IMAGING SERVICES NORTH
Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

This PDF was created from the British Library's microfilm copy of the original thesis. As such the images are greyscale and no colour was captured.

Due to the scanning process, an area greater than the page area is recorded and extraneous details can be captured.

This is the best available copy

DX
172349

THE BRITISH LIBRARY
BRITISH THESIS SERVICE

TITLE .. The Development of Semantic Memory
A Spatial Model of Animal Concepts
in Schoolchildren, Novices and Experts

AUTHOR .... Denise Ann Hale ....

DEGREE ....

AWARDING BODY City of London Polytechnic
DATE .... (1991) C N A A ....

THESIS
NUMBER ....

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

The quality of this reproduction is dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction.

Some pages may have indistinct print, especially if the original papers were poorly produced or if the awarding body sent an inferior copy.

If pages are missing, please contact the awarding body which granted the degree.

Previously copyrighted materials (journal articles, published texts, etc.) are not filmed.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no information derived from it may be published without the author's prior written consent.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

cms 1 2 3 4 5 6

REDUCTION X 20
CAMERA 5
No. of pages

The Development of Semantic Memory

A Spatial Model of Animal Concepts in Schoolchildren, Novices and Experts

Denise Ann Hale

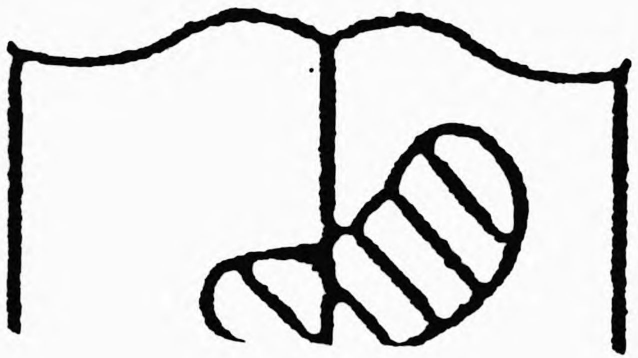
October 1991

A thesis submitted in partial fulfilment of the requirements of the Council of National Academic Awards for the Degree of Ph.D

City of London Polytechnic in collaboration with
The London School of Economics

CONTENTS

VARIABLE PRINT QUALITY



THE DEVELOPMENT OF SEMANTIC MEMORY

DENISE A HALE

Children s and adults judgements of semantic relatedness amongst a set of animal terms were compared in order to examine the validity of a spatial model of conceptual development in schoolchildren.

The individual differences multidimensional scaling model, SINDSCAL, represented the animals as points in an interpretable 3-dimensional configuration. Individual differences did not emerge strongly, but predictions that children would attach more importance to concrete dimensions and display less taxonomic clustering relative to adults were partly supported. A model relating times taken to make judgements about the animals to their locations in the configuration received little support. The spatial model displayed individual differences only when children and adults were compared with experts.

To test the sensitivity of judged semantic relatedness to variations in knowledge, features of the animals were elicited from new groups of children and experts. The groups produced similar data, but experts made more use of features relating to formal taxonomy. Comparison of feature overlap proximities with judged relatedness indicated that relatedness judgements do not reveal the complexity of children s semantic structures. When both types of data were analysed by ADDTREE, a clustering procedure, trees were found to provide a better fit than SINDSCAL, particularly for the children s data.

The similarities and differences between children, adults and experts suggested that while they share complex conceptual structures reflecting superficial similarities amongst everyday objects, experts have access to deeper knowledge underlying scientific definitions and taxonomies. This result is compatible with an existing feature model of semantic memory in which some features of a concept are defining, while others are merely typical. The findings are interpreted in terms of other developmental research indicating that concepts are best understood as beliefs or theories about regularities in the world.

ACKNOWLEDGEMENTS

I would like to thank all who participated in this research and to express my gratitude to the parents and teachers of the children involved.

Grateful thanks are also due to Graham Clarke for his assistance with computing and to my supervisors, Janet Stockdale and Veronika Coltheart and Paul Willner.

I would particularly like to thank Veronika, my husband, Anthony Hale, and my family for their support and encouragement over a long period.

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

CHAPTER 1

SEMANTIC MEMORY

1. A Developmental Approach to Semantic Memory

Research into the representation of word meaning in memory has generated various of models describing how information about words is stored and retrieved. Central to this area of language research is the notion that people must have stable internal conceptual structures which underly their everyday understanding of words. Tulving (1972) used the phrase 'semantic memory' to refer to this long term store, distinguishing it from 'episodic' memory which, as the name suggests, refers to the more everyday understanding of memory as the mechanism by which specific events are recorded and recalled.

In contrast with early memory research which focused on the mechanics of recall, work in the later area of semantic memory involved tasks designed to reveal organisational principles of knowledge and test the extent to which structural theories can predict linguistic behaviour. The organisation of knowledge was generally inferred from sorting and rating tasks, while structural theories, constituting the core of the area known as semantic memory, were derived primarily from experiments measuring how long people take to make simple decisions about the meaning of words or short sentences. Although most experiments dealt with semantic representations of single words, theories of semantic memory were seen as addressing the representation of concepts or categories. It was generally accepted that the linguistic equivalents of some concepts can be descriptions rather than single words (Lachman and Lachman, 1979).

The developmental studies reported in this thesis were carried out because although semantic memory research had identified appealing principles for conceptual structure in general, these had not been tested outside laboratory settings or with varied groups of subjects. Also early theory had been rapidly revised to accommodate effects obtained both with similar material in different tasks (Rosch, 1973) and with a broader range of linguistic material in standard semantic memory tasks (Wilkins, 1971; Conrad, 1972, Rips et al, 1973). It was therefore sensible to ask whether the revised principles of semantic structure could represent behaviour across a wide range of subjects, differing in age, educational background or culture.

Developmental issues had not been tackled within traditional semantic memory research. Indeed, research topics and theories appeared to be based on the understandable assumption that the important thing about concepts is their relative stability and that change owing to individual differences might be intrinsically uninteresting. However, concepts develop somehow and studies of language acquisition and development were a source of hypotheses about how word meaning might develop and what a model of semantic memory should encompass. There was therefore a need for descriptive work on the development of semantic structure and for experiments to test the validity of existing models in this context.

Chapter 1 outlines the scope of semantic memory research, reviewing the major theories and their limitations. The importance of semantic distance and its representation in spatial models of semantic memory is identified as a framework for developmental study. Chapter 2 examines findings in language acquisition and linguistic development in order to generate hypotheses about how semantic structures might change with age. Chapter 3 describes the use of spatial models to represent semantic structures in adults (Henley, 1969; Rips et al, 1073, Caramazza, Hersh and Torgerson, 1976) and the ways in which the approach can be improved.

2. The Role of Semantic Distance

In the 1970s research in semantic memory was dominated by two major and apparently competitive theories, the Feature Comparison theory put forward by Smith, Shoben and Rips (1974) and the Spreading Activation theory described by Collins and Loftus (1975). Contemporary with these were similar less general models designed to explain discrepancies in processing times when sentences contain quantifiers such as SOME of ALL (Meyer,1 970; Glass, Holyoak and O'Dell, 1974).

A central feature of all of theories was the notion of semantic distance between concepts as an explanation for differences in processing times. Similar concepts such as ROBIN and SPARROW are semantically 'close', while a dissimilar pair like ROBIN and SHOE would be distant in semantic space. Rips et al (1973) invoked semantic distance to explain what was called the 'subset' effect, in which 'subset' sentences like 'a robin is a bird' are verified faster than 'superset' sentences like 'a robin is an animal'; and also to explain effects of semantic relatedness in other speeded tasks.

The overall effect of semantic distance may be seen in same-different tasks in which subjects are asked to decide whether words presented as pairs belong to the same or different categories. For example, ROBIN-SPARROW would elicit a 'same' response, while ROBIN-ELM would be treated as 'different' in decisions involving the categories BIRD and TREES. In both of these cases, semantic distance is thought to facilitate or speed up the response: people are quick to decide that similar concepts belong to the same category and that different concepts belong to different categories. If, however, a third category, say MAMMALS, is introduced, semantic distance is seen to operate in the opposite manner. That is, subjects are slow to respond 'different' to ROBIN-CAT because although in a different category, these items are semantically close as they are both animals. When items are semantically close, same judgements are facilitated and different judgements are impeded, but for semantically distant items, same judgements are impeded and different judgements are facilitated (Schaeffer and Wallace, 1969; 1970).

Following these initial observations, research in semantic memory expanded hugely. Numerous findings were reported concerning same-difference judgements and other categorisation tasks. It is sufficient to note that semantic distance effects are reliable, but they do not operate in the same way for all material. For example, Schvaneveldt, Durso and Mukherji (1982) reported that while relatedness facilitated positive responses for a range of categories, it did not inhibit response equally for negative responses. Thus the nature of the material is important. Various theories were proposed, but no definitive model emerged. Research then declined in a pessimistic atmosphere noted by Chang (1986), whose review of the major findings and models in semantic categorisation reflected the complexity of the field. Even in 1981, Medin and Smith noted that to attempt an adequate review of the feature comparison and spreading activation models would 'bog us down in a quagmire of experimental details' (p83).

Johnson-Laird (1983) later summarised the difficulty saying that semantic memory had suffered the fate of all paradigm-driven research, that is, the central observations were given alternative explanations and then empirically shown to be extremely complex. Some of the significant findings are discussed here in the context of the major theories to assess the achievements of the field.

1.3 Theories of Semantic Memory

1.3.1 The Spreading Activation Network

Ideas in artificial intelligence (AI) shaped theory in semantic memory. The contribution of AI was that it offered an alternative to traditional experiments on humans: proponents of AI argued that a computer programme that functions like a person represents a psychological theory (Quillian, 1968) and that the study of such programmes can generate useful hypotheses about mental processes which may then be tested using human subjects. This approach stimulated a spate of research investigating the extent to which computer models of information storage and retrieval can predict human behaviour.

Collins and Quillian (1969) reported an influential test of their spreading-activation model of semantic memory. This theory assumes that concepts are nodes in a network where the interconnections are relational labels, such as ISA, HAS or CAN, which specify the properties of the concepts. For example, 'a canary ISA bird' represents one property of canaries. Once a node is activated, for example, by presentation of the word corresponding to it, activation spreads along the links in a manner analogous to nervous conduction, thereby activating neighbouring nodes. This means that processing times for lexical decisions should be a functions of the spread of activation and of the relative locations of concepts in the network.

Collins and Quillian (1969) assumed that information would be stored in a rational manner such that subordinate-superordinate or ISA relations would form a hierarchy and properties applying to concepts at various levels in the hierarchy would be stored only once at the highest level to which they apply. In their well-known example of the hierarchy relating CANARY to BIRD and BIRD to ANIMAL via ISA links, the property SINGS is stored with CANARY, while FLYS is stored with BIRD and HAS SKIN is stored with ANIMAL. The predicted verification times for this structure were confirmed: that is, subjects were faster to decide that A CANARY IS A BIRD than that A CANARY IS AN ANIMAL because the first judgement required one ISA link while the second required two links in the hierarchy; equally, verification of A CANARY CAN SING was faster than A CANARY CAN FLY, which was in turn faster than A CANARY HAS SKIN because these properties were stored at successively higher levels in the hierarchy.

This attractive formulation soon proved to be an oversimplification, although in fairness, it was later claimed that the Collins and Quillian paper had only described aspects of the theory relating to their experiment and that the subsequent challenges resulted from misunderstanding this specificity (Collins and Loftus, 1975). Wilkins (1971) argued that controlling for Thorndike-Lorge frequency was inadequate and that the frequency of cooccurrence of subordinate and superordinate was critical. Using published production norms measuring the frequency with which instances are generated in response to a category name, he found that this 'conjoint frequency' predicted verification times, thereby refuting the implication of the Collins and Quillian report that all ISA links are equivalent.

Conrad (1972) put forward a similar argument for the storage of properties when she rejected the assumption of cognitive economy which states that properties are stored once and must be retrieved via inferences for all conceps except the one with which they are directly stored. Again, production frequency of properties in response to concepts predicted verification time, not the position of the logically essential level in the hierarchy.

Rosch's complementary research on the internal structure of natural categories (Rosch , 1973: Mervis and Rosch, 1981) presented further difficulties. In a series of papers beginning in the 1970's, Rosch and her collaborators described characteristics of natural categories such as ANIMALS and TOYS. They demonstrated that natural conceptual behaviour differs from formal models which were developed to explain concept formation for simple perceptual stimuli which vary systematically on a few easily identified dimensions (Anglin, 1977).

Rosch's most significant finding was that category membership is not all or none; rather it is graded with some instances being better examples of the category than others. For example, ROBIN is better that PENGUIN as an instance of the category BIRD. Known as typicality, this appears to be a semantic distance effect which operates within concepts.

Rosch also identified a unique type of category known as 'basic'. These, the first to be learned, are used in most everyday situations. They also readily give rise to visual images and can be drawn by most people. A good example of a basic category is the term CAT: most people have a strong general concept of the objects that fall into this category. Although a minority may distinguish between the varieties of CAT, even these people would probably use the basic terminology for most purposes.

The appeal of Roschs' work was that it related directly to everyday concepts. Similar findings have been reported in anthropological studies of conceptual structure (Berlin, 1978). This was in direct contrast to research in semantic memory which concentrated on conceptual structure in educated individuals and has no obvious parallels in other areas of linguistic research.

In answer to these criticisms, Collins and Loftus (1975) described a revised version of the original spreading activation network model. As before, concepts were related by links such as ISA between ROBIN and BIRD and people were thought to retrieve the ISA relation when verifying sentences.

However, the idea of a strict hierarchy was abandoned and differences in verification times were attributed to the lengths of links and to their 'criteriality'. Long links led to slow reaction times, but links varied in the extent to which they facilitated the spread of activation and this facilitation - or lack of it - was unidirectional. The result was an asymmetry in the relation between concepts: for example, activation might spread more rapidly from ROBIN to BIRD than from BIRD to ROBIN.

1.3.2 The Feature Theory

Meanwhile, Smith, Shoben and Rips (1974) had published a Feature model in which concepts were viewed as bundles of features. In verifying sentences, people were thought to compare the features of ROBIN with those of BIRD and answer affirmatively if sufficient positive evidence accrued from the comparison.

The Feature theory was initially less attractive than the Network theory because it requires concepts and features to be interchangeable: any concept is a bundle of features, each of which is a concept, itself defined by a set of features. For example, WINGS is a feature in relation to the concept BIRD, but it is also a concept in its own right, possessing features such as FEATHERS. This duality of status did not arise in the Network model which treated all entities as concepts.

In the Feature model differences in verification times were attributed to the comparison process. Using the category BIRD as an example, Smith et al (1974) claimed that certain features are necessary for an instance to be categorised as a BIRD and that these features may be considered 'defining'. In reality, however , birds possess a variety of additional features, some of which are highly typical or 'characteristic' of the concept BIRD, while others are not.

In comparing ROBIN with PENQUIN, one would find that both concepts share the defining features of BIRD, but ROBIN would have a greater number of characteristic features. In order to categorise PENGUIN, only the defining features constitute positive evidence, while in classifying ROBIN both types of feature would indicate a match. This is the basis of a 2-stage process: in stage one all features are compared, while in stage two only defining features are compared. For typical instances of the category, sufficient evidence for a positive response would accrue from a stage one comparison, but for atypical instances, the second stage would be necessary.

The effect of semantic relatedness was seen in the latencies to false items: unrelated false pairs were correctly verified quickly, while related false pairs, such as TREE and ANIMAL were error prone because they shared a number of defining features. In this way the Feature theory accounted for processing times when semantic relatedness and typicality are manipulated.

1.4 Limitations of the Network and Feature Theories

A fundamental problem was the lack of differences between the apparently competing Network and Feature theories. Hollan (1975) pointed out that any feature model can be reconstructed in terms of a network model. In Hollan's view, Smith et al's feature model is a network in which a given concept is linked to each feature. The value assigned to each link indicates the importance of the feature to the concept and defining versus characteristic features can be differentiated by comparison with a threshold value. Thus the network consists of ordered pairs of concept and feature.

Smith et al's criticism that networks fail to provide for graded category membership was thought to derive from a misunderstanding of networks as described by Collins and Quillian and by subsequent authors since this restriction only applies to some networks. Collins and Loftus (1975) recognised the similarity of network and feature formulations, noting that networks were originally devised as a means of storing features in a computer.

