Chapter 15

Expert Conceptual Structure: The Stability of Pathfinder Representations*

John G. Gammack

This chapter describes the use of Pathfinder in modeling an expert's conceptual structure. The aim is to produce networks that represent domain knowledge underlying and informing expert classification decisions. To this end, several elicitation tasks are described, each of which produces a network relating domain concepts. These tasks have been developed in psychology and include relatedness estimates, repertory grid, and recall tasks. The techniques address distinctions among British steam locomotives, exemplifying the classificatory aspect of domain conception.

Practical application of these techniques in knowledge elicitation are described elsewhere (Gammack, 1987a, 1987b). Here the concern is with the role of Pathfinder within the approach and the stability of its representations under different elicitation conditions.

The elicitation techniques used in this study attempt to observe the results of information-based decisions without determining or otherwise influencing these results. Since only the results of cognition can be observed in this way, any conceptual structure in memory presumed to determine this may only be indirectly inferred. Such psychological tasks have traditionally been used to reveal pre-existing memory structures, on the assumption that they neither originate nor misrepresent that structure. Factual information is assumed largely to inhere in conceptual structures that are stored in memory and that reflect the actual relations of domain objects. Since each elicitation task provides a relationship network, these may be compared. By overlapping the (network) structure revealed by different techniques addressing the same stored information, it is assumed that spurious relationships due to experimental noise or error may be identified, and thus a core organization of stable relationships established.

Although the dictates of actual circumstance may modify the applicability of a given set of relationships, this constraint is minimized in an experimental laboratory setting. By abstracting the elicitation from any pragmatic context, any pre-existing structure in the mind is more likely to emerge uncompromised by workplace influences and expedient considerations.

These assumptions imply that a stable organization of related domain concepts will emerge. Two aspects of stability are considered here. The first concerns the stability of the Pathfinder representation of conceptual relationships under different elicitation methods. The second considers the stability of information in memory and of formal representations of its structure.

---

*I would like to thank Roger Schvaneveldt, Nancy Cooke, and Richard Young for help during preparation of this chapter.

The Elicitation Tasks

Five psychological tasks provided data for analysis by Pathfinder. The proximity matrices were also analyzed using cluster analysis (single-linkage, nearest neighbor) and multidimensional scaling (MDS), which respectively indicated prominent groupings and dimensions of conceptual variation. The five tasks are described more fully in Gammack (1987a, 1987b) but are introduced briefly here. Using the dominance metric ($r = \infty$) and omitting weights, the minimal Pathfinder graphs from each task are illustrated, and in the next section these will be formally compared.

Proximity Ratings

This task elicited estimates of the relatedness between pairs of locomotives to directly provide a proximity matrix. One locomotive was presented as a target and its relatedness to each of the others was considered in turn, thus providing the proximity estimates. Notwithstanding Tversky (1977), asymmetries were not considered to have psychological significance here, and targets were not replaced. To make the procedure easier, the expert presorted the remaining photographs into three piles according to degree of relatedness (strong, medium, or weak), and then assigned numeric values within each range in turn. This procedure was repeated until each locomotive had been used as a target, and the resulting half-matrix produced the graph shown in Figure 1.

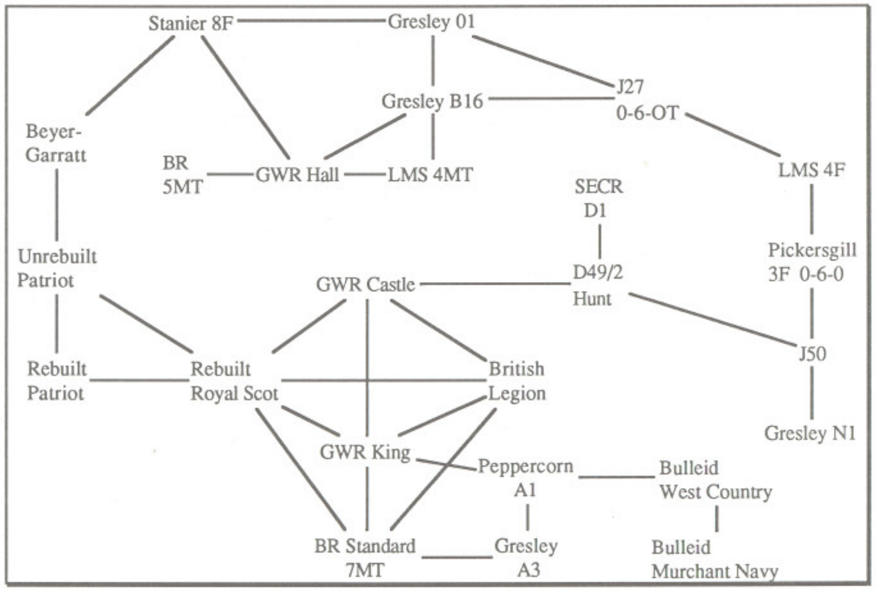

Figure 1. Pathfinder network from the proximity ratings.

Much domain information is implicit in the network, which reasons of space preclude detailing here; however, it is superficially clear that (for example) the two Bulleid engines are strongly related, as are the three Great Western Railway (GWR) engines. The expert later described the grounds for each link, and the detailed domain knowledge thereby emerging justified the relationships shown.

Repertory Grid

Relatedness estimates are not absolute values and are subject to contextual variation. Two locomotives highly related by geographical region might be unrelated in their design features. Such context will affect whether two locomotives are considered to be highly related or not.

In the absence of any context (other than the target locomotive itself) to provide criteria for relatedness ratings, it is probable that relatedness decisions will have been determined by consideration of the characteristic properties of engines, such as their age, size, and duty, and an overall estimate made on that informed basis.

To test this, and to establish the underlying information being used, the repertory grid technique was employed. This involved taking three locomotives at random from the set, and asking in what way two were alike and thereby different from the third. For example, two locomotives might be primarily involved in freight duty, but the third was rather more a passenger engine. The dimension (or construct) of freight/passenger thus elicited provided a bipolar scale along which the rest of the sample was rated from 1 (purely freight) to 7 (purely passenger). Repeating this procedure elicited a set of constructs, such as freight/passenger and older/modern, and provided a rectangular matrix profiling each engine in terms of these constructs. The city block distance through the multidimensional space implied by this grid was calculated for each pair to provide a full proximity matrix. This is illustrated in Figure 2, superimposed on the MDS plot and cluster analysis of the same data. Links in common with the previous solution are shown as bold lines, comprising about half the minimal network.

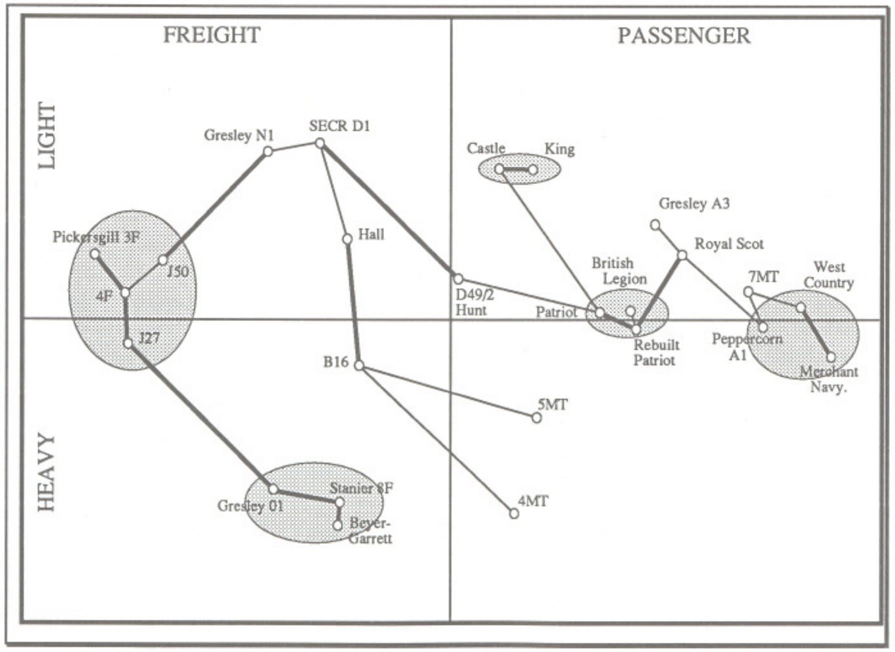

Figure 2. The analysis of the repertory grid data showing a Pathfinder network on an MDS layout with clusters of nodes.

Seeded Recall Task

This task follows the method of Reitman and Reuter (1980). In it the expert is given the name of one locomotive as a cue and is required to recall the rest of the set. This is repeated using the other items as cues (or seeds) and the output order is noted. The resulting data may then be converted to a proximity matrix by calculating the overall distance between each pair of items. The rationale behind the technique is that related items will be recalled consecutively, so that if two items are recalled consecutively on each of 25 trials, they will have a (very close) proximity of 25 units. The analyses are shown in Figure 3.

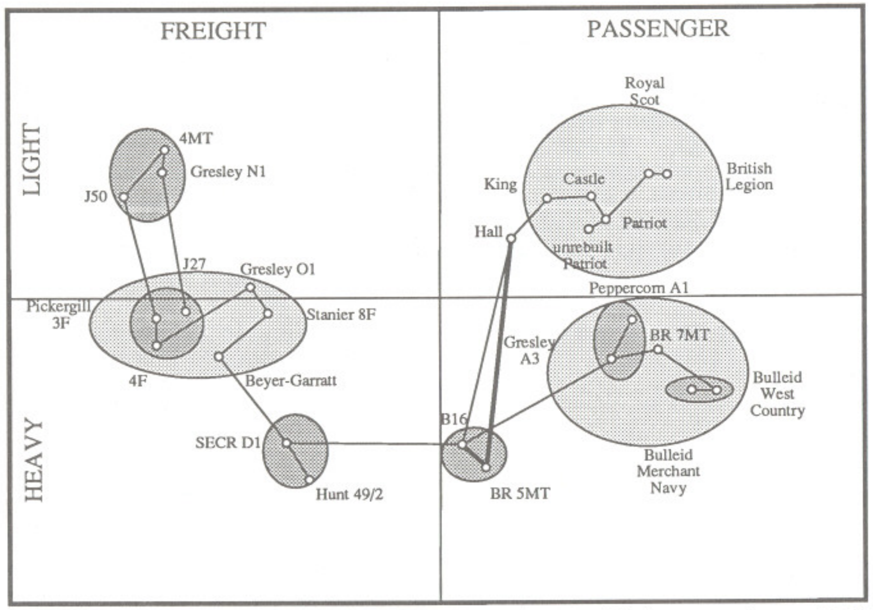

Figure 3. The analysis of the seeded-recall data showing a Pathfinder network on an MDS layout with clusters of nodes.

Bipolar Rating Scales

This is a simple consistency check on information given in the repertory grid and implicit in the MDS dimensions. Following a method reminiscent of the semantic differential (Osgood, Suci, & Tannenbaum, 1957), seven bipolar scales (e.g., freight/ passenger) were constructed. The expert marked the appropriate points on these for each locomotive, and this information was quantized as seven-point scales. These were again converted to a proximity matrix, using the same procedure as for the repertory grid, namely by taking the city block distance in multidimensional space. The analyses are shown in Figure 4.

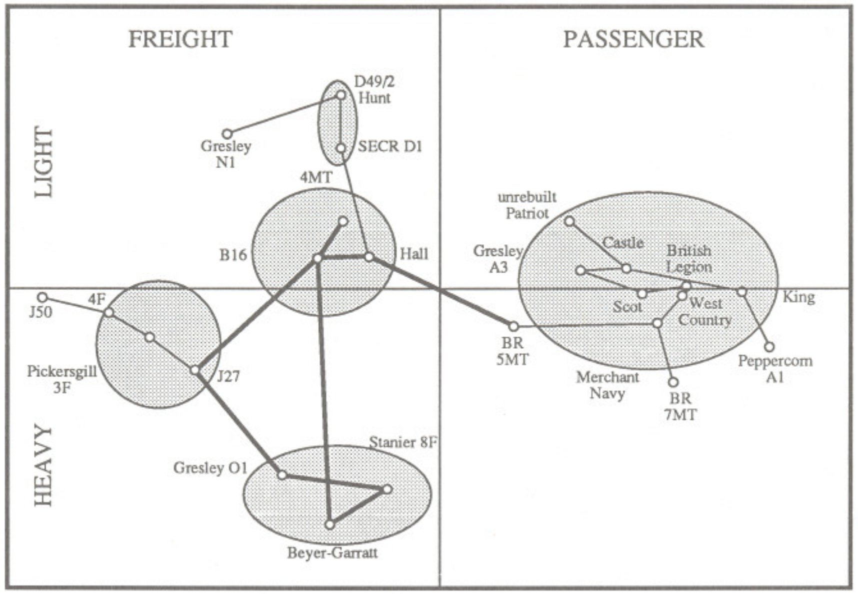

Figure 4. The analysis of the bipolar scale data showing a Pathfinder network on an MDS layout with clusters of nodes.

Twenty Questions

In this task, adapted from a parlor game, the experimenter chose one locomotive as a target, which the expert had to guess by asking questions phrased to be answered yes or no. For example, to the question "Are you thinking of a freight engine?" a "no" response would eliminate a whole class of items. In addition to being specifically informative, the elimination patterns were recorded. This allowed the similarity of locomotives across targets to be gauged to provide another proximity matrix. Using the rationale that locomotives with similar profiles would tend to be eliminated by the same criteria, this proximity matrix was calculated in two stages. First, a rectangular matrix was constructed, showing for each target item the question number at which each locomotive was eliminated. After this was done the city block distances for all pairs were taken as before, and the subsequent analyses are shown in Figure 5.

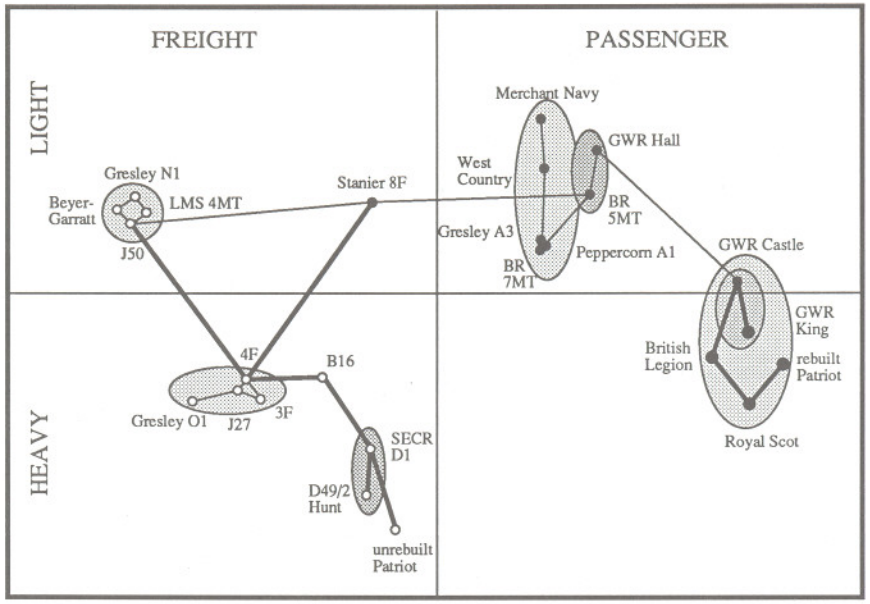

Figure 5. The analysis of the Twenty-questions data showing a Pathfinder network on an MDS layout with clusters of nodes.

Summary

These five tasks have provided five different networks showing relationships (in particular, similarity) among locomotives. Any or all of these may be (and have been) used to elicit the specific domain information they imply and concisely represent. Each network approximately fits the same two-dimensional space (e.g., Alscal's simple euclidean model in two dimensions explained virtually all the variance with little stress), but only a minority of specific relationships are preserved across tasks, and visually the networks differ considerably. In the next section some measures of the agreement between networks are introduced, prior to a more general discussion.

Analysis of Agreement Among Tasks

The previous section introduced the elicitation tasks and presented the minimal network for each. In this section some methods for comparing these networks are described.

Correlation

At first this method was not applied directly to the networks but to the original data matrices. Comparing these matrix pairs point for point using nonparametric statistics, such as Spearman's ρ or Kendall's τ, rarely produced correlation coefficients of greater than 0.2. Kendall's τ is appropriate to the comparison of interest since it attends to disarray in rank-ordered data. This test is as powerful as Spearman's ρ, but its conservative nature often

leads to lower values. Because of the high number of points involved, however, even such low values were statistically significant, suggesting that at least the datasets are related.

However, Schvaneveldt (personal communication) points out that a straightforward relationship between raw proximities and network distances is not captured by linear correlations of the raw data. Since noisy raw data may produce spuriously low correlations, it may be more meaningful to compare the minimal networks directly by correlating matrices reconstructed from the minimal network. This information is implicitly available from the Pathfinder program, although an extra procedure must be written to have it output. The distances through the minimal network can be computed for cells which would otherwise be empty in Pathfinder's output matrix. The full matrix, despite numerous ties, may then be compared with its counterpart from another elicitation task. Although to date the mathematical properties of this reconstructed matrix remain unexplored, the correlations were again disappointingly low for these datasets, appearing rather sensitive to properties of the raw data. For instance, an unusually extreme value would propagate through many paths to profoundly affect the overall pattern of rank orderings and lower the amount of correlation. Cooke (personal communication) suggests that the number of ties can typically be reduced by summing the weights along the shortest paths, rather than using the dominant link weight in each path. If psychologically justified, this procedure may be more appropriate.

Correlations were also performed on matrices of graph-theoretic distances taken from the illustrated minimal graphs. For each pair of locomotives, the number of links in the shortest path connecting them was computed and entered in a separate matrix for each of the five tasks. The matrices thus derived from each task were then correlated in pairs, comparing first the proximity ratings with each of the other four tasks and then the repertory grid with the other tasks. These compared favorably with the previous correlations; for instance, the correlation between the graph-theoretic distances for the proximity ratings and the repertory grid tasks was $\rho = 0.59$. The other correlations averaged about 0.3, and all correlations were significant or highly significant. This use of correlation is probably the most satisfactory one.

Hypergeometric Test

This simple statistic indicates the significance of the overlap between two networks. It views the networks as samples drawn from a known population, in this case the 300 possible pairwise links among 25 objects. The minimal network represents (typically) 24 of these links. Given a number of links common to two networks, the probability of this overlap due to chance alone can be calculated from the hypergeometric distribution. For example, the 12 links shown in bold in Figure 2 show the overlap between minimal networks from the proximity ratings and repertory grid tasks. This highly significant ($p < 0.001$) overlap is extremely unlikely to be due to chance, indicating the relatedness of the two solutions.

Two things, however, should be noted here. Firstly, links common to both tasks account for only about half the total number of links, so highly significant overlap can be achieved between rather different looking networks. Secondly, the hypergeometric test indicates significance given a relatively small number of common links. For instance, a mere 6 links in common between two 24-link networks is significant at $p < 0.01$. This means that the test will successfully indicate the presence of nonrandom overlap, but is not a strong indicator of agreement between networks.

It seems reasonable to expect that even minimal graphs over the same domain objects will overlap to a high degree. The significance of the statistic can often be increased by using lower values of the q-parameter, for example, using $q = n-1$ is conservative. However, in most of the networks illustrated here there was little difference between the number of links for different values of q, and neither the statistic nor the conclusions were significantly affected by adding more links.

Chi-squared Test of Association

The next test allowed different assumptions about the sensitivity of the data to be considered, and used the chi-squared statistic as a test of association between two datasets (in effect a sort of McNemar test). This test used information about the link weights rather than mere connectivity, but without being overly sensitive to properties of the raw data.

Each task has provided a matrix of proximity values for the relative relationship strength between each pair of locomotives. These ordinal data can be compared with their counterparts from other tasks by binning them into ranges, then noting the frequency of matches in a contingency table, as shown in Table 1.

Table 1. Frequency of matches for two elicitation tasks.

	Repertory Grid	
Proximity Ratings	*Strong*	*Weak*
Strong	129	18
Weak	39	114

Thus in the simplest case of two ranges, all cells with proximities below 50 (or the median) are designated "strong," and all those above 50 (or the median) "weak," and two such rescored binary matrices are then directly compared. The frequencies are noted in a 2 × 2 contingency table, showing the number of times a relationship was "strong" in both tasks, "weak" in both, and the number of mismatches in one direction or the other. The more consistently relationship strength is preserved across tasks, the larger the frequencies on the main diagonal, and the fewer the mismatches. Applying the chi-squared statistic indicates the significance of this consistency across tasks. In this case there was a highly significant agreement between the directly rated proximities and those derived from the repertory grid ($\chi^2 = 117.96$, $p < 0.0001$).

This basic procedure may be variously enhanced, for instance, by using more than two ranges or by treating values around the median (splitting criterion) more cautiously. Alternatively, distance matrices derived from Pathfinder networks can be used. While this test might benefit from more thorough investigation, the present data showed significant consistency across tasks using the median as the splitting criterion between strong and weak ranges.

Cluster Analysis

Having observed networks containing groups of items linked by small weights, distanced from other such groups by a single larger weight, the similarity to cluster analysis became obvious. It seemed likely that often the linkage structure of some subgraph comprised only one possible selection of relationships within that group of nodes, and in fact the whole group, as a group or clique, was stable. Comparing only partial selections within such a group would fail to indicate this overall stability. For example, the GWR King, the Castle, A1, A3, and Scot are similar express passenger engines, with different pairings of these linked in different networks. However, although the A1 is linked directly

to the King in two networks, and is never far away in others, measures sensitive only to common links will miss this.

By cutting the network successively at its longest links, such groups may be identified, resulting in a form of cluster analysis. The clusters emerging between tasks may then be compared directly at a chosen grain size of clusters. This is illustrated in Table 2.

Table 2. Common elements[a] in clusters from proximity ratings and clusters from seeded-recall.

	Seeded Recall Clusters						
Proximity Clusters	A P	L K H	B G M X Y	I J R Q S T U	C D E	O N	F V W
A P Q	A P			Q			
L K I		L K		I			
B M G X Y			B G M X Y				
J R S T U				J R S T U			
C D E					C D E		
H O N		H				O N	
F V W							F V W

[a]Individual engines are designated by individual letters.

The relation between the minimum network and single-linkage cluster analysis is formally homomorphic (Dearholt, Schvaneveldt, & Durso, 1985; Dearholt & Schvaneveldt, Chapter 1, this volume), but the network is more informative, since the cluster analysis can be derived from it, but not vice versa. However, the value of identifying the clusters as done in Table 2 is to show the grain size at which there is essential identity between two networks. Although the detailed networks of Figures 1 and 3 differ, at a slightly higher level of abstraction (grain size) they are barely different, as the main diagonal alignment of Table 2 shows.

Incidental information on unstably locatable items is also evident from the cluster comparison of Table 2. For instance, the GWR Hall (item Q) is justifiably associated with the mixed traffic engines A and P, but has a competing claim to be associated with the other GWR engines (items R and S) and their group of passenger engines. Comparing the cluster structure derived from the networks thus indicates agreement between tasks at another level of abstraction.

Summary

These measures are all candidates for establishing the agreement between pairs of Pathfinder networks derived from different tasks. Although the hypergeometric test and the correlation coefficient typically indicate that statistical agreement is significantly non-random, this may not be the comparison of interest. Two networks may have very little linkage structure in common and still be shown to be statistically related.

Chi-squared analysis and comparison of clusters aim to show that independently derived datasets are not significantly different from one another, more in terms of their implied content than of their structural details. While more satisfactory measures may be developed in the course of time, it may be sufficient for the moment to note that statistical agreement can be shown among these networks, despite surface differences in structure. In the next section the significance of these surface differences will be emphasized, as the stability of Pathfinder networks as a representation is discussed.

The Stability of Pathfinder Networks

This section asks whether differences between networks are important. Consideration of what a network actually represents is useful in this regard.

Firstly, each network is a statistical summary of an originally more informative set of data, and the algorithm selects the most salient relationships in that dataset. Unfortunately, which relationships are esteemed salient will vary across contexts, and relationships strong in one set of circumstances may not apply in another. Secondly, over longer paths, both cumulative distances and any implied conceptual relation may become meaningless, so that, as an overall picture the network's value is severely limited. To illustrate this, Tversky's (1977) example serves well:

$$\text{Jamaica} \xrightarrow{\quad 2 \quad} \text{Cuba} \xrightarrow{\quad 2 \quad} \text{USSR}$$

In a linked set of countries, Jamaica may be considered similar to Cuba with a strength of (say) 2 units, and Cuba may be considered similar to USSR, also with a strength of 2. Particularly under the dominance metric, Jamaica and USSR are very close in network terms, but such implied proximity is psychologically hardly very meaningful. For such multifaceted concepts, relatedness may be judged on numerous grounds, and this must be considered when interpreting a network. The criticism may not apply to all networks but is particularly relevant to those with semantic properties. De Groot (1983) found no evidence to suggest that semantic priming persists between indirectly linked concepts, such as Jamaica and the USSR in the above example. This may imply that values of $q > 2$ should be treated with caution.

As a representation of a dataset, a Pathfinder network is subject to noise and other experimental influences in those data. The actual values of certain cells assume critical importance when the algorithm is run and lead to the formation or rejection of direct links. Small changes to these critical cell values may dramatically affect the final structure and lead to apparent instability between comparable networks. In the final section, reasons for the evident instability of Pathfinder representations are discussed.

Some Reasons for Instability

Until closely examined, it seems disturbing that the networks are so structurally different, despite demonstrable statistical agreement. Whereas the effects of noise or task context presumably contribute to this phenomenon, there is nevertheless the assumption that the network represents only a slightly distorted form of a network structure actually existent in memory. Questioning this assumption suggests no reason to suppose direct correspondence between a network and a pre-existing memory structure. This implies that any apparent instability may be a fair reflection of psychological flexibility and not merely a methodological bugbear. However, to make the point clearer, both methodological and psychological reasons for instability are examined.

Methodological Reasons

Trivially, it may be unsurprising that the networks differed from one another, after all, the elicitation tasks differed, and the proximity data were variously derived. If this is the explanation, it implies that task or experimental manipulations affect, if not determine, the relationships finally represented in a Pathfinder network. Far from these traditional procedures being passive measuring tools, they actively shape the resulting data. If this is extensive, the implications for methodology go beyond Pathfinder, since the source dataset is itself already compromised. For the moment it seems preferable to believe that the experimental task is not the major determinant of the data it produces.

Another possibility is that the variability results purely from noise, reflecting the finding that people are just not good at estimating numbers. While this may be part of the explanation, it is unlikely to be all of it. For instance, in the repertory grid and rating scales tasks, a locomotive's properties were considered individually, and rated on a scale of only 7 points. It seems unlikely that expert estimates could be seriously wrong under these conditions. Even if they were, the recall and twenty-questions tasks did not involve assigning numbers at all.

To say that people are not good at estimating numbers does not explain why estimates are unreliable in the first place. Probably here it is precisely because both relationship strength and the saliency of properties are relative to context, and there is no absolute value for relatedness judgments. The task demands of estimating relatedness are likely to be met by rough computations, guided largely by the expert's own experience and biases. For the data presented here, similarity of duty and size clearly contributed much to the consistency. However, relatedness judgments by an expert for whom design and vintage were important would produce a very different relationship pattern.

A third possibility is that all these networks are merely selections from a more comprehensive representation, for example, the total union of all defensible direct associations. It may be that no particular network is the only correct structure in memory, but that their complete combination is. Although a completely linked network might be a truer representation of connectivity, it must have differentially weighted links if it is not to be vacuous. For if everything is connected to everything else, and there are no intrinsic distinctions of weight, then the network in itself is utterly uninformative. However, in practical domains it is unlikely that all connections would be defensible, and for most intents and purposes it is sensible to establish the subset of useful ones. Pathfinder is a principled attempt at this.

However, there seems no reason to suppose that weighted links as such have any a priori existence in memory. Weightings that reflect the importance of relationships are likely to be computed online and in specific contexts (cf. Barsalou, 1987; Kahneman & Miller, 1986). Furthermore, the particular links formed between domain concepts are likely to be affected by factors extrinsic to properties of those concepts (Barsalou, 1985; Murphy & Medin, 1985; Schank, Collins, & Hunter, 1986). These and other psychological factors affecting observed structure are considered below in more detail. So even though regarding networks as selections from a totality at first seems promising it may lead to empirical vacuity (cf. Johnson-Laird, Herrmann, & Chaffin, 1984).

Psychological Reasons

A link-weighted network of concepts may be viewed as descriptive of semantic memory (in this case of an individual expert's category structure). However, the explicated network results from the interaction of pre-existing memory with various external demands. Recent research has identified various external factors which bear on the stability of conceptual representation and therefore may affect the resulting network. These include goals and needs (Gluck & Corter, 1985), context (Barsalou, 1987), and personal theories (Murphy & Medin, 1985).

Gluck and Corter (1985), for instance, note that traditional measures of category structure (such as cue validity) ignore the contexts and needs of people who create and use concepts. They propose a context-sensitive measure of the utility of categorizations. Schank et al. (1986) make similar points in their discussion of inductive category formation. They note that the function for which categories are formed helps to determine their structure, and this lies outwith the scope of purely inductive category formation systems, which operate on the intrinsic properties of concepts.

Murphy and Medin (1985) emphasize the importance of personal theories in shaping conceptual structure and give examples showing how such theories provide a coherent context in which concepts can meaningfully relate. Thus the seemingly disparate set of shrimps, moths, and grasshoppers becomes cohesively related given a deeper understanding of biological structure, in this case, a theory of the class of arthropods. Analogously, an individually held belief set gives a context of meaning within which conceptual relationships have cohesion. An expert's set of attitudes to a domain will affect its conceptualization, and some examples of this are demonstrated in Gammack (1988).

Barsalou (1982, 1985, 1987) has perhaps provided the most sustained arguments against the static definition of concepts. In Barsalou (1987), for example, his primary concern is with the instability of graded structure, that is, the extent to which exemplars typify a category. Barsalou shows how this varies with context and suggests that categories do not have an invariant representation that we should be trying to discover. Instead he proposes that concepts are temporary constructions in working memory, drawing upon (rather than retrieving intact) knowledge in long-term memory. This research suggests that long-term memory is relatively unstructured, supporting the flexible construction of concepts rather than containing knowledge in well-bounded packets of characterizing properties. This view is consistent with the present results.

Role of Pathfinder in Knowledge Elicitation

If the instability noted in this chapter signifies an inescapable issue not only for Pathfinder but for psychology in general, it would be easy to be pessimistic and conclude that a static description represented nothing more than the specific product of a few arbitrary conditions. However, this negative view ignores the many valuable gains to be made by considering instability seriously.

Supposing that the expert has a uniquely configured, stable network of 25 domain objects, which looks in retrospect, naive. Furthermore, the assumption that this knowledge structure would be transparently elicitable by objective methods having no influence on underlying structure also looks questionable. Since the tasks were different in nature, it may have been unreasonable to expect anything else. Clearly the tasks have had an effect in that they have elicited different structures, suggesting that they may be addressing different aspects of the same knowledge. Knowledge itself is surely stable, but representations of it are not, despite their specific value in particular contexts. Thus the bipolar scales and the repertory grid tasks act as mutual consistency checks on a constrained set of properties, whereas the proximity estimates permit other bases for relatedness than similarity of a small set of properties. The elicitation tasks do appear to have individual properties which make them nonequivalent and thus uniquely valuable. The extent to which they help shape (if not determine) the form of elicited knowledge is a methodological issue which requires fuller investigation.

One argument against the existence in memory of an absolute weighted structure reflecting a "true state" is the effect of circumstantial changes on relatedness estimates. Although the relating criteria used here were freely chosen by the expert, an instruction to rate proximities in terms of vintage or design would have produced different patterns, as noted earlier. Exigencies arising in practical settings will have a similar effect. While some conceptual relations are likely to be more strongly held than others, and likely to emerge frequently across contexts (such as the Bulleid Merchant Navy and Bulleid West Country in this domain), there will be occasions when this relationship is relevant, and others when it is not. A weighted structure in itself describes something about an expert's domain knowledge, but one must recognize its range of applicability. Accordingly, when eliciting an expert's knowledge of the relatedness of domain concepts, the context in which two items are related must also be identified and acknowledged. It is likely that this will not be given merely by a linear combination of "intrinsic" properties, but by extrinsic considerations which give the concept its current definition. The meaning of a concept is not fully determined as a function of its analyzed parts, but is assigned within a context and background outwith a static representation (see Gammack & Anderson (submitted); Gregory, 1986; Winograd & Flores, 1986).

Theoretically, Pathfinder is likely to prove very useful in investigating and representing the stabilization of a systematic set of conceptual relationships. Instead of viewing Pathfinder as faithfully mapping a stable conceptual structure in memory, a narrow view in which personal interpretations, changing contexts, and methodological properties are problems to be reduced as far as possible, Pathfinder can become a technique in investigating those very areas.[1] For instance, a novice's domain theory, mapped out using Pathfinder and augmented with descriptions justifying the structure, can be compared with the same individual after training or with another expert to identify the effect of experience on domain conception. Whereas practicing experts are not limited to a single inflexible structure (see Murphy & Wright, 1984), such structures are likely to be particularly useful in formal teaching, routine tasks, and many other situations for which expert knowledge is elicited.

Given the existence of domains or parts of domains of expert knowledge that can be profitably systematized, Pathfinder has many advantages in producing a concise, visualizable, and informative representation. Several expert domains have already been modeled in which Pathfinder networks were an important part of the represented knowledge. In addition to domains described by Roger Schvaneveldt and his colleagues, these include choice of statistical analysis techniques and information scientists' sources (Gammack, 1988), central heating systems (Gammack, 1987a), and locomotive classification (Gammack, 1987b).

Being suitable for modeling the knowledge of experts as individuals or in groups, and having a variety of possible sources of input data, Pathfinder is a particularly flexible way of representing implied relationships among domain entities. The formal and quantitative aspects of Pathfinder networks are also attractive as a representation which knowledge engineers can use in producing expert system code.

[1] I am grateful to Nancy Cooke for reminding me of this.

Conclusion

In this chapter, an attempt was made to model an expert's conceptual structure using Pathfinder (and other) representations. However, across different elicitation conditions, no unique conceptual structure emerged when it might have been expected. Methodological and psychological reasons were considered as explanation for this instability, and with hindsight, it seems sensible to expect structural variation for a variety of reasons.

While the underlying knowledge may be relatively stable, structured representations of it are variable, context specific, and considered to be secondary phenomena of constructive processing, rather than homologues of any pre-existing memory structure. This implies that any given Pathfinder network is unlikely to represent a true state of memory, but a product of memory's interaction with task, contextual demands, and other factors extrinsic to similarity of defining properties. With a static description of relationships such as a network, one must also acknowledge the contexts in which domain objects relate coherently.

The instability of representation is neither caused by, nor unique to Pathfinder, which provides a tool for investigating the sources of instability in conceptual representation, such as individual theories and biases, and task or other demands. Despite some surface variation, Pathfinder descriptions were found to be particularly useful in the context of knowledge elicitation where concise and meaningful representations of expert domain conception were reliably produced.