonomy for Tactical Flying.* AFHRL-TR-78-42(III), AD-A061 478. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

NILSSON, N. (1965). *Learning Machines*: *Foundations of Trainable Pattern-Classification Systems.* New York: McGraw-Hill.

QUILLIAN, M. R. (1969). The teachable language comprehender. *Communications of the Association for Computing Machinery*, **12**, 459–475.

REITMAN, J. S. (1976). Skilled perception in Go: deducing memory structures from interresponse times. *Cognitive Psychology*, **8**, 336-356.

RIPS, L. J., SHOBEN, E. J. & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, **12**, 1–20.

SCHANEVELDT, R. W. & DURSO, F. T. (1981). *Generalized semantic networks.* Paper presented at the meeting of the Psychonomic Society, Philadelphia.

SCHVANEVELDT, R. W., DURSO, F. T. & DEARHOLT, D. W. (1985). *Pathfinder*: *Scaling with network structures.* Memorandum in Computer and Cognitive Science, MCCS-85-9, Computing Research Laboratory, New Mexico State University.

SHEPARD, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, **5**, 33–48.

SHEPARD, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrica*, **39**, 373–421.

SMITH, E. E., SHOBEN, E. J. & RIPS, L. J. (1974). Structure and process in semantic memory: a featural model of semantic decisions. *Psychological Review*, **81**, 214–241.

TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-351.