Perhaps owing to the complexity and numerous assumptions of the Network theory, it has proved difficult to refute and subsequent investigations have tended to concentrate on the more specific proposals in the Feature theory. In the course of the current study, various findings arose which cast doubt on the adequacy of the major theories and the methodologies that support them.

For example, McClosky and Glucksberg (1979) tackled the Feature theory. They noted that when the subject and predicate in a verification task are highly related in a false sentence, there should be a tendency to produce errors. An example of this kind of sentence is ALL BIRDS ARE ROBINS. In order to avoid errors, the internal criterion for evidence would have to be high enough to prevent stage one confirmations based on characteristic features. Correct responses could only result from stage two comparisons.

They suggested that if all comparisons use both stages the relatedness effect for true sentences should disappear since this is due to the difference in processing time between one comparison and two. Using a task in which false sentences were as related as true sentences, they found that the relatedness effect remained: high related/true sentences were verified faster than low related/true sentences, but both of these were processed more slowly than unrelated/false sentences.

A further difficulty for the Feature model came from Hampton's (1979) finding that for six out of eight categories, no set of features were, in conjunction, necessary for category membership. The existence of features which may be considered 'defining' applied only to the categories BIRD and VEHICLE. Hampton also found that typicality was associated with the degree to which all features are present and not just with the presence of apparently 'characteristic' features. However, when the relationship between the featural definition of concepts and their rated typicality was examined, a common model did not suffice for concrete and abstract concepts, suggesting a lack of generality in formulations based on concrete concepts, such as the Network and Feature theories (Hampton, 1981).

Another problem for both Feature and Network approaches was Hamptons (1982) finding of intransitivity in class inclusion judgements. For example, people will agree that A CHAIR IS FURNITURE and that A CAR SEAT IS A CHAIR, but not that A CAR SEAT IS FURNITURE. In the Smith et al (1974) model, the defining features would be nested so that this result should not occur, while in Network model, the class inclusion relation, ISA, is always specified, again making intransitive judgements impossible.

McCloskey (1980) then raised a more fundamental objection to models of semantic memory and the sentence verification methods used to test them. As a result of examining the materials used in three experiments, including his own with Glucksberg (1979), McCloskey suggested that stimulus familiarity might be responsible for many accepted effects in semantic memory. For example, Smith et al (1974) had attempted to vary category size and relatedness independently in a sentence verification task and found that relatedness and not category size predicted latency of response. A COLLIE IS A DOG was faster than A COLLIE IS AN ANIMAL, where COLLIE is more related to DOG than to ANIMAL and DOG is a smaller category than ANIMAL; but a CHIMPANZEE IS A PRIMATE IS slower than a CHIMPANZEE IS AN ANIMAL, because CHIMPANZEE is more related to ANIMAL than to PRIMATE, even though PRIMATE is the smaller of the superordinate categories.

Smith et al had obtained relatedness ratings for items within small category and large category pairs from different groups of subjects because subjects tended to use category size as a cue in making the relatedness judgement. McCloskey argued that different groups of subjects might use the scale differently, making it invalid to conclude, for example, that COLLIE is more related to DOG than to ANIMAL or vice-versa. Using two groups of subjects, each rating relatedness of small and large category pairs, McCloskey found that Smith et al's instances were all more related to the small superordinate than to the large superordinate. Why then were the latencies for CHIMPANZEE-PRIMATE slower than those for CHIPMANZEE-ANIMAL? This finding was difficult to interpret, because PRIMATE is both the smaller and the more related superordinate, which according to numerous other studies should ensure the faster latency.

McCloskey replicated Smith et al's results and then altered the procedure to test the hypothesis that the unfamiliarity of words like PRIMATE lay behind the apparent inconsistency. Subjects viewed the superordinate before pressing a key to display the instance, thereby permitting a measure of comprehension time. Different subjects rated the familiarity of the superordinate terms. The result was that small superordinate categories that were also unfamiliar took longer to comprehend, but categorisation latencies followed the usual pattern of being faster for small categories and for more related pairs.

McCloskey pointed out that his three examples of familiarity confounds were not obtained from an exhaustive search through the literature and that in two cases, the stimulus materials had been carefully controlled for frequency and word length. There is therefore a clear case for caution in interpreting some of the more complex semantic memory effects and a need for experimental designs which facilitate the control of extraneous variables.

1.4 The Role of a Developmental Study in Semantic Memory

After an promising start, semantic memory had become an extremely complex field with an unsatisfactory theoretical basis. Research was characterised by the use of limited sets of semantic material in restrictive tasks and, in most cases, the tasks had been completed by highly sophisticated subjects.

A consequence of this narrow approach was a tendency for theory to become unparsimonious. This was particularly evident in the Network theory, which had tended to embrace new findings by increasing its assumptions. The initial difficulties had arisen because certain tasks and material produced inconsistent results, thereby indicating that the major models lacked generality. It was also noticeable that neither the Network nor Feature theory dealt with the issues of individual differences and development. As Ornstein and Corsale (1979) said, they take 'a static view ......with no real treatment of how new information becomes integrated within the existing knowledge structures or of how these structures are modified by experience (p247)'. Therefore a developmental study could provide a further test of the generality of semantic memory models and supply basic information about how concepts develop, which existing models lacked.

Despite the emerging complexity in semantic memory, basic findings about relationships between semantic distance and some aspects of linguistic behaviour were reliable. In particular, some researchers had used subjective ratings of semantic distance to create models of semantic space and established that the models could predict reaction times in various tasks. Therefore a spatial model offered a framework for extending hypotheses about semantic memory in adults to children. In addition a spatial model provides dimensions which should correspond to the important features of semantic space. Therefore it has the potential to identify how semantic structures change with age.

The individual differences multidimensional scaling model (MDS), INDSCAL (Carroll and Chang, 1970) was chosen because it can incorporate both differences between individuals and the semantic distances between words in a single model of semantic memory. MDS is a family of techniques for representing relations between objects as distances in space, such that similar objects are close together and dissimilar objects are distant. If the objects are words and the relations between them are peoples' judgements of semantic similarity, an MDS model can act as a representation of semantic space.

A further rationale for this approach was that structural models similar to MDS had been used to investigate some aspects of conceptual development. This work and relevant issues in language acquisition are reviewed in Chapter 2. The use of MDS in modelling adult sematic memory is described in Chapter 3.

CHAPTER 2

SEMANTIC DEVELOPMENT

2.1 The Research Context

The primary aim of this study was to examine the extent to which a spatial individual difference model of semantic memory would represent developmental changes in the subjective lexicon. Structural representations of semantic development using cluster analysis and factor analysis had been reported, but there was little integration between these models and the much larger body of work on language aquisition and semantic aspects of cognitive development. Therefore trends in linguistic development had to be identified and related to the structural studies in order to generate hypotheses about what the spatial model might show.

The review of language acquisition concentrated on early word meaning. In some cases information about linguistic development had to be extracted from studies of learning and memory, as there was comparatively little research on development itself after the acquisition of basic linguistic skills. The major theme to emerge was that early language is dominated by perceptual features and thematic relations, with the ability to handle non-perceptual information and taxonomic relations developing gradually in schoolchildren. Evidence for this is discussed in the first part of this chapter.

Of the structural studies, the best known was Anglin's (1969) use of MDS and cluster analysis to compare children's and adults' behaviour in tasks involving sorting and free recall of words. Lesser known investigations followed the Semantic Differential tradition in which the meanings of concepts are measured on bi-polar scales (Osgood, Suci and Tannenbaum, 1957). Divesta (1966) used factor analysis to examine the importance to children of the various types of scales which describe adult semantic structure. In a similar study, Saltz, Dunin-Markiewicz and Rourke (1975) used an early individual differences scaling model, CANDECOMP (Carroll and Chang, 1970), to examine developmental differences in correlations between attributes. Anglin was explicit in describing his results in terms of a concrete-abstract progression and the findings of the factor-analytic studies, although interpreted differently by their authors, were consistent with this view.

Conceptual development as a focus for research grew in popularity during and after the studies reported in this thesis were completed. Some key findings are mentioned in this chapter, but a broader account is presented in chapter 10 and interpreted together with conclusions from the findings of this research.

2.2 Semantic Development: From Concrete to Abstract

2.2.1. Separating Language Acquisition and Development

Much of the literature on language acquisition reports investigations of children's first utterances and focuses on overextensions and underextensions as indicators of semantic features which are salient for young children. Overextensions occur when infants incorrectly apply a single word to a large category of superficially similar objects. For example, Bowerman (1976) noted that her daughter used 'moon' to refer to a piece of grapefruit, an overextension based presumably on colour and shape. Underextensions are the use of a word to refer incorrectly to a restricted category, often reflecting some irrelevant thematic context. For example, the word 'car' might be applied only to moving cars.

Studies with older linguistically competent children were more concerned with learning and how it is affected by the way in which children encode and organise linguistic stimuli. An example is Means and Rohwer's (1976) study of recognition memory in which the effects of acoustic, visual and semantic distractors were compared in order to determine the relative importance of these factors in the memories of children aged seven and twelve. They found that while visual distractors affected both groups, acoustic distractors led to false recognitions amongst the seven year olds and semantic distractors attenuated the performances of both older children and seven year old girls.

Although based on different paradigms and concerned with children at various stages of development, from prelinguistic infants to competent adolescents, studies of language acquisition and memory development displayed a common theme which may be described as the concrete-abstract progression (Anglin, 1969). This refers to the general finding that language and memory in young children is characterised by concrete attributes, while older children display an ability to handle abstract attributes.

That concrete attributes predominate in language acquisition was not of central interest in this research. Of greater concern was whether concreteness would characterise semantic development in linguistically competent children. Evidence for the persistence of a visual coding system (Pavio, 1971) or at least for the utility of visual coding as superficial strategy rather than fundamental representation (Pylyshyn, 1973) indicates that concreteness affects cognition in adults, but not the extent to which it characterises semantic memory. Perhaps semantic structures at different ages vary in terms of how much they hold or are organised by concrete features. When adults use imagery this might represent a choice to concretise abstract information for mnemonic again, while children use imagery out of necessity rather than choice.

2.2.2. Language Acquisition

An important issue in language acquisition was whether infants acquire information about form before function as suggested by Clark (1973), or whether early word meanings are built around the functions of objects in relation to the child (Nelson, 1974). Although evidence existed for both theories, Tomikawa and Dodd (1980) noted that there was more support for the pre-eminence of form. A problem with some of Nelson's stimuli was that objects which differed in function also appeared to differ in form. For example, plastic eating utensils were probably more similar to one another perceptually than they were to plastic animals, which might explain why infants' choices appeared to follow the functional difference between the two groups of objects.

Tomikawa and Dodd varied form and function independently and found strong support for the perceptual view. Since perceptual attributes such as shape and colour are intuitively concrete relative to the variety of functional relations that exist between the child and external objects, the weight of evidence in this debate appeared to support a concrete-abstract progression.

2.2.3 Linguistic Development

A similar developmental trend appeared when older children were confronted with items which may be grouped according to colour, shape or membership of a superordinate category. Melkman, Tversky and Baratz (1981) demonstrated a clear shift from colour, through shape to taxonomic or conceptual organisation between the ages of four and nine. However, when colour, shape and superordinate cues were supplied as an aid to recall, superordinate cues were most effective for all age groups, followed by shape and then by colour.

The authors suggested that when children group objects they abstract features upon which to base their groupings and that easily abstracted features such as colour and form have an early advantage. In contrast, retrieval involves selecting the correct item from a number of possibilities and in this situation, colour is least helpful while form and superordinate terms both narrow the field considerably.

In order to trace the continuation of the concrete-abstract progression with age, it is important at this point to compare the terminology used in infant studies with that applied to schoolchildren. In language acquisition research the term 'functional' refers to uses of objects and what the objects themselves do (Nelson, 1974). In this sense, infants' choices based on functional organisation may be viewed as abstract relative to choices based on form which depend on the physical or perceptual attributes of objects. For example, a group of toys or eating utensils displays a 'functional' relation, while round or red objects may be grouped 'perceptually' or by 'form' according to shape or colour, irrespective of function.

In studies with toddlers and school children, the term 'functional' also describes use, but these studies demonstrated that younger children display 'functional' organisation, for example, grouping animals according to their use as pets or for food, while older children prefer groupings described as 'abstract', for example, plant eaters versus meat eaters (Storm, 1978). This view followed Parker and Day (1971) in defining organisation according to use as 'functional' and that according to a class name as 'abstract'. Storm justified the dietary distinction as abstract because it is fundamental to zoological classification. Therefore, for older children, functional categorisations are seen as concrete relative to those based on formal methods of classification, in contrast to younger children whose functional groupings are abstract relative to categories based on form.

Another manifestation of the development of abstract organisation during school years is the thematic-taxonomic shift (Tenny, 1975; Anglin, 1969), which contrasts associations based on contiguity or characteristic behaviour with taxonomic organisation deriving from class-inclusion heirarchies. For example, Anglin (1969) described the young child's choice of FLOWER-GROW in a sorting task as thematic in contrast to the adult preferences for FLOWER-HORSE.

Smiley and Brown (1979) used picture triads to test preferences for thematic versus taxonomic organisation. For example, subjects were asked whether NEEDLE goes with THREAD or PIN. Pre-school and first grade children preferred thematic pairs such as NEEDLE-THREAD, while fifth grade and college age subjects chose the taxonomic alternative. However, this study also illustrated the difficulties that bedevil this approach, namely the confounding of reasons for categorising stimuli. It is difficult to decide whether NEEDLE and PIN go together because they look the same, because they are both used in sewing, or because they fall into some general category which may be described as long, thin, sharp, silvery objects and which could include physically dissimilar objects such as STILLETTO and SWORD.

There were further interesting components to this study. All subjects were asked to justify their choices and these were rated by undergraduates with respect to thematic and taxonomic understanding. With the exception of those supplied by the pre-school children, justifications could be rated easily. Subjects were also asked if there was any way that the standard and non-preferred item could go together and these justifications were also rated. First graders and above were able to justify the non-preferred choice with clear understanding of the alternative mode of categorisation. In a second experiment it was shown that pre-schoolers were able to learn to make taxonomic choices, but reverted to their original thematic preferences when tested on the following day.

The researchers also tested a group of elderly subjects whose preferences, like the preschoolers, were for thematic pairs. Without the influence of formal education, the elderly were thought to have reverted to child-like modes of thought, but it could not be shown that those particular elderly persons had ever held a preference for taxonomic organisation.

Storm's (1978) study of functional versus abstract organisation of animal terms was of greatest relevance to the current study. Subjects aged eight to thirteen were asked to memorise animal terms presented either as a list or as a hierarchy in which the animals were classified according to function (pets, fur, game) and then aggregated into meat-eaters or plant-eaters. As predicted, the older subjects were better at reproducing the higher level of the hierarchy (meat versus plant-eaters) in a subsequent sorting task.

A secondary finding was of greater significance. Storm's eight year olds readily reproduced a structure that they had recently been exposed to, but without such exposure their behaviour indicated underlying dimensions for animal terms similar to those that have emerged from adults in MDS studies. Subjects who had received either random training or no training at all sorted into categories which appeared to reflect size, ferocity and habitat. Size and ferocity have emerged from all MDS studies with adults. Whenever another dimension has occurred, it has either been hard to interpret and best described as 'nearness to man' (Henley, 1969) or it has clearly reflected habitat (Coltheart, Hale and Walsh, 1985).

2.2.4 Anglin's Work on the Growth of Word Meaning

The concrete-abstract progression was directly demonstrated by Anglin's studies of conceptual structures in childhood (Anglin, 1970; 1977; 1978). Being primarily concerned with the development of abstract organisation, he selected stimuli specifically to explore this dimension of meaning.

The studies reported in The Growth of Word Meaning (Anglin, 1970) compared children's and adults' behaviour with a stimulus set consisting of six nouns, five verbs, five adjectives and four prepositions. Within each part of speech items were drawn from a hierarchy that varied in concreteness.

For example, the nouns - BOY, GIRL, HORSE, FLOWER, CHAIR and IDEA - were instances of noun categories arranged in a hierarchy. At the bottom and concrete end of the hierarchy was the category CHILD, from which the items BOY and GIRL were selected; the next category was ANIMAL, which was represented by the item HORSE and included BOY and GIRL; the next category was BEING, which permitted FLOWER to join the stimuli; next came OBJECT, exemplified by CHAIR; and finally, the most abstract category was ENTITY, which included all the previous stimuli and gave rise to the last noun, IDEA. It is important to note that although Anglin referred to the hierarchy as a concrete-abstract progression, he intended this to describe only class-inclusion relations in the sense that superordinates increase in 'abstractness', rather than the words themselves.

Anglin hypothesised that as semantic development proceeds, children are able to appreciate increasingly abstract relations between words. He therefore predicted that adults and children would differ in the categories they use in a sorting task involving the twenty words. The results of a hierarchical clustering analysis (Johnson, 1968) supported this view in that adults sorted according to part-of-speech while children tended to prefer thematic groupings, such as FLOWER-GROW. Adults also tended to produce larger categories.

More interesting, however, were the results of Anglin's final experiment in which subjects were given pairs of words and asked to 'write about each pair what both words share in common, what makes them similar in meaning'. The stimuli consisted of the original twenty words, paired within part-of-speech groups to form noun pairs, verb pairs and so on. In analysing the host of relations produced, Anglin excluded thematic responses such as SIT for BOY-CHAIR, classifying the remaining predicates into eight groups: BOTH ARE (OR DO).; BOTH ARE KINDS OF ...; BOTH DESCRIBE ...; BOTH ARE PERFORMED BY ...; BOTH IMPLY ...; BOTH PRODUCE ...; BOTH ARE SIGNS OF ...; BOTH HAVE ... (p.90).

The major finding was that the number of responses rose monotonically with age and that as the abstractness of the relation (or distance in the stimulus selection hierarchy) between members of a pair increased, children became less likely to produce acceptable responses. For example, nine and ten year old children produced more responses to BOY-GIRD and to BOY-HORSE than to BOY-CHAIR and BOY-IDEA.

One possibility was that some children may not have known the meanings of words like IDEA, but although not tested directly, this was unlikely. First, in a previous experiment, Anglin demonstrated that children of comparable ages were able to put the words into appropriate sentences when given a choice of four sentences with blanks. This was not a conclusive test of the subjects' understanding of the words because in order to make a correct choice, subjects need only have been aware of the part of speech of the work; but there were no indications that the children may not have understood the words in a wider sense.

Second, the words are all intuitively within the vocabulary of nine and ten year old children, IDEA being perhaps the 'hardest' word in the set.

A more plausible hypothesis was that children's definitions of words may be primarily concrete and thematic, such that abstract properties are either absent or less readily available for retrieval. One possible relation between BOY and IDEA is that BOTH ARE...produced by people. In order to arrive at this response, a subject would have to retrieve the facts that a BOY is produced by parents and that an IDEA is something that people generate. While a child may readily know that a person can have an IDEA, his definition of BOY may be dominated by concrete features such as SMALL; HAS SHORT HAIR; WEARS SHORT TROUSERS, so that in searching his lexical entry for BOY, he is unlikely to retrieve the more abstract fact that a person can have a BOY.

This is not to say that young children do not possess some version of these abstract facts, but it may be that their semantic memories are not organised to permit the ready retrieval of such information in simple production tasks. Anglin summarised his finding thus: "A child does almost as well as an adult at generating predicates for words bound by a concrete relation; a child does not do nearly so well as an adult when the relation is abstract".

2.2.5 The Semantic Differential in Conceptual Development

Following Osgood, Suci and Tannenbaums' (1957) influential work on the structure of connotative meaning, several studies focused on the development of the Evaluative, Potency and Activity dimensions which Osgood et al described. The use of factor analytic techniques for representing relationships between words meant that these developmental studies were similar in approach to the current research.

Divesta and Stauber (1971) showed four year old children a clown's face with one eye open and the other closed, describing the open eye as the GOOD eye and the closed eye as the BAD eye. Nouns representing extremes of the Evaluative dimension as rated by seven and eight year olds (Divesta, 1966) were then presented and the children were asked to press the eye corresponding to the meaning of the word. For example, they were asked 'Do you think AMERICA is GOOD or BAD?

The clown's eyes were then identified as STILL and MOVING and, finally, as WEAK and STRONG for testing judgements of Activity and Potency terms respectively. Judgements of Evaluative terms were significantly more accurate than those for Potency and Activity terms, although judgements for all terms were more accurate than one would expect on the basis of chance alone. Another indication of the salience of these factors to young children came from Ervin and Foster (1960) who found that six year olds used Evaluative terms synonymously, for example, using GOOD as if it meant the same as CLEAN. The same applied to Potency terms such as BIG and STRONG.

Of greater relevance to the current study were investigations of the structure of attribute spaces. Divesta (1966) conducted three studies in which hundreds of children were asked to rate concepts such as ME, MUSIC and ENEMY on a number of bi-polar adjectives. Principal components analyses followed by varimax rotations of the data sets indicated that Evaluative, Potency and Activity factors accounted for about 70%of the variance and that contrary to his hypothesis that the Evaluative dimension would decrease in importance with age these proportions remained constant for groups aged seven to twelve.

Saltz, Dunin-Markiewicz and Rourke (1975) noted that Divesta felt that some of the younger children may not have known the meanings of all the concepts rated. This would have contributed to noise in the data. In addition, they suggested that a technique for rating scales against scales is more appropriate than Divesta's indirect concepts on scales approach which is dependent on the distribution of the variance for each attribute across the concepts. For example, if all concepts are rated as GOOD, the result is zero variance on this scale and correlations with similarly rated evaluative scales would collapse to zero.

Saltz et al (1975) asked subjects aged six to college age to judge pictures of fictitious animals on sixteen bipolar scales drawn from studies using the semantic differential technique (Snider and Osgood, 1969). Subjects were shown a pair of fictitious animals, drawn so that one animal was the mirror image of the other and told that one animal was, for example, CRUEL and the other was KIND. How, then, did the remaining scales apply to the animals? If one animal was CRUEL and the other KIND, was one animal FRIENDLIER than the other and if so, which one? Or were they the same?

CANDECOMP (Carrol and Chang 1970), an MDS procedure which was incorporated into the later INDSCAL model, was used to determine semantic structure at each of four age levels. The dimensionality of the attribute space was shown to develop from two dimensions at six years of age, through three and four dimensions at nine and twelve years respectively, to five dimensions at college age. That is, a single 5-dimensional solution was found to provide a good fit for all age groups; and although, the authors did not report any test of the differences between subject groups in the weights attached to each dimension, the number of dimensions contributing to the variance in the data was thought to increase with age. The dimensions from I to V were labelled Evaluative; Potency; Activity; Inner and Outer Qualities; and Extraversion. All but two scales had high loadings on dimension I, which was taken as evidence of higher intercorrelation of scales amongst the young.

A drawback of this study, as with many others using similar methods, was the subjective manner in which the solution and interpretation were derived. The decision to attribute a given dimensionality to each age group was based on an examination of the amount of variance in the data of each group that was explained by each dimension in turn. In situations like this it is difficult to defend any particular choice of cut-off point below which dimensions are considered unimportant.

The main finding from Saltz et al's study was that attributes tend to be correlated in young children and that this produces 'a large general dimension (essentially evaluative) which tends to weaken with age.' (pp 913). As evidence of attribute correlation amongst the young, Saltz et al noted that the absolute values of the indices of co-relation between scales which formed the input to CANDECOMP decreased from the youngest to the oldest subjects. In addition, high proportions of the six year olds rated an animal as BIGGER when it was described FRIENDLY, TAME, CAREFUL, ACTIVE, CLEAN or BRAVE and it was found that younger children were generally more likely to state that two scales were related. Correlation was also more evident in boys than in girls and it is possible that this reflected the early verbal superiority of girls. Intercorrelation of scales also produced the first dimension which was extremely important to the six and nine year olds and in which almost all scales had high loadings.

There appeared to be the two factors underlying this effect. First, it was plausible that, according to some general process of evaluation, strong and active terms are considered good while weak and inactive terms are thought to be bad. The loadings of scales on dimension I were all high with the exception of YOUNG/OLD and BIG/SMALL, which of the set of sixteen scales, displayed the most tenuous links with a general concept of evaluation. Second, the youngest children appeared to be using their everyday experience of the world to provide thematic rationales for correlating scales. Saltz et al reported that one child said that a TAME animal would be CLEAN because if it were TAME someone would look after it and keep it CLEAN.

A further finding of this study was that additional dimensions were needed to account for the judgements of older children and adutls. Examination of the pattern of loadings of scales across dimensions also suggested that there may be a systematic process whereby scales become uncorrelated to form these dimensions. Saltz et al's discussion was in terms of semantic differential dimensions, inner-outer qualities and extraversion, but an alternative framework for this process may be more germane to an understanding of semantic development.

While dimension I displayed high positive loadings for all scales, dimension II showed high positive loadings for STRONG/WEAK, BRAVE/COWARDLY and HARDOWRKING/LAZY in contrast with high negative loadings for KIND/CRUEL, FRIENDLY/UNFRIENDLY, BEAUTIFUL/UGLY and YOUNG/OLD. Thus the scales that normally characterise an Evaluative dimension such as KIND/CRUEL were still strongly represented by dimension II, but were contrasted with the equally represented but negatively correlated Potency scales such as STRONG/WEAK.

Dimension III was a clear representation of Activity in that the

Evaluative and Potency scales had low loadings and were, therefore, independent of this dimension. Dimensions IV and V showed high positive and negative loadings for scales which had all positively loaded on dimension I. For example, dimension IV displayed a high positive loading for KIND/CRUEL and a high negative loading for BEAUTIFUL/UGLY, suggesting that personality can contrast with physical appearance. Dimension V had, for example, a high positive loading for FRIENDLY/UNFRIENDLY and a high negative loading for SMART/SILLY, indicating a contrast of two abstract characteristics which could be termed personality and competence.

This interpretation of the developmental trend in Saltz et al's attribute space is consistent with the concrete abstract progression. Although the first differentiation to occur was described in terms of Potency (dimension II), the key factor may be that this type of discrimination may be achieved from a comparision of concrete and perceptual attributes such as size and shape. The Activity dimension is based on discrimination of overt behaviour involving movement and more abstract characteristics associated with and deriving from movement. Dimension IV saw the first appearance of a pattern indicating discrimination based on perception of physical qualities in contrast with more abstract personality attributes. Finally, dimension V appeared to indicate discriminations based on entirely abstract attributes which may be termed personality and intelligence or competence.

2.2.6 More Recent Findings on Taxonomy and Abstractness

As shown in section 2.2.3, young children demonstrate an awareness of taxonomic relations when asked, but do not tend to generate them spontaneously (Smiley and Brown, 1979). Also at an anecdotal level, it is clear that children have some understanding of abstract words like 'nice' and 'nasty', although their definitions may not be identical to those of adults. The shift from themtic to taxonomic organisation is therefore more properly descibed as a preference.

Markman and Hutchinson (1984) demonstrated that while two to three year old children prefer thematic organisation, they can nevertheless switch to construing material taxonomically when this is of advantage to them. Children were asked to find a picture that was 'the same' as a standard, given the choice of a thematic or taxonomic associate. In the No Word condition the standard was not named, but in the Novel Word condition it was given a nonsense name like DAX. In the No Word condition children chose thematically, but when children thought they were learning a new word they chose the taxonomic alternative. This was interpreted in terms of a sensitivity to constraints on word meaning, which allows children to switch from an ordinarily useful thematic view of the world to a taxonomic strategy which is efficient for learning new words.

Keil (1979) developed a theory about the structure and development of ontological knowledge, carrying out some studies which, although superficially similar to Anglin's, were contradictory in outcome . Ontological concepts define basic categories of existence, representing understanding of what fundamental kinds of things exist and how they relate to each other. The largest category is 'all things'; this includes 'things with spatial location' which includes 'physical objects' and so on. At each level there are instances. For example, an 'idea' is a 'thing'; a 'thunderstorm' is a 'thing with spatial location'; and 'milk' is a 'physical object'. Ontological classes form a hierarchy of abstractness which is the same as that studied by Anglin, whose categories were ENTITY, OBJECT and so on down to CHILD.

Keil's work was based on philosophical studies of ontological classes focusing on the extent to which a predicate can be sensibly applied to a term. For example, 'the door is brown' makes sense, as does 'the child ran across the room', but 'the door ran across the room' is anomalous. The anomaly is a 'category mistake' rather than a 'falsehood' because it is inconceivable except in a deliberate act of fantasy. This is one way of identifying that door and child come from different ontological classes. In a study with children aged five, seven, nine and eleven, Keil asked the children if various sentences 'were OK'. The children's recognition of 'OK' versus 'silly' sentences indicated their understanding of properties of various ontological classes. While Anglin's methods showed that older children did not generate either properties of abstract classes or the classes themselves, Keil's anomaly tests demonstrated that seven year olds distinguish the properties of animals, plants, inanimate objects and non-physical things.

These findings supported the view that young children express a preference for thematic organisation, but are capable of taxonomic judgements in certain situations. If, as Markman and Hutchinson showed, toddlers can make a taxonomic discrimination, the ability may exist at an even earlier age and underly some of the contradictory findings on form and function in infant studies (Markman, 1989). The significance of understanding taxonomic relations at an early age is discussed in Chapter 10, along with the findings of the studies reported in this thesis.

2.3 MDS and Semantic Development

The dominant characteristic of semantic development to emerge from this review was the concrete-abstract progression. Related to this was the change from thematic to taxonomic organisation. Also, studies using factor analysis indicated that adults discriminate amongst concepts using more dimensions than children. It was important to choose a model of semantic development that could encompass all of these phenomena.

Individual difference MDS models can represent semantic relations in a space where the dimensions reflect the ways in which individuals differ most. Subjects can weight the dimensions differentially and weights on dimensions can approach zero, thereby allowing some subjects to utilise more dimensions than others. If dimensions can be identified, they can be assessed with respect to concreteness and the dimensional weights will indicate if some subjects' representations are more concrete than others. Since the model also provides each individual's semantic distances between items, it allows individuals to be compared in terms of the strength of taxonomic clustering.

CHAPTER 3

SPATIAL MODELS OF SEMANTIC MEMORY

3.1 Structural Models of Semantic Memory

Much research in semantic memory has been concerned with the validity of structural models. Investigators have used various methods for revealing structure in data, the general aim being to reduce complex data to a small number of interpretable entities. The major techniques are cluster analysis, factor analysis and multidimensional scaling (MDS).

Cluster analysis was used by Miller (1969) to examine subjects' sorting behaviour with a set of disparate nouns. The raw data were the number of subjects putting a pair of items into the same group. Osgood, Suci and Tannenbaum (1957) employed factor analytic techniques to explore the components of affective meaning. Data for this type of analysis are more complex, consisting usually of concepts rated on bi-polar scales.

Shepard (1974) combined clustering with MDS in order to model the relations amongst a set of animal terms. Shepard's data were associative strengths between animals derived from their relative positions in lists that subjects had generated; but any measure of proximity between objects would suffice for this type of analysis. Kruskal (1977) also described the complementary nature of scaling and clustering procedures.

Of the three techniques, MDS has featured most frequently in semantic memory research (for example, Henley, 1969; Rips, Shoben and Smith, 1973; Caramazza, Hersh and Torgerson, 1976; Hutchinson and Lockhead, 1977) and is the topic of this chapter.

3.2 Multidimensional Scaling

MDS analysis represents relations amongst a set of objects by distances between the objects in n-dimensional space. Using iterative procedures on a measure of proximity between objects, it constructs a configuration in which the distances between objects are maximally congruent with the original proximity estimates. In semantic memory research subjects are generally asked to provide judgements of similarity for the members of all possible pairs of a set of words. Such data are then submitted to a scaling analysis which produces a semantic space in which the distances between words reflect judged semantic relatedness.

MDS models are eithter nonmetric or metric. Nonmetric models seek to represent the rank ordering of proximity in the data, while metric models reflect the proximities themselves. Weeks and Bentler (1979) compared these methods with data from random configurations, finding that when assumptions of linearity are met, the metric model performs best, but when assumptions of linearity are not met, the two methods perform equally well.

Models designed to represent individual differences constituted a major development in scaling methodology. Originally, MDS models analysed only the relations amongst a set of objects and were known as 2-way models because the data consisted of measures of the proximity between pairs (the two ways) of objects. In semantic memory research this describes the judgements of one subject or the average judgements of a group of subjects. Individual difference or 3-way models, such as INDSCAL (Carroll and Chang, 1970) permit the analysis to include the effect of replication, which in most psychological studies, refers to judgements from different subjects.

MDS has been widely used, both in psychology and in other disciplines, to reveal hidden structure in data. An example in social psychology is the study by Nygren and Jones (1977) on political perceptions, whilst the most famous application is probably the work by Kendall (1971) in which the locations of long-lost parishes were recovered from records of marriages. Kendall's thesis was that distances between parishes would be reflected in the frequency with which inter-parish marriages took place and, using MDS to analysis these data, he was able to predict the locations of ancient parishes.

MDS has also been used to represent comprehension of prose passages. Subject performed similarity ratings on nouns before and after reading a passage containing the nouns, while under instructions to carry out the second set of ratings with respect to similarity in the passage. The 'before' and 'after' scaling solutions demonstrated the influence of the passage (LaPorte and Voss, 1979) and same-different reaction time judgements for the nouns were also related to distances in the MDS space (Bisanz, LaPorte, Vesonder and Voss, 1978).

3.3 MDS Models of Semantic Memory

There are two major reasons why MDS is an attractive tool for the study of semantic structures. First, semantic relations are inherently amenable to distance representation, given that a major concept in current semantic theory is that of semantic distance. With the context of MDS studies, Rips et al (1973) employed semantic distance to explain the subset (or category size) effect in reaction times: subjects are faster in deciding that A CANARY IS A BIRD than that A CANARY IS AN ANIMAL (Collins and Quillian, 1969). The implication here is that CANARY is 'closer' to BIRD than to ANIMAL in some underlying spatial semantic structure (Rips et al, 1973).

The second advantage of a scaling approach to semantic modelling is that it allows subjects' perceptions of word meaning to emerge largely unconstrained by the experimenter' ideas, so providing a uniquely realistic measure of semantic relations. The alternative is to determine semantic distance by reference to some common-sense scheme or widely held taxonomy.

For example, the model developed by Collins and Quillian (1969) assumed that any given instance of a mammal, such as HORSE,would be closer to the term MAMMAL than to the term ANIMAL, in accordance with the Linnaean system of biological classification. If this structure were valid, speeded verification tasks would reveal the subset effect in which an instance's membership of a subset is verified faster than its membership of a superordinate set. Collins and Quillian were aware at the outset that some 'hierarchies are not always clearly ordered' (p242) and that this could create difficulties for their model. They avoided such heirarchies. As expected, a 'reverse subset' effect was eventually found by Collins and Quillian and others for heirarchies containing MAMMAL, subjects taking longer to verify instances as mammals than to verify them as animals (Rips, Shoben and Smith, 1973).

One possibility was that this was due to frequency, but Rips et al (1973) disagreed, noting that like MAMMAL, BIRD is also less frequent than ANIMAL, but does not give rise to a reverse subset effect. Their study provided a scaling solution in which all but three of the mammals were closer to ANIMAL than to MAMMAL, thus suggesting that the effect arose from the discrepancy between the experimenters' determination of semantic distance and that of their subjects'. McCloskey's (1980) later work on the role of stimulus familiarity provided another explanation: MAMMAL may be less familiar than ANIMAL, where familiarity influences the time needed to comprehend a word. However, even without evidence of this kind it would be reasonable to suppose that peoples' conceptions of common nouns may differ from those inherent in some external logical structure and, furthermore, that there may be important individual differences in conceptual structure.

Research has focused on the predictive value of subjective structures obtained from scaling and other structural analyses. In general, findings indicate that MDS models can predict certain aspects of performance when subjects retrieve information from the modelled semantic domain. In a review of reaction time studies Hutchinson and Lockhead (1977) suggested that a spatial model "based on the similarity of word meaning is sufficient to account for most of the present data relating reaction time to semantic memory (p664)". Other studies have demonstrated that the predictive capacity of scaling solutions extends to aspects of free recall and reasoning. For example, Rumelhart and Abrahamson (1973) were able to predict subjects' choices in an analogical reasoning task from Henley's (1969) configuration of mammal terms and Caramazza, Hersh and Torgerson (1976) found that their scaling solution predicted proximity of recall for animal terms.

Researchers in semantic memory have displayed a distinct preference for modelling the domain of animal terms and dimensions apparently reflecting size and ferocity have emerged consistently from studies investigating adults (Henley, 1969; Rips et al, 1973; Caramazza et al, 1976). However, these dimensions have not been convincingly demonstrated and other less interpretable dimensions have been reported. In addition, some investigators appear to have been less rigourous than others in their application of scaling methodology and, consequently, some of the models reported maybe sub-optimal.

3.4 A Review of the Major Studies

3.4.1 Henley (1969)

Henley (1969) obtained dissimilarity ratings for thirty mammals on an 11-point scale from twenty-one adults. The unfortunate subjects had to perform the task of rating the 435 possible pairs of mammals twice to provide reliability correlations and to check that ratings were unaffected by reversing the within-pair order. Having examined individual differences via hierarchical clustering, Henley created a mean dissimilarities matrix based on judgements from eighteen subjects and submitted this to TORSCA, a non-metric 2-way model (Young and Torgerson, 1967).

The badness-of-fit measure in TORSCA is 'stress' (Kruskal, 1964), later defined as 'stress 1' to distinguish it from a modified version, 'stress 2' (Davison, 1983). Henley reported stress for 4-, 3- and 2-dimensional solutions at .068, .094 and .154 respectively. At the time there was little guidance on the interpretation of stress. Kruskal (1964) had suggested that stress less than .01 indicated an 'excellent' solution; values between .01 and .05 were 'good'; values between .05 and .10 were 'fair'; values between .10 and .15 were 'moderate'; and stress above .15 indicated a 'poor' fit. Kruskal and Wish (1978) later provided more detailed guidance, particularly in relation to the number of objects scaled which can dramatically affect the interpretation of stress. For example, they pointed out that although stress of .02 is normally 'good', if seven objects are scaled in three dimensions, stress of .02 occurs "for contentless random data about 50% of the time" (p52). Therefore for such a small number of objects, stress of .02 is not small enough to indicate a good fit.

As Henley scaled thirty mammals, her stress value of .094 at three dimensions was 'moderate' and this may have influenced her choice of that solution, although she did not explicitly refer to Kruskal's criteria. Her choice was also probably also influenced by the interpretability of the solution. Comparisons of animals at opposite extremes of each dimension indicated that dimension I was size and dimension II was mildness versus ferocity. Dimension III was 'harder to label', but because the primates were at one end it was considered 'resemblance or relatedness to man or something like that' (p180). Henley also noted that the mammals were roughly grouped into four categories: small and wild: large, hooved herbivores; primates; and carnivores. The exceptions were PIG and ELEPHANT.

Henley's overall aim was to compare techniques which are commonly used to study semantic fields, therefore she obtained data about semantic associations between the animals from the same subjects via a range of methods, of which the dissimilarities task was only one. Dissimilarities correlated well with similarities derived from triadic ratings and with a proximity measure based on the number of animal associates that mammals shared. Dissimilarities correlated less well with inter-item distances derived from subjects' lists of animals.

When subsets of the mammals were scaled using data from triadic ratings, the solution for six items showed a ferocity dimension and an uninterpretable dimension, while the solution for twelve items contained both size and ferocity dimensions, plus another uninterpretable dimension. There were high correlations between data produced for the various subsets, indicating that the context in which rating took place did not affect the relationships among mammals.

A problem with Henley's interpretation is that although sensible, it was obtained from a scaling method in which the axes are arbitrarily positioned and need not reflect psychological dimensions. Her method of comparing items with extreme locations is appropriate (Kruskal, 1977), but rotation can alter the positions of animals on dimensions radically. For example, in a 2-dimensional configuration, an object with extreme scores on both dimensions will have a moderate score on one dimension after a 45' rotation, therefore while it would enter into the interpretation of both dimensions in the first case, it would only be relevant to interpreting one dimension after rotation.

3.4.2 Rips, Shoben and Smith (1973)

Rips et al (1973) obtained ratings of 'relatedness' on a 4-point scale for twelve birds and twelve mammals separately. The subjects were 24 undergraduates. The task entailed rating a standard item against the remaining items, therefore subjects could have adapted to the standard. Another unusual feature was the inclusion of ANIMAL, and BIRD or MAMMAL in each set. Using INDSCAL to analyse birds and mammals separately, they concluded that two dimensions were adequate for both sets. Dimension I of the bird solution was size, while dimension II contrasted predatory birds with game birds. In the mammal configuration, dimension I was again size, while dimension II separated predatory mammals from farm mammals. However, predacity in the mammal solution was not convincing because MOUSE was almost as extreme as LION. The bird and mammal solutions explained 46% and 42% of variance in the original data respectively.

Rips et al had previously obtained sentence verification latencies from different subjects for instance-category pairs such as 'A robin is a bird'. The true latencies correlated reasonably well with the instance-category relatedness ratings submitted to INDSCAL (r=-.73 for birds and r=-.52 for mammals) but their correlations with inter-item distance derived from INDSCAL were close to zero, implying that INDSCAL had provided a poor solution. The items were re-scaled in three dimensions which improved the fit for birds to 56% and that for mammals to 51%, but not the correlations between latencies and derived distances. Solutions in higher dimensionalities would have resulted in better fit, but were reported as uninterpretable.

Rips et al attributed this poor outcome to the fact that only a few distances in the scaling solutions - the instance-category pairs - were used. Therefore another experiment was carried out in which same-different judgements for instance-instance pairs were correlated with a second set of INDSCAL solutions. These new bird and mammal solutions were obtained by scaling a subset of the ratings collected earlier; they contained six instances, ANIMAL and BIRD or MAMMAL. The 2-dimensional configurations explained 68% and 59% of the bird and mammal variance respectively; and the same dimensions were evident. This time true latencies were predicted by the solutions, the best predictors being the instance-category distance relating to the first instance in the pair. Little or no improvement resulted when the other instance-category distances and the instance-instance distances were added to the equation.

Although these results were promising, the use of scaling methodology presented certain problems. A 4-point scale is unusually short and no check on subjects' reliablility in the rating task was reported. The variances accounted for by INDSCAL solutions of increasing dimensionality were not systematically compared to identify the optimum solution and the selected configurations were not convincingly interpreted. Furthermore, there is evidence that the inclusion of superordinate items creates difficulties for subjects in a scaling task because judging the similarity of instance-category pairs is awkward in the context of instance-instance pairs (Coltheart, Hale and Walsh, 1985). Medin and Smith (1981), referring to Tversky's work,also felt that superordinates should be excluded because they are rated as most similar to all instances and the resulting distortion in data requires a solution of very high dimensionally. Finally, although the same dimensions were said to have appeared in solutions of twelve item and six item subsets the scaling of a subset of data is likely to produce results which differ from the scaling of independently collected data for a subset of items.

3.4.3 Caramazza, Hersh and Torgerson (1976)

Caramazza et al (1976) asked whether a single scaling model could deal with items from different categories, namely mammals, birds and fish. Thirty animals, ten from each category were rated on a 9-point scale by fifteen students. A form of principal components analysis was used to examine individual differences before the data were collapsed across subjects to form a single matrix. Like Henley, they used TORSCA, reporting stress of .013 at four dimensions.

The solution was rotated to maximise category membership in the plane formed by dimensions I and II. Dimensions III and IV were then interpreted as a size-ferocity plane although these attributes were at 45' to the actual dimensions. This interpretation was not entirely convincing as ROBIN and CANARY appeared to lie between LION and ELEPHANT on the size dimension. Other features of the solution were that MOUSE and RABBIT were far from all other items and the mammal distances were larger and more variable than other within category distances.

Latencies for same-different judgements were then obtained from different subjects and correlated with both rated dissimilarity and inter-item distances derived from TORSCA. The correlations between same judgement latencies and rated dissimilarity were .48, .24 and .61 for mammals, birds and fish respectively; and the correlation between different judgements and rated similarity was -.17. Similar correlations were obtained between latencies and derived distances.

Another group of subjects were tested on free recall of the thirty items and inter-item distances in the recall lists were correlated with distances in the scaling solution. These correlations were .55 for mammals, .394 for birds and .163 for fish.

Caramazza et al concluded that subjects operate with comparable subjective structures across categories and that although semantic distance in scaling solutions can be a powerful predictor of performance in latency and recall tasks, prediction varies between categories, sometimes accounting for a very small proportion of the variance. Some of this variability could plausibly be attributed to the material as people know much more about mammals than they do about fish.

3.4.4 The Scaling Methods: Scope for Improvements

Consideration of these three studies as a whole supported the conclusions drawn by Caramazza et al, but suggested that the adequacy of the scaling solutions themselves might be an additional important factor. If a scaling solution is a poor model of semantic memory distances deriving from it are unlikely to predict cognitive performance. There were several problems with the applications of scaling methods in these studies. Minor issues related to subject reliability and the treatment of individual differences, while the major issues concerned the use of MDS itself.

Henley was the only one to report reliability correlations ranging from .29 to .85. Also, she rejected three subjects because they were idiosyncratic in relation to others, but whether these subjects were also the least reliable was unclear. Henley then scaled a mean dissimilarities matrix.

Rips et al did not report tests of subject reliability, but they were unique in performing individual differences scaling analyses. However, they reported no examination of individual differences and used the solutions to predict performance in different subjects. Caramazza et al scaled mean dissimilarities without rejecting idiosyncratic or unreliable subjects; instead they accepted that 30% of the variance in their data might be attributable to stable individual differences and to subject unreliability.

Finally, Henley correlated distances in her scaling solution with other judgements from the same subjects, but in the other studies scaling solutions from one set of subjects were used to predict performance in different subjects. These observations indicated, first, that subject reliability may have contributed to 'noise' in the data and, second, that the role of individual differences in the predictive power of scaling solutions had not been explored.

Turning to the application of MDS, a general point is that scaling analyses do not usually produce unique solutions. The scores of items on dimensions result from an iterative procedure in which distances in the data are correlated with distances in successive test solutions, the process terminating when a pre-set criterion of fit is attained. The final solution should represent a minimum point in the fit function from which any further iteration worsens fit, but sometimes a local rather than global minimum is reached.

INDSCAL users are generally advised to run the analysis several times from different random starts and choose a solution from amongst a group of highly similar solutions, on the basis that these are likely to represent the global minimum (Pruzansky, pc). TORSCA has a rational starting configuration, but this is no guarantee of a unique result. In pilot studies in the current study it was found to produce different results depending on the starting dimensionality. That is, a 3-dimensional solution in a run starting at four dimensions will differ from a 3-dimensional solution in a run starting at three dimensions. This can also occur in INDSCAL analyses (Schiffman, Reynolds and Young, 1981). The difficulty with the scaling studies of animal terms is that none of them contain details of how this issue was tackled.

In addition, there are questions about the investigators' use of goodness of fit measures, their interpretation of the dimensions and their methods for determining the optimum number of dimensions. Henley and Rips et al reported fit measures for solutions in three and two dimensionalities respectively, but Caramazza et al reported only the fit of their chosen dimensionality. Therefore, with the possible exception of Henley's study, systematic comparisons of fit at different dimensionalities were not reported and may not have taken place. This implies that identification of the optimum number of dimensions was a largely subjective process.

The interpretation of dimensions was probably also a factor in determining dimensionality, but not consistently so, for while Rips et al rejected high dimensions as uninterpretable, Henley accepted 'ineffable' dimensions. Common to all studies, however, was an apparent reliance on subjective interpretation based on visual inspection.

Taken together, these observations indicated that optimum solutions may not have been achieved. Determining dimensionality is a central issue in scaling analysis and all three studies may be criticised for their failure to employ objective techniques (Coltheart, Hale and Walsh, 1985).

3.5 An Objective Approach to Scaling Studies

3.5.1 Identifying the Best Configuration

Determining the dimensionality of a solution is a complex task and there is a number of techniques designed to reduce reliance on intuition (Coxon and Jones, 1979b, Schiffman et al, 1981). The primary consideration are goodness-if fit and interpretability, but since an increase in dimensionality will always lead to an increase in fit it is important to be able to detect the point at which dimensionality is optimum, such that contribution of additional dimensions is negligible.

As a first step the relation between dimensionality and fit should be examined in order to establish whether or not there is a clear discontinuity or 'elbow' in the curve. If solutions are run for a range of dimensions, each solution in n dimensions will account for a greater proportion of the variance in the data than a solution in (n-1) dimensions, but for successive dimensionalities, the gain in fit diminishes and sometimes the resulting curve will display an 'elbow' that is sufficiently marked to indicate the dimensionality beyond which increases in fit become negligible. In reality, such curves rarely possess clear elbows and are more likely to indicate the range within which the true dimensionality lies.

A more comprehensive technique is the use of property fitting (Carrol & Chang, 1970; Wish, Deutsch and Biener, 1972). This is a multiple regression procedure which assesses the fit between specific properties and the configuration, while at the same time indicating which properties offer the best interpretations for each dimension. Data must first be collected indicating the value of each object on each potential property. In many cases these are subjects' ratings on unidimensional scales, but they can be more objective characteristics, such as molecular weight applied to a configuration of chemical tastes (Schiffman et al, 1981, p280). If the multidimensional configuration is an accurate representation of the relations between objects, it should be possible to locate lines (or vectors), each representing the object's scores on a salient property, somewhere in the multidimensional space (Kruskal and Wish, 1978, p36).

In statistical terms, this means that a stepwise regression analysis, in which the dimensions are independent variables and the unidimensional rating is the dependent variable, will result in a high multiple R. If the vector is highly correlated with one dimension to the exclusion of others, it may also be used as a label for that dimension. The chief value of property-fitting is as an aid to interpretation, but it can also indicate optimum dimensionality: as with internal goodness-of-fit measures there will be a point beyond which an additional dimension fails to significantly increase the fit between potential labels and the solution.

A sophisticated approach for identifying dimensionality involves comparing the obtained solution with solutions from a Monte Carlo (simulation) study which used data of known dimensionality and known levels of error. This procedure, known as M-space (Spence and Graef, 1974), provides an estimate of the best dimensionality of an obtained solution by comparing its goodness-of-fit values with those resulting from the analysis of fictitious data. In principle this approach could be applied to the whole range of scaling techniques, but in practice, simulation data have been published only for 2-way scaling and the goodness-of-fit measure known as 'stress'. Coxon and Jones (1979b) noted that basic simulation results did not exist for individual differences scaling and were unlikely to become available.

If solutions defy interpretation, Coxon and Jones (1978; 1979) suggested that it is possible to identify 'troublesome' subjects or objects whose removal from the analysis would improve the data. A subject or object would be considered 'troublesome' if, when different dimensionalities are compared, it appears to move around the space in a dramatic manner. This may mean that the judgements associated with that object or person are inconsistent relative to the remaining data.

3.5.2 The Implications for Modelling Semantic Memory

As none of these techniques has been employed in semantic memory research, it is possible that published results are sub-optimal and that the predictive capacity of scaling models of semantic structure is greater than findings so far suggest. A further improvement in the predictive power of MDS models should also result from the incorporation of individual differences. While one strategy would be to submit each subject's data to a separate scaling analysis, a more coherent and economical approach would be to use an individual difference model.

One such widely used model is INDSCAL (Carroll and Chang, 1970), which represents both subjects and objects in the same multidimensional space. INDSCAL assumes that all subjects share the same cognitive structure, but it allows them to weight the dimensions idiosyncratically. Furthermore, since the dimensions are located to maximise inter-subject variation in the sense of principal components, they are generally found to be interpretable.

An INDSCAL analysis in conjunction with property fitting should, therefore, provide an interpretable model of semantic space in which the important semantic features are represented by the dimensions. Subjects' weights on the dimensions would indicate the relative importance of various features. Finally, by applying an individual's weights to the dimensions, a private semantic space may be created which represents the relations between stimuli as perceived by that individual. This allows semantic distances between stimuli to be estimated for individuals. Such data may be examined for individual differences in clustering of stimuli and can also be related to the individual's handling of the stimuli in other cognitive tasks.

3.6 Using MDS with Children

3.6.1 The Difficulty of the Task

Since studies of linguistic development suggest that semantic structures may be age-dependent and, within semantic memory research, MDS has been widely used to explore semantic structures, it was initially surprising to find no direct comparisons of semantic structure in children and adults using MDS. This partly reflected the differing interests of researchers in cognitive development and semantic memory, but it may also have been due to the difficulty of obtaining the judgements required by an MDS analysis from children.

Three factors affect children's ability to produce suitable data. First, the number of judgements is more than young children can be expected to complete without considerable deterioration in concentration. Second, the easiest method of data collection is to obtain direct estimates of similarity on an interval scale and young children may be unable to appreciate the properties of such a scale. Third, below a certain age children may be incapable of behaving in the systematic manner required for scaling judgements.

Of these potential problems, the third is the easiest to solve, for if judgements are obtainable, an indication of subjects' ability to be systematic is their consistency over repetitions of items. Scaling tasks for both children and adults should contain a proportion of repeated items so that some form of reliability score can be computed for each subject and grossly inconsistent subjects discarded from the analysis.

Problems relating to quantity of judgements and to the difficulty of the judgement require more thought, as reducing the latter tends to lead to an increase in the former. The traditional solution to the problem of a numerical scale is to use the method of triads as discussed by Torgerson (1958). Unfortunately, however, this leads to a large increase in the number of judgements. The complete method of triads requires that each triad be presented three times. For n objects, n(n-I)(n-2)/6 different triads are generated; and the subject must make 3.n(n-I)(n-2)/6 judgements. As quite small numbers of stimuli can obviously generate hundreds of judgements it is clearly desirable to look into ways of reducing the demands this makes on subjects.

One approach is to split the experimental session into two or three shorter periods. For this to be effective in avoiding fatigue it may even be necessary to run subjects over several days. In a study concerning visual memory in children and adults, the children were given a total of 132 triads over two consecutive days, the adults completing all the triads in one session (Arabie, Kosslyn and Nelson, 1975). However, while this method undeniably eliminates some aspects of subject fatigue, its effect on the judgemental process is unclear. The implications for similarity judgements of protracted breaks between sessions are not generally known, although it is useful to note that Henley (1969) obtained a high correlation between two sets of ratings of the same 435 pairs of mammals produced by adults one week apart. However, in a developmental context, this would be an undesirable practice since subject's notions of the stimuli may become contaminated between sessions by the introduction of novel information; and even without such input, subject's own cognitive and inferential processes may cause their conceptual structures to be modified as a direct reslut of the processing demands of a previous judgemental task.

A second method of reducing demands of the task is to present subjects with a subset of the total number of judgements required. The use of incomplete designs is possible because it has been shown that a configuration may be recovered even when large sections of the input proximity matrix are missing (Spence and Domoney, 1974).

Levelt, van de Geer and Plomp (1966) used balanced subsets of thirty-five triads instead of the full 455 triads generated by their fifteen stimuli. The subsets were balanced in the sense of each stimulus appearing equally often. Similarly, Arabie et al (1975), in addition to splitting their task into separate sessions, used a balanced subset of sixty-six triads so avoiding the full 200 triads and 660 judgements necessitated by their twelve stimuli. The problem with this method is that however balanced the subset, if its size is arbitrarily decided, there is no guarantee that sufficient data will be collected to ensure the recovery of structure.

One non-arbitrary method of determing the minimum size of subset required is a Mone Carlo approach advocated by Spence and Domoney (1974). Their results apply, not to triads, but to paired comparisons for thirty-two to forty-eight stimuli. Using data of known structure and with known error, they examined the effect of various deletion patterns in the dissimilarity matrix on the ability of TORSCA-9, a 2-way method, to recover the data's structure. One finding, for example, is that up to 33% of a matrix for thirty-two stimuli may be deleted before recovery begins to decline. However, this merely reduces the task to 330 judgements, a number too great to contemplate asking of children, even if it could be assumed that they were capable of making dissimilarity ratings.

3.6.2 The Use of Rating Scales

Since even small numbers of stimuli generate untenable numbers of triads and neither split sessions nor subsets could be used with any confidence, it was decided that children's ability to make dissimilarity ratings for pairs of objects on an interval scale should be examined. As no direct evidence existed on the ability of adults or children to use such scales, the first task was to discover how children who have acquired basic literary and numerical skills behave when asked to assign numbers to their judgements. It is generally assumed that adults understand the implications of an interval scale.

In order to use an interval scale, a subject must be able to both classify and seriate. The assignment of numbers to objects is a classification task, while the ordered nature of the numbers involves seriation. An additional requirement of interval scaling is that subjects treat the intervals as equal, but this is often assumed and rarely tested. Whenever doubt arises it is usually possible to analyse the data using non-parametric procedures that do not require the assumption of equal intervals in the data. In this case, interval data was required by the model, but if this could not be achieved, Weeks and Bentler's (1979) demonstration that linear and monotone methods perform equally well with non-linear data indicated that the result would be no worse than if a monotone model had been attempted from the start.

Although dissimilarity scaling involves classification and seriation, subjects need not actually assign numbers to their judgements. Ramsey (1978) cited a study by Simard and Taylor (1976) who asked eleven year old children to draw a tick on a line. The line was 11 cm long and the ends were labelled VERY SIMILAR and VERY DIFFERENT respectively. They measured the distance of the ticks from one end of the line, a procedure which arguable results in interval measurement. This suggested that young subjects might be assisted by a technique which encourages them to use the distance analogy and imagery of a physical scale.

A further requirement of dissimilarity scaling is that subjects understand the meaning of word like SIMILAR and DISSIMILAR or ALIKE and DIFFERENT. Mansfield (1977) presented antonym and synonym pairs to four and five year olds, asking them to say whether the pairs were the SAME or DIFFERENT and to define the words. She found that "although the definitions were formally inadequate, it was apparent that the words were recognised and understood" (p61).

Meredith (1977) reviewed research on children's understanding of antonyms concluding that studies showing that young children treat the terms SAME and DIFFERENT as equivalent suffer from ambiguities in the questions asked and that other studies show that young children can differentiate these terms. As these debates generally relate to children under five and the current research was concerned with linguistically competent children of school age, inability to understand the poles of the scale was not anticipated. However, as a precaution children's and adults' understanding of such terms was tested in a pilot experiment. Following this, a test was devised to examine childrens' ability to classify, seriate and make dissimilarity judgements using both triads and a rating scale.

3.7 Devising a Dissimilarities Pre-test

Children's understanding of the terminology of similarity was initially explored by asking a small sample of children and adults to choose objects that were similar or dissimilar to a standard. The stimuli were red and blue circles of three sizes and there were two examples of each stimulus. Subjects were shown a standard stimulus selected by the experimenter and asked the following questions: (a) find one that is the same as the this; (b) find one that is a lot like this, but not exactly the same, that is, one that is very alike; (c) find one that is a little like this, but not very much; (d) find one that is very different from this. Subjects were asked to explain their choices.

Question (a) could only be answered correctly by choosing the identical stimulus, but each of the remaining questions possessed a number of potentially acceptable answers, depending on which stimulus was the standard. Answers were quantified by treating the change in the stimulus dimensions equivalently. It was apparent from the outset that the subjects - five children aged five to seven years and three adults - all understood the task. As their choices and accounts were very similar, indicating no difficulties with the terminology of dissimilarity judgements, a paper and pencil test was devised to test subjects' ability to make the various types of judgement involved in rating dissimilarity.

The test is shown in Appendix I. It asked children to classify words into subsets; to arrange words on a linear dimension of meaning; to decide which of two objects was most like a third (triads); and to make dissimilarity ratings on a 1 to 7 scale for pairs of items drawn from the category CLOTHING. Subjects completing the pre-test were asked to explain some of their decisions so that the reasonableness and rationality of the judgements could be assessed.

The test was administered to ten seven year olds from a local primary school. As eight of these children had no difficulty with the test, it was decided that the test could serve as a screening device to select children for a scaling study in which dissimilarity of items within pairs was judged on 7-point scale. The use of a rating scale meant that a reasonable number of objects could be scaled, while keeping the demands of the task within acceptable limits.

CHAPTER 4

A MULTIDIMENSIONAL SCALING MODEL OF SEMANTIC DEVELOPMENT: THE SEMANTIC DOMAIN AND THE TASK

4.1 Introduction and Objectives

Experiments I, II and III were undertaken to test the usefulness of a MDS approach to the study of development in semantic memory. In Experiment I a spatial model of semantic memory for animal terms in children and adults was generated and individual differences were examined. In Experiments II and III the model's ability to represent specific developmental effects was tested. The value of the excercise is that a scaling model has the potential to provide a single representation of individual differences in cognitive structure; and it can incorporate explanations in terms of both feature lists (the dimensions) and networks (the inter-item distances).

The experiments were designed to provide three tests of the spatial approach. First, an adequate scaling model should display individual differences with respect to age if these exist between children of different ages and between children and adults. Age differences in the number of dimensions and in their characteristics would be reflected in the relative importance of dimensions to each of the subject groups. Second, interpretation of the dimensions should indicate whether there is a shift from concrete to abstract organisation with age. Third, the relative locations of items in the space should indicate whether taxonomic organisation increases with age.

The domain of animal terms was selected because it has been explored in similar contexts using adult subjects. It was also considered to possess a clear taxonomic structure and to be a ready source of familiar items.

4.2 Experiment I: The Individual Differences Model

4.2.1 INDSCAL

MDS can represent semantic relations between words by distances in n-dimensional space. Individual differences are represented by allowing subjects to have different weights on a common set of dimensions. The individual difference scaling model, INDSCAL (Carroll and Chang, 1970) provides a single stimulus configuration in n dimensions and salience weights for each subject on each dimension. Analysis of variance and related techniques may then be used to examine group differences in salience weights.

4.2.2 Stimulus Selection

An important stage in the scaling of objects is the selection of objects to be scaled. This is because characteristics of the objects and their relation to one another set limits on the number and type of dimensions that may emerge from the analysis. Ideally the stimulus set should represent an everyday notion of the range of the concept under investigation; in reality items are chosen to reflect extreme and moderate levels of the dominant features of the concept. Thus stimulus selection involves a degree of informal scaling and should be based upon hypotheses about the structure of the concept, leaving subjects participating in the scaling task itself to display the extent to which they are aware of - or consider important - the attributes embodied in the set.

If items naturally divide into two or more groups it is particularly important that differences between groups are not so large that they minimise differences between items within groups. This would encourage subjects to take a simplistic approach in their judgements since it would be feasible to treat all group members as equivalent to one another and equally different from the members of other groups. It is also important that the stimuli all be reasonably familiar to subjects, although it must sometimes be accepted that the aims of familiarity and representativeness can conflict, since in order to represent all aspects of a concept, it may be necessary to draw on unfamiliar instances.

The selection of familiar and representative items is facilitated if a pool of potential items can be defined. In this study such a pool would consist of recognizable animals that between them represent the range of animal life included in our concept of the term ANIMAL. One way of producing such a pool is the method of semantic investigation by which category norms are elicited. Category norms are the set of items that readily come to mind in response to a category name.

Battig and Montague (1969) gave subjects thirty seconds in which to write down as many examples of a category as they could; owing to the large numbers of subjects, the authors were able to describe their results in normative terms and researchers have subsequently drawn on these norms whenever they have needed good examples of categories. The most frequently produced items in a normative task would fulfil the criteria of representativeness and familiarity, while the least frequently produced or non-normative items should be those that represent the readily available boundaries of the concept. Therefore the procedure for the production of category norms may provide candidates for stimulus selection if both normative and infrequently emitted items are considered.

A possible objection to this approach is that infrequently produced items may be unfamiliar to the majority of subjects, but while there is some correlation between production frequency and familiarity, there is no necessary connection: SPIDER may be an infrequent response to the category ANIMAL, but this does not necessarily render it less familiar than normative responses and it may well be a frequent, albeit incorrect, response to the category INSECT. Hampton and Gardiner (1983) found that SPIDER was the fourth most frequent response to the category INSECT and was the first response for nine of their seventy-two subjects.

A second approach to the selection of stimuli is to set up a framework which, by common consensus, identifies important aspects of the concept. The framework may be defined in terms of attributes or sub-categories, whichever is most appropriate in an everyday sense. If attributes are used, a representative stimulus set is one containing items that possess the attributes with systematic variation; if sub-categories are used, the selected items should characterise each grouping within of the framework. For the present investigation, the number of attributes that potentially describes animals was considered too large for an attribute framework to be useful, but the breakdown of the animal kingdom into MAMMALS, BIRDS, FISH, REPTILES and INSECTS provided a feasible alternative. The stimuli for this study were therefore selected by combining the normative approach, involving the collection of category norms, with framework specification: that is, responses to the category ANIMAL were sampled to represent the major animal families.

The category norm procedure was used in what is termed a category emission task. This is because there were too few subjects for the results to have normative generality in the sense of the Battig and Montague data, but it was reasonable to consider them normative for the population of subjects who produced them and who also participated in the scaling task.

The category emission task was described to subjects and a verbal example was given. Subjects were then asked to write down as many instances of a named category as they could in thirty seconds. When the specified time was over, another category name was provided and the procedure was repeated until five categories had been covered. The categories were ANIMALS, BIRDS, FURNITURE, FOOD and PLACES IN LONDON. All subjects attended the same primary school in the City of London and were drawn from the full age range of pupils. There were ten subjects aged between seven and eight years, ten aged between eight and ten years and ten aged between ten and twelve years. The youngest children were recommended by their head teacher as being 'good writers'.

Examination of the ANIMAL items indicated that mammals, birds, reptiles and fish, but not insects, were represented. Not surprisingly, mammals were produced more frequently than any other group. The responses were arranged in order of frequency as shown in Table 4.1.

Since one aim of this study was to investigate developmental differences in taxonomic structure, instances of the four superordinate animal categories were selected. In addition, the selection included items produced by all age groups of children as far as this was possible. Inevitably, the non-mammalian instances tended to be of low frequency and this was reflected in the final selection. Since the only birds produced in response to the prompt, ANIMAL, occurred with very low frequency, emissions for the BIRD category were used to select two bird instances.

The number of items selected was determined by the need to ensure comparability with previous studies, while at the same time minimising the number of judgements required from subjects. The use of eleven items was considered suitable balance between these conflicting aims. The stimulus set consisted of six mammals, two

Table 4.1: Animal category emmision frequencies (n=30)

No. Ss	Age 10-12	Age 8-10	Age 7-8
9	lion, tiger, cat, dog, rabbit	cat, dog, elephant	
8	bear, sheep, elephant	lion	
7	giraffe, horse, pig, monkey, snake	tiger, monkey	monkey
6	gorilla, wolf	horse, giraffe, zebra, rabbit	lion, cat dog, panda
5	whale, shark, mouse, zebra, guinea-pig, cow	bear, camel, fish, goat	horse, rabbit
4	seal, hamster, lamb, gerbil, crocodile	sheep, cow hare, deer, snake, fox, mouse	goat, fish penguin, bear elephant, bird
3	leopard, ape, penguin, fox, duck, parrot, rat, mole, alligator	whale, bird, crocodile	pony, mouse, hamster, frog, shark, parrot

birds, two reptiles and one fish as follows:

Mammals - CAT, ELEPHANT, MONKEY, MOUSE, TIGER, WHALE

Birds - SEAGULL, SPARROW

Reptiles - CROCODILE, SNAKE

Fish - SHARK

4.2.3 Unidimensional Scales

A novel feature of this application of scaling methodology to the investigation of semantic structure was the labelling of dimensions using unidimensional scales in a property fitting analysis. The choice of unidimensional scales is a further area of stimulus selection, but in this case the items represent hypotheses about the nature of the dimensions and while it is desirable that they be comprehensive, there is no necessity for the scales to be selected according to rigorous criteria of representation and familiarity. In this context they are used as an interpretational tool and not as an end in themselves. If a severely limited set of scales is used - which might be the result of understandable attempts to minimise the number of judgements in the task as a whole - it may be difficult to justify interpretation based on these scales alone. Alternatively, if the plausible range of attributes is represented, interpretation primarily within the context of the unidimensional scales can proceed.

The number of judgements required of subjects is again a major consideration. If there are n objects and u scales, subjects must rate the dissimilarity of every pair of items, [n(n-1)/2];

and the relation between each item and scale, [n.u]. For example, eleven objects and twenty scales would result in a total of 55+220=275 judgements for each subject. Pilot experiments indicated that tasks of this magnitude were within the capabilities of the young children, particularly if carried out under close supervision.

The decision to use a relatively large number of unidimensional scales was taken in order to maintain an objective approach to interpretation. The unidimensional scales were chosen from items generated by children in two pilot studies. In the first study six subjects aged between six and nine years were asked to sort twenty animals into categories. Subjects were asked to repeat the task "a different way" as many times as they could. This resulted in eighteen different category descriptions shown in Table 4.2.

In the second pilot study the same subjects were shown triads selected from a group of thirty animals and asked to say why two animals in the triad were like one another and different from the third. The results are shown in Table 4.3. This task produced nineteen reasons for differentiating between animals, nine of which were equivalent to categories produced in the first task. In both tasks the frequency with which attributes were elicited was taken as an indication their salience to the subjects.

From the combined results of these two tasks it was decided that

Table 4.2: Categories obtained from 6 children aged 5 to 8 sorting 20 animals

Category	No of subjects
water/swim	5
land/can't swim	3
crawl/wriggle/slide	3
fat	2
4 legs	1
chase	1
big	1
fur, feathers	1
smooth skinned	1
bird	1
climb trees	1
don't climb	1
shy	1
pets	1
eaten	1
vicious	1
night	1
woodland	1
farmyard	1
abroad	1
desert	1

Table 4.3: Reasons given by 6 children aged 5 to 8 for differentiating the standard from the pair in animal triads

Reason	No of subjects
big/little (overall/legs/tail)	6
4/2/0 legs	4
long (body/tail/legs)	4
fur	4
flies	3
water/swims	3
wings/fins	3
bird	2
beak	2
face	2
climbs/lives in trees	2
horrible/nasty	2
don't know	2
running	1
moves slowly	1
fat/thin	1
same mouth	1
whiskers	1
African	1
home	1
eats animals	1

NB The reason listed is the first utterance. In some cases the polar opposite or a contrast was supplied, while in others, the absence of the characteristic in other members of the triad was mentioned.

twenty unidimensional scales would adequately represent the attributes that children readily use in order to discriminate amongst animals. The total of twenty-eight different attributes was reduced by selecting the most frequent and condensing the remaining idiosyncratic responses into a smaller number of attributes. In achieving this condensation it was sometimes useful to use a word or phrase that had figured frequently in the transcripts of informal conversations with the subjects. The unidimensional scales are listed in Table 4.4.

4.2.4 The Subjects

In order to obtain ten subjects in each of three age groups for this experiment a total of thirty-eight subjects were tested. Of the eight subjects rejected, seven did not display satisfactory consistency in the rating task and one failed to complete the task. As part of the rating task each subject rated nine items twice, first, in a block at the beginning in the form of practice items and, second, as real items dispersed throughout the rating task. The Pearson correlation between these ratings was used as a measure of subject reliability.

Of the youngest subjects, all of whom were aged six to eight years, two were rejected, one with low reliability at .62 and another because she failed to complete the whole task. The ten displaying satisfactory reliability ranged in age from six years and ten months to eight years and one month. Their reliabilities ranged from .64 to .95 with a mean of .81. As their mean age was seven years and five months, this group was called the 'seven

Table 4.4: The unidimensional scales

Eats other animals	Eats plants
Noisy	Quiet
You don't know what its thinking	You know what its thinking
Animal-like	Not animal-like
Not man-like	Man-like
Friendly	Unfriendly
Small	Large
Lives in water	Lives on land
Walks easily	Can't walk
Not furry	Furry
Tame	Wild
Ugly	Attractive
Swims easily	Can't swim
Lives abroad	Lives in this country
Nice	Nasty
Lives in trees	Lives on the ground
Dangerous	Harmless
Thin	Fat
Lives far from people	Lives near people
I've never seen one	I've seen one lots of times

year olds'. Of these subjects, two attended a local primary shcool, while the remained were the children of colleagues and acquaintances.

Amongst subjects in the nine to twelve year age range, reliability ranged from .31 to .99. Two subjects were rejected owing to unsatisfactory consistency. The remaining ten children displayed reliabilities above .77 with a mean of .86 and ranged in age from nine years and seven months to twelve years and nine months. With a mean age of ten years and ten months, this group became the 'eleven year olds'. Six of these children attended another local primary school, the remaining four again being the offspring of acquaintances.

Fourteen adults - undergraduate volunteers - were tested. Their reliabilities ranged from -.58 to 1.00. Reliabilities for the selected ten ranged from .80 to .94, with a mean of .87. They were all aged over eighteen and the majority were in their twenties.

Children were tested using the Dissimilarities Pre-Test described in section 3.6 and shown in Appendix 1. Much of the pre-testing was carried out in schools and it was the original intention that all subjects would be drawn from the schools. This was not possible, first, because although piloting of the pre-test indicated that the majority of seven year olds could do it and, by implication, the scaling task, it transpired that only 50% of that age group pre-tested at school met the required standard.

Second, this and subsequent experiments required a minimum of two sessions with subjects and some of the originally selected subjects were not available for subsequent testing.

The high failure rate on the pre-test amongst the school population of seven year olds raised issues about the applicability of the technique to normal populations. It appeared that only the more competent seven years olds could do the test. At this stage IQ tests were considered, but rejected because of the already lengthy nature of the tasks required. Also, it was apparent that subjects tested in a classroom with others were less likely to concentrate and it was thought that this accounted for some of the failures.

Another way in which the data lacked generality was in the predominance of females amongst the subjects. Although the original intention was to include equal numbers of males and females, eight of the seven year olds, six of the eleven year olds and six of the adults were female. Amongst the young children, boys were more likely to fail the pre-test, to fail to complete the rating task or to be excluded for lacking consistency in their ratings. Amongst the adults, the men were more inconsistent than women and were excluded on that account.

4.2.5 The Dissimilarities Task

The dissimilarity pairs were presented in a booklet: each page displayed a pair typed in lower case side by side above a horizontally typed numerical scale ranging from "1" labelled

"Very Alike" to "7" labelled "Very Different". There were nine practice pairs followed by the fifty-five pairs representing every possible pair of the eleven animals. The fifty-five pairs followed a 'Ross order' which optimally spaces items (Ross, 1939).

The unidimensional scales were also presented in booklet form, this time each page being headed by an animal name typed in lower case, below which were twenty bi-polar seven point scales. The animals were randomly assigned to one of two orders of unidimensional scales for each subject. In both tasks subjects circled the appropriate numbers on the scales.

All subjects were given essentially the same instructions with some variation and repetition for the children in order to ensure comprehension. The full instructions are shown in Appendix 2. The example used to illustrate dissimilarity judgements referred to "things you spread on bread" such as butter, margarine, Marmite and jam". With the children, a twelve-inch wooden ruler was used to illustrate the meaning of scale. The ruler's inch divisions were covered by a scale with prominent "1" to "7" markings. The "1" end was labelled "Very Alike" and "7" end was labelled "Very Different".

Subjects were then given the list of animals and told that there would be some practice pairs. After the practice pairs questions about the task were dealt with and the test pairs commenced. Interuptions were permitted for clarification and so that subjects could think aloud, but any greater digression from the main task was firmly discouraged. Following the dissimilarity judgements, subjects took a short break before going on to the unidimensional scales.

The children were tested either individually or in groups of up to four subjects. Of the twenty children tested, two in the younger group and five in the older group were unable to finish the tasks in one session, but were available to do so within one week. The ten adults were tested either individually or in small groups. Both tasks were completed in a single session, each subject working at his or her pace from written instructions, following a short verbal introduction.

CHAPTER 5

A MULTIDIMENSIONAL SCALING MODEL OF SEMANTIC DEVELOPMENT

5.1 The MDS Analysis

5.1.1 From INDSCAL to SINDSCAL

The dissimilarities matrices were initially analysed using INDSCAL, the Carroll and Chang (1972) individual differences scaling model. It then transpired that INDSCAL had been superceded by SINDSCAL. Although applications of INDSCAL had been widely reported, Monte Carlo simulations indicated that the programme was prone to producing sub-optimal solutions; of particular concern was the finding that subjects weights are not totally recoverable, which suggests that they are, therefore, not interpretable (Coxon and Jones, 1979b). SINDSCAL is a more efficient programme and was recommended in preference to INDSCAL by Bell Telephone Laboratories (Schiffman et al, 1981). It therefore seemed prudent to re-analyse the data using SINDSCAL.

5.1.2 Dimensionality: Expectations and Limitations

In common with most studies of this kind, the dimensionality of the semantic space was not known in advance. Rather, the aims of the analysis were, first, to discover how many dimensions were needed to represent the stimulus dissimilarities and, second, to generate a meaningful label for each dimension identified. Results of previous studies indicated that two dimensions might suffice in the scaling of animal terms (Rips et al, 1973; Caramazza et al, 1976) and that a third dimension could emerge (Henley, 1969). However, as discussed in Chapter 3, the authors of these studies reported only rudimentary methods of determining dimensionality, therefore their findings could be taken as guidelines only.

Potentially more important in creating expectations of dimensionality was that the number of dimensions which may be extracted in a scaling analysis is limited by the amount of data analysed. The stimulus and subject coordinates (the parameters in 3-way scaling) must be estimated from the subjects' judgements (the data). It is, therefore, desirable for the number of parameters to be small relative to the amount of data (Ramsey, 1978). However, there are no clear criteria for the relationship between parameters and data.

In 2-way scaling there is only one dissimilarities matrix to be analysed and this means that the issue may be simplified to the ratio of stimuli to dimensions. Coxon and Jones (1979b) briefly reviewed rules of thumb for 2-way non-metric scaling, from which one may deduce that the ratio of stimuli to dimensions should exceed 5 and that the corresponding ratio of data to parameters should exceed 2.5. If 2-way scaling were proposed in the current study, the eleven stimuli could support only two dimensions, but Coxon and Jones went on to say that further dimensions could be extracted in 3-way metric or individual difference differences scaling, as intended here. Subject replication in 3-way scaling substantially increase the ratio of data to parameters, thereby justifying the extraction of more dimensions than a given number of stimuli would support in the 2-way case.

The amount of data in a scaling excercise is given by N.s.(s-1)/2, when N is the number of subjects and s is the number of stimuli, the formula s(s-1)/2 providing the number of pairs judged. In the current study, the amount of data was fixed at 1650. The number of parameters to be estimated by the scaling programme is (N.d)+(s.d), when N and s are as before and d is the number of dimensions. Calculation of the ratio of data to parameters for a range of dimensionalities in this study indicated that up to fourteen dimensions could be extracted without violating Coxon and Jones' (1979b) suggested ratio for 2-way scaling. However, solutions of this order of dimensionality would be impractical and also unwise in the absence of guidelines for 3-way scaling. Therefore, it was decided that dimensionality would be chosen with regard to the results of previous studies and on the basis of objective assessments of the model's goodness-of-fit and interpretability.

5.2 Stages In Determining Dimensionality

There are two stages in a typical analysis. In the first stage solutions of different dimensionality are obtained in the expectation that one of these will be the 'correct' solution. It is common practice to generate solutions in dimensionalities ranging from d-3 to d+3, where d is the expected dimensionality (Davison, 1983).

In the second stage the candidate solutions are compared using criteria relating to goodness-of-fit and interpretability. MDS specialists writing in the 1980s have also stressed 'reproducibility' as a third criterion. This may be viewed as a third stage in that, once a solution has been obtained, one repeats the analysis with a completely new set of subjects, or alternatively, with sub-sets of the original subjects, in order to ascertain whether the same dimensions emerge (Schiffman et al, 1981; Davision, 1983). Reproducibility tests were not carried out in this study, partly because it was not common practice at the time when most of the data was analysed and also because there were plans to repeat the experiment with different subjects as part of a subsequent stage in the research, which meant that reproducibility would eventually be tested.

5.3 Generating Potential Solutions

Although the aim of the first stage is to produce a single series of candidate solutions ranging in dimensionality around the expected number of dimensions, these analyses must be carried out not once, but several times, starting the programme on each occasion from a different starting configuration. This is necessary because of the way in which SINDSCAL calculates a solution.

Given a starting configuration, the algorithm iterates until it finds a solution meeting pre-set criteria of fit with the input data. Ideally, the programme should be able to start from several different random configurations and still converge on a single solution. Confidence in the final solution is justified if this occurs. However, in practice different random starts produce different solutions, therefore rules of thumb are required in order to decide when these different results may be considered equivalent. Pruzansky (p.c) suggested that solutions are the same if the stimulus coordinates are identical up to two places of decimals. If different starting configurations produce very different solutions, the iterative function is said to have reached local minima rather than the global minimum indicating the best solution. In this case the criterion for fit should be made stricter and the analysis tried in lower dimensionalities.

In this study two or three dimensions were expected, therefore candidate solutions in one to five dimensions were generated. (Some 6-dimensional solutions were run, but the systematic production of these was not attempted owing to their demands on computer time and space.) An optimal series of solutions emerged from amongst the first three runs in one to five dimensions. Two of these runs were identical according to the 'two decimal places' rule and the third was identical up to one decimal place. This efficient performance by SINDSCAL contrasted with INDSCAL, which had not produced such a clear result, despite many more runs and modifications of the fit criterion.

5.4 Goodness-of-Fit

The overall measure of fit in SINDSCAL is the correlation between dissimilarities in the data and distances in the solution. This is printed for every iteration in the analysis at a given dimensionality. The square of the final correlation is the percent variance accounted for (%VAF) by the model at that dimensionality. Although %VAF inevitably increases with dimensionality, an elbow may sometimes be observed in this function indicating the point beyond which increases in dimensionality are not associated with reasonable increases in %VAF. A plot of this function is shown in Figure 5.1. It may be seen that the correlation between data and solution increased gradually, with a minor discontinuity at dimension II, thus indicating that the solution could be 2-dimensional.

However, detection of 'elbows' and the assessment of their importance relative to other criteria are notoriously subjective procedures. For example, Kruskal and Wish (1978) published an array of elbows with much advice on interpretation; and Davison (1983) described a scaling analysis of hypothetical data which yielded a spectacular elbow at two dimensions, while noting that "In most real data, the elbow is more ambiguous and, consequently, the fit measure alone does not suffice as a basis for deciding dimensionality." (p131). As there can be much debate about elbows, the function in Figure 5.1 was viewed as evidence that the eventual solution might be 2-dimensional, but that higher dimensionalities could not be ruled out.

5.5 Importance of Dimensions

Another way of assessing dimensionality is to look within a solution at the relative importance of each dimension to

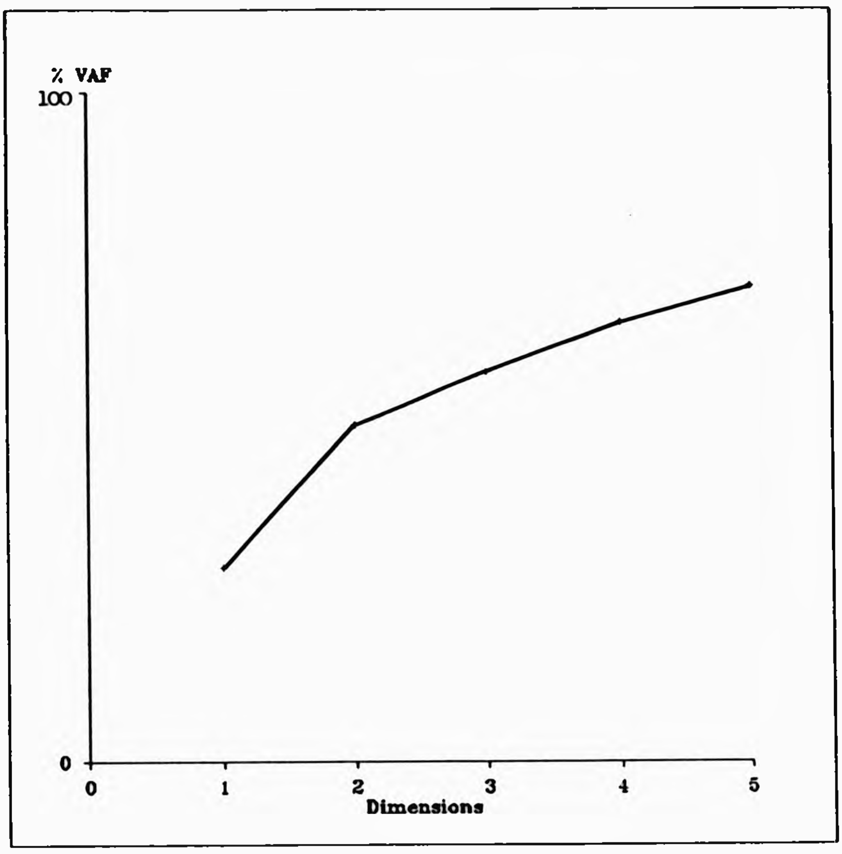


Figure 5.1: Percent variance accounted for (%VAF) by SINDSCAL solutions of different dimensionality for children and adults (n=30)

subjects. Dimensions which are unimportant to subjects may be unnecessary to the solution. Conversely, one may not be justified in discarding dimensions which are important to a substantial minority of subjects. A dimension's importance to the group of subjects as a whole is measured by the amount of variance explained by that dimension, while its importance to an individual subject is given by the subject's weight on the dimension.

The importance of dimensions is best evaluated in terms of individual subjects. In the 3-dimensional solution five subjects weighted dimension III as most important and, for a sixth subject, all three dimensions were equally important. In the 4-dimensional solution only three subjects weighted dimension IV as most important and for two of these subjects, another dimension in the solution was equally important. This comparison indicated that a fourth dimension could be discounted on the basis of lack of importance, but that a third dimension should be considered because it was important to six of the thirty subjects. Furthermore, between two and three dimensions, nineteen subjects' correlations increased by .05 or more. Nine of these subjects were adults, five were eleven year olds and five were seven year olds.

5.6 Independence of Dimensions

Yet another guide to dimensionality is the degree of correlation between dimensions within a solution. Ideally, dimensions should be orthogonal, but this does not always occur. The correlation between dimensions in the 2-dimensional solution was -.065. Within the 3-dimensional solution the correlations were higher, being .001 for the first and second dimensions, -.107 for the first and third dimensions and .317 for the second and third dimensions. As these were all small correlations, neither solution was at first considered very non-orthogonal and further interpretation of the 2- and 3-dimensional solutions proceeded on this basis.

The significance of these correlations was reviewed following the appearance in 1981 of Schiffman et al's basic text on MDS. Schiffman et al reported a 2-dimensional INDSCAL analysis of young and elderly subjects' judgements of food flavours in which the correlation of .30 between the two dimensions was described as non-orthogonal. Nevertheless, Schiffman et al interpreted this solution in terms of meaningful groups of flavours (rather than dimensions) and an age difference in the ability to discriminate: in addition, the configuration emerged consistently in re-analysis of the data using other MDS procedures (KYST, ALSCAL, POLYCON and MULTISCALE).

In applying this report to the current study, it was concluded that lack of orthogonality did not bar interpretation. However, correlation between dimensions raised questions about a dimensional interpretation, as suggested by Schiffman et al's use of categories rather than dimensions in their interpretation of the non-orthogonal food flavour configuration.

5.7 Property Fitting

5.7.1 The Rationale for Property Fitting

Property fitting is primarily an aid to interpretation, but there are two ways in which it can also assist in determining dimensionality in cases where the scaling model's internal goodness-of-fit measures are ambiguous. First, the interpretability of dimensions is a useful criteria, since it has often been said that dimensions which cannot be interpreted probably do not exist (Schiffman et al, 1981). Second, if unidimensional scales are fitted to solutions of increasing dimensionality using multiple regression, the gain in fit for individual scales from one solution to the next can be evidence for accepting an additional dimension. Gain in fit is assessed by testing the significance of the increase in %VAF, that is, the significance of the change in R square. Additional dimensions are justified as long as the fit between solutions and unidimensional scales continues to improve significantly.

5.7.2 Individual Differences in Property Judgements

If differences between children and adults were expected in the SINDSCAL solution, it was also possible that there would be important individual differences in the unidimensional scale ratings. This was examined at an early stage using Principal Components analysis on a subset of sixteen subjects, consisting of four seven year olds, six eleven year olds and six adults. The data for these subjects were collapsed across animals to yield a matrix of twenty scales by sixteen subjects.

Principal Components analysis of the subjects resulted in four factors with eigenvalues greater than one, which together explained 80.8% of the variance. All but one subject - an eleven year old - had high loadings (>.5) on factor 1, which itself accounted for 55.3% of the variance. Examination of the Varimax rotation indicated that the children's judgements were less consistent than those of the adults. Loadings on factor 1 were highest for five adults, two eleven year olds and one seven year old. This left one adult, four eleven year olds and three seven year olds for whom the remaining factors were as important or more important than factor 1. However, as the pattern of loadings was not related to age and there appeared to be a dominant first factor, it was concluded that major age differences in the use of the unidimensional scales had not occurred and that a single matrix could be used for property fitting.

5.7.3 The Fit between Properties and SINDSCAL Solutions

The unidimensional scale matrix for the whole sample consisted of twenty scales by thirty subjects by eleven stimuli. Owing to missing data from one seven year old and one eleven year old, the matrix was collapsed across twenty-eight subjects to yield a mean value for each animal on each scale. Stepwise multiple regression analyses were then used to determine the relation betwen each scale and the 2-, 3- and 4-dimensional solutions.

Twelve unidimensional scales were predicted by the 4-dimensional solution - three at $p<0.001$, three at $p<0.01$ and six at $p<0.05$ level. For the 3-dimensional solution, fourteen scales had significant multiple Rs: one scale fitted at $p<0.001$, with another seven at $p<0.01$ and six at $p<0.05$. This pattern is summarised in Table 5.1. Also shown in Table 5.1 are the scales which significantly increased in fit between dimensions when subjected to the R square change test.

At this point it is useful to outline the convention adopted hereafter for describing the fit betwen scales and solutions and also between scales and dimensions. When referring to the scales in general and to their fit with solutions the scales are described using the terms listed in Table 5.1. These are short labels which identify one end of each scale. For example, when describing the overall fit between a scale and a solution, the label DANGEROUS might be used with the opposite pole, HARMLESS, implied, but not stated. However, when examining the fit between a scale and specific dimension, it was felt that clarity would be enhanced by consistently describing one end of the dimension - arbitrarily the positive end - and referring to the scale by whichever pole is associated with that end.

The three scales which gained in significance at four dimensions - WATER, SWIMS and FURRY - had also been significant for the 2-dimensional solution, therefore the 4-dimensional solution provided no advantage in interpretation. However, of the four scales to gain significance between two and three dimensions,

Table 5.1 A summary of the fit between between 20 unidimensional scales and SINDSCAL solutions (n=30) of different dimensionality, showing the scales fitting each solution and those which increased in fit between solutions

Unidimensional Scale	- Significant Rsq - 2D	3D	4D -	Sig Rsq Change D2-D3	D3-D4
Far from people	**	*	*		
Water	***	**	***		***
Trees					
Abroad	*	*	*		
Wild					
Nasty		*		*	
Friendly					
Dangerous	*	*			
Swims	***	**	***		**
Walks	**	**	**		
Noisy		**	*	**	
Eats animals					
Furry	**	**	**		*
Attractive	**	*	*		
Large	**	***	***	**	
Fat		*	*	**	
Man-like					
Animal-like	**	**	**		
Know thoughts	*				
Seen lots	**	**	*		

*	p<.05
**	p<.01
***	p<.001

three were non-significant at two dimensions. These were FAT, NOISY and NASTY.

5.7.4 Interpreting the 2-Dimensional Solution

The fit between between unidimensional scales and the solution in two dimensions were examined in detail, with particular attention to the simple correlations between scales and dimensions. Whereas a high multiple correlation indicates that the configuration as a whole predicts a scale's variation, a high simple correlation between a dimension and the scale is necessary before the scale can be treated as a label for that dimension. Davison (1983) suggested that the correlation between scale and dimension should be greater than .70 in either the positive or negative direction and that, ideally, the scale should be uncorrelated with other dimensions.

Simple and multiple correlations for all twenty scales with the 2-dimensional solution are shown in Table 5.2 and a plot of the stimulus locations is shown in Figure 5.2. It may be seen that five scales defined Dimension I uniquely - that is, they all displayed correlations above +/-.70 with Dimension I and below +/-.16 with Dimension II. This indicated that the characteristics of the positive end of Dimension I were LIVES NEAR PEOPLE, LIVES IN THIS COUNTRY, HARMLESS, SMALL and I'VE SEEN ONE LOTS OF TIMES.

The difficulty with this result was that it did not permit a simple interpretation of Dimension I. In particular, it combined

Table 5.2 Simple and multiple correlations between the unidimensional scales and the 2-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

Unidimensional Scale	---- Simple r --- DI	DII	Multiple R
Far from people	-.830	.160	.837**
Water	-.682	.641	.907***
Trees	.334	.431	.564
Abroad	-.769	-.134	.791*
Wild	-.584	.341	.659
Nasty	-.688	.231	.714
Friendly	.658	-.223	.683
Dangerous	-.821	.018	.822*
Swims	-.760	.543	.907***
Walks	.590	-.679	.886**
Noisy	.319	.066	.330
Eats animals	-.354	.146	.376
Furry	.574	-.720	.893**
Attractive	.718	-.468	.833**
Large	-.900	.005	.902**
Fat	-.367	.032	.367
Man-like	.250	-.434	.488
Animal-like	.449	-.758	.858**
Know thoughts	-.014	-.793	.796*
Seen lots	.886	-.162	.892**

* p<.05
** p<.01
*** p<.001

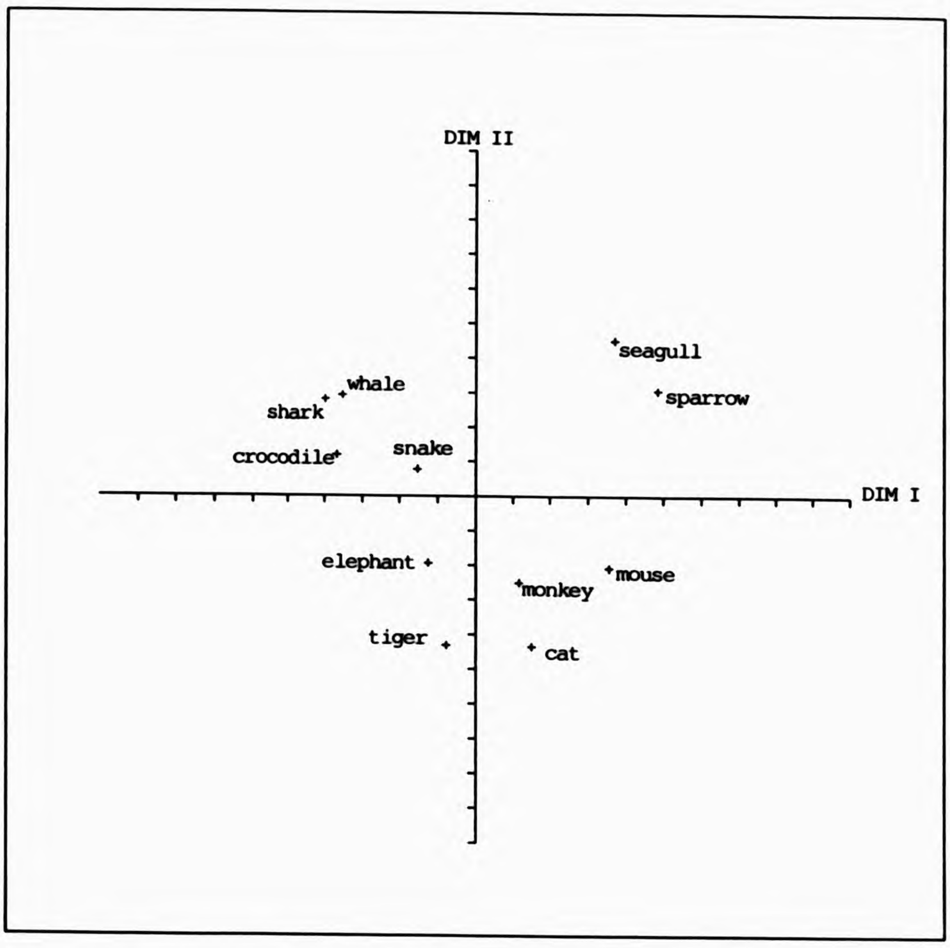


Figure 5.2: Dimensions I and II of the 2-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

size and an element of ferocity in one dimension, when these have allegedly emerged as independent dimensions in previous studies. Additional correlates of this dimension were CAN'T SWIM and ATTRACTIVE, but these scales did not add to the interpretation and they were also correlated with Dimension II.

Dimension II was uniquely defined by only one scale, YOU DON'T KNOW WHAT ITS THINKING. Also contributing to Dimension II were NOT FURRY and NOT ANIMAL-LIKE, but these scales were also correlated with Dimension I. There was no obvious single label for Dimension II, but it appeared to distinguish mammals, with which one can communicate, from the more inscrutable aquatic and avian creatures.

Scales which displayed a significant multiple correlation with the 2-dimensional solution, but which did not meet the criterion of a simple r exceeding +/- .70 for any one dimension were WATER and WALKS. This pattern, together with the non-unique correlates of Dimensions I and II, indicated that a water/land distinction was represented by the solution, but not as an independent dimension.

5.7.5 Interpreting the 3-Dimensional Solution

Examination of the fit between scales and the 3-dimensional solution identified two dimensions which were very similar to Dimensions I and II in the 2-dimensional solution, plus a third interpretable dimension. The simple and multiple correlations obtained in this analysis are shown in Table 5.3.

Table 5.3: Simple and multiple correlations between the unidimensional scales and the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

Unidimensional	------	Simple r	------	Multiple
Scale	DI	DII	DIII	R
Far from people	.118	-.797	-.531	.854*
Water	.581	-.539	-.643	.904**
Trees	.455	.300	.175	.562
Abroad	-.146	-.846	-.315	.861*
Wild	.279	-.437	-.570	.671
Nasty	.115	-.364	-.847	.853*
Friendly	-.115	.359	.794	.803
Dangerous	-.073	-.612	-.741	.852*
Swims	.466	-.575	-.731	.910**
Walks	-.011	.345	.741	.924**
Noisy	.202	-.121	.751	.893**
Eats animals	.054	-.066	-.590	.603
Furry	-.694	.540	.465	.911**
Attractive	-.377	.483	.777	.871*
Large	-.009	-.975	-.405	.980***
Fat	.111	-.691	.226	.852*
Man-like	-.389	.095	.399	.530
Animal-like	-.672	.185	.700	.922**
Know thoughts	-.779	-.102	.212	.804
Seen lots	-.101	-.101	.768	.647**

*   p<.05
** p<.01
*** p<.001

An immediate difficulty with the 3-dimensional solution was that the only scale which uniquely defined Dimension I - YOU DON'T KNOW WHAT ITS THINKING - failed to achieve a significant multiple R with the solution as a whole. However, this scale had been significantly related to the 2-dimensional configuration, where it had characterised Dimension II. Overall, the pattern of correlations between scales and Dimension I of the 3-dimensional configuration was similar to that described for Dimension II in the 2-dimensional solution. Therefore it seemed reasonable to conclude that the original dimension had emerged again, if in somewhat attenuated form.

A similar difficulty arose in the interpretation of Dimension II. Although the pattern of correlations mirrored that of Dimension I in the 2-dimensional solution, on this occasion the scales were not as uncorrelated with remaining dimensions. The best descriptions of Dimension II in the 3-dimensional solution were SMALL and LIVES IN THIS COUNTRY, therefore, it could also be argued that this result simplified interpretation by narrowing the choice of labels.

The positive end of Dimension III was uniquely defined by NOISY and to a lesser extent by NICE and ATTRACTIVE. Non-unique correlates of this dimension were HARMLESS, CAN'T SWIM, WALKS EASILY and ANIMAL-LIKE. Other scales which correlated uniquely with this dimension, but which did not attain significance with respect to the configuration as a whole were FRIENDLY (R=0.803) and EATS PLANTS (R=0.590). Of all dimensions, this was the one most clearly related to ferocity.

5.8 Choosing a Configuration

In comparing the 2- and 3-dimensional solutions, it was evident that, as they shared two dimensions in common, they were, at least, equally interpretable and that choice of dimensionality would rest on the utility of Dimension III. That is, the addition of a third dimension did not alter the first two dimensions so much as to make them any less interpretable; if anything, it simplified the interpretation of what had been Dimension I in the 2-dimensional configuration by absorbing some of its properties. The major question was whether the third dimension was a justified and desirable addition to the model.

The decision to adopt the 3-dimension solution was taken on grounds of interpretability, importance to subjects and consistency with previous findings. First, the addition of Dimension III led to a significant increase in fit as a whole for the important LARGE scale. Second, it permitted the emergence of a significant fit for the NASTY and NOISY scales. Third, it was itself interpretable as a form of ferocity dimension. Fourth, it was the most important or equally important dimension to 20% of subjects. Fifth, and finally, by separating size and ferocity, it permitted an overall interpretation which was consistent with previous findings.

5.9 Labelling the Three Dimensions

A vector representing the fit of a unidimensional scale to the stimulus configuration may be plotted using the standardised regression coefficients (beta weights) and the original as coordinates. The closer the vector lies to a dimension, the more plausible it is as an label for that dimension. The length of the vector is also important because the longer the vector, the more it lies within the plotted 2-dimensional plane; short vectors appear thus because their direction is influenced by remaining dimensions. Plots of the I x II and I x III planes of the 3-dimensional configuration with fitted unidimensional scale vectors are shown in Figure 5.3 and Figure 5.4 respectively. The II x III plane is shown in Figure 5.5.

Dimension I was difficult to interpret in strictly dimensional terms. Although it was related to the water/land scales, it was not uniquely defined by them. From the pattern of scale correlations with dimensions, it appeared that aquatic features were too intercorrelated with size and ferocity within the stimulus set for a water/land dimension to emerge independently.

Another possibility was that Dimension I might reflect relatedness to man. However, if so, why did it not correlate more clearly with either or both of the unidimensional scales MAN-LIKE and ANIMAL-LIKE? The answer to this lay in how subjects interpreted these scales. MAN-LIKE was not correlated with the configuration at all because all animals except MONKEY received a low score. This also suggested that subjects were treating the scale in a superficial visual sense. People may, nevertheless,

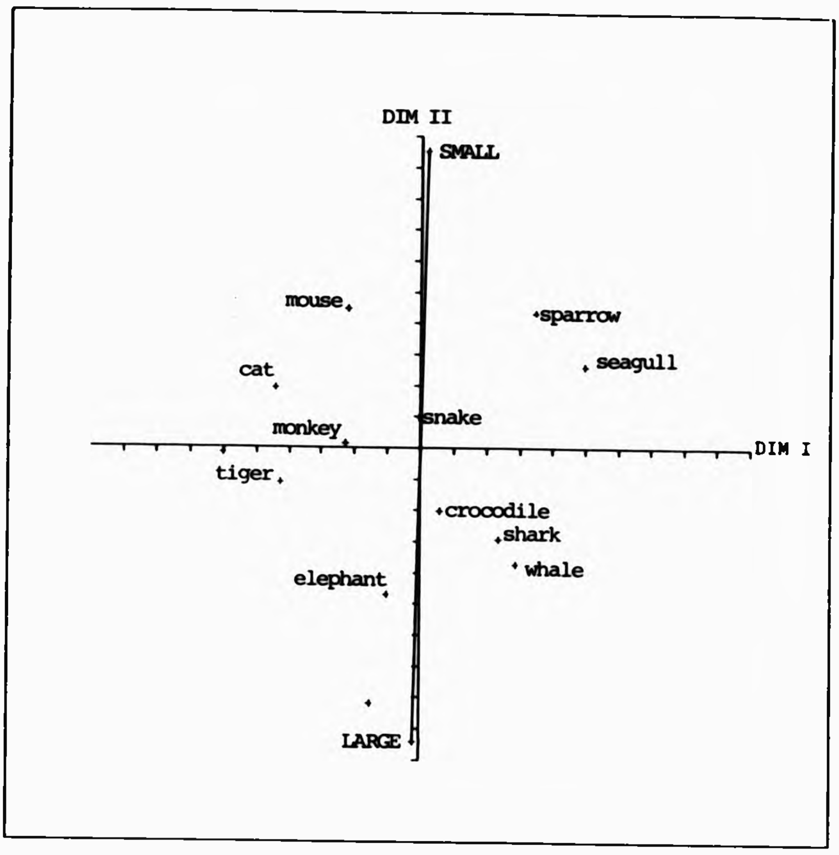


Figure 5.3: Dimensions I and II of the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

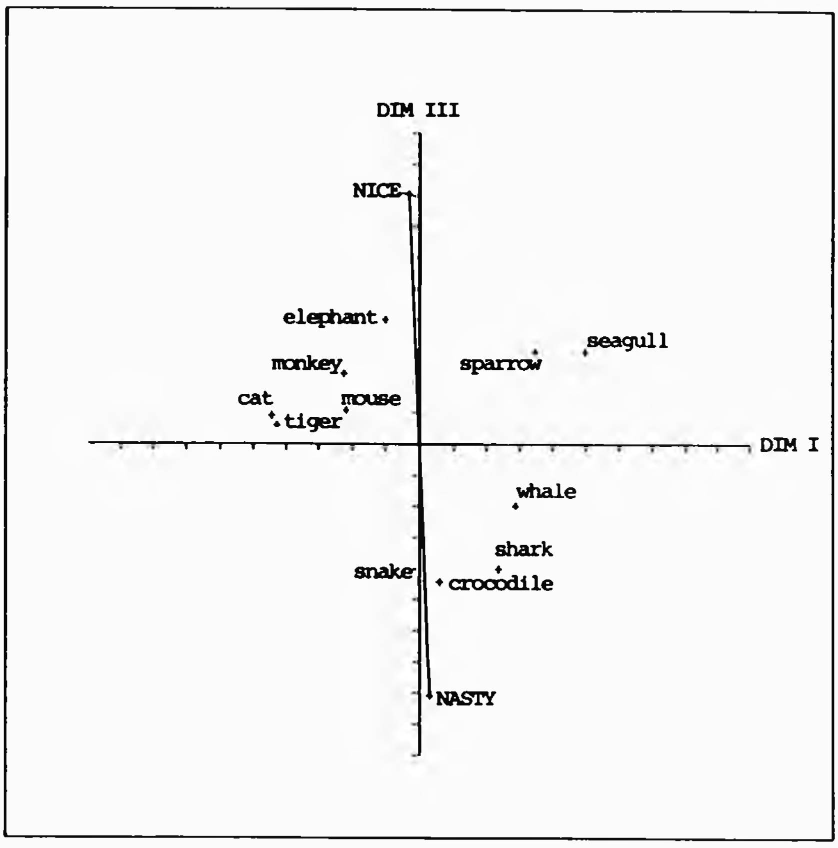


Figure 5.4: Dimensions I and III of the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

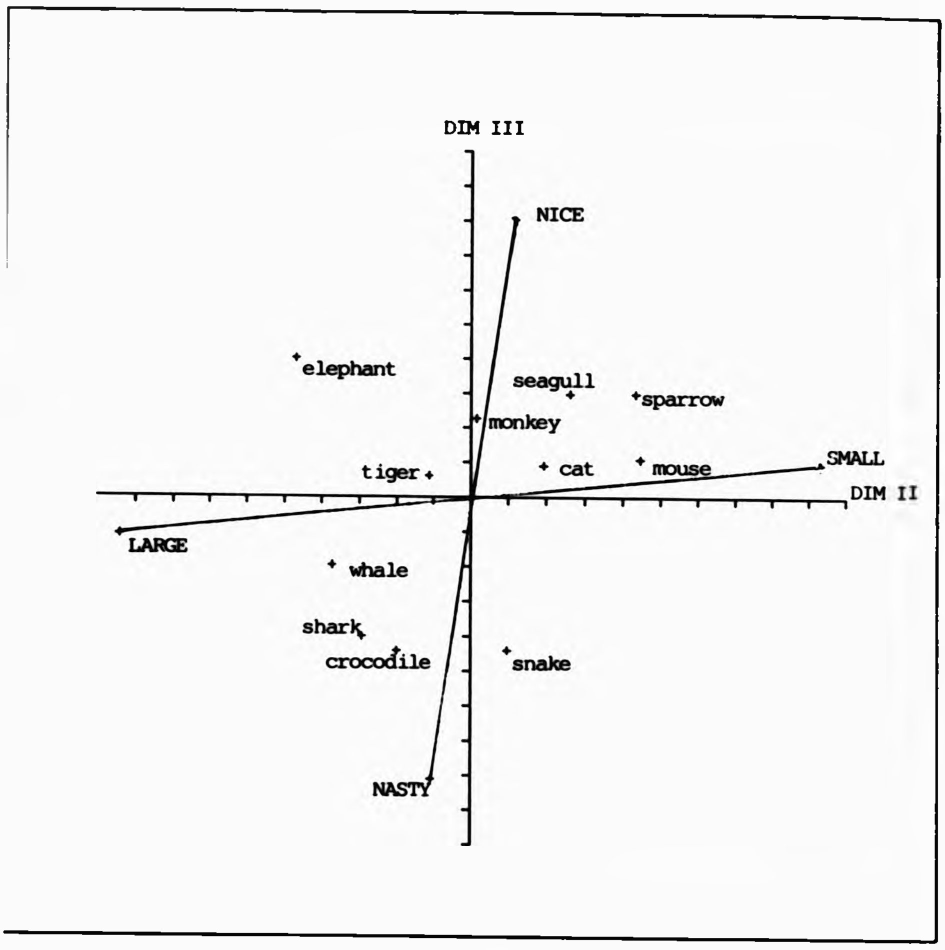


Figure 5.5: Dimensions II and III of the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

have an implicit sense of man's overall relatedness to animals which emerged via the scaling task and in the use of the ANIMAL-LIKE scale. Judgements on the ANIMAL-LIKE scale were not the inverse of those on MAN-LIKE: the multiple correlation between ANIMAL-LIKE and the configuration was .85 ($p<.01$) and land mammals were seen as more animal-like than the other creatures. Therefore, while subjects treated MAN-LIKE categorically, they judged gradation on the ANIMAL-LIKE scale. ANIMAL-LIKE displayed a good correlation with Dimension I, but it also correlated with Dimension III.

A further possibility resulted from examination of the locations of animals in the I x II plane. This suggested the presence of a categorical structure separating land mammals from birds and water-living creatures. Hierarchical cluster analysis of the mean dissimilarities matrix for each age group confirmed that this was a strong element in the structure.

There are several methods of clustering which differ in their definition of the distance according to which a set of items is partitioned into clusters. A researcher's choice of method can be somewhat arbitrary (Everitt, 1974). The major options are the single linkage, complete linkage and average linkage methods. A single linkage analysis defines distance between two clusters as that between the nearest items from each cluster. The complete linkage method uses the opposite approach, finding two items, one from each cluster, which are the furthest apart of all inter-cluster pairs. The average method, as its name implies, defines the distance between clusters as the mean of distances between all pairs composed of one item from each cluster.

The average method was attractive because it would not be unduly sensitive to outlier effects. However, all three methods were used and their results compared on the basis that a stable categorical structure should emerge consistently, irrespective of the clustering method.

The three methods produced very similar results for adults, identifying first a water grouping, followed by land and air clusters. The only difference between methods was that the single linkage tree put ELEPHANT in the water cluster.

For eleven year olds, the complete and average methods produced identical results, distinguishing water from land and air animals. However, the single linkage result was not easily interpreted as it put MONKEY, ELEPHANT and SNAKE each into its own single-item cluster before splitting water animals from the air and land creatures.

For seven year olds, the single and average methods produced similar clusters, isolating first ELEPHANT and then producing water, land and air clusters. The complete linkage result was least interpretable as its major division put CAT, TIGER, MOUSE, SEAGULL and SPARROW into one cluster, while ELEPHANT and MONKEY went with the water animals.

The overall result was that in each case, the average linkage method produced interpretable clusters which were replicated by at least one of the other methods. The average linkage clusters for each group of subjects are shown in Figure 5.6. Clusters of land, water and air habitats were evident for each age group, although for the youngest subjects, ELEPHANT was in the water cluster before forming a single-item cluster on its own. On this basis the best label for Dimension I appeared to be Relatedness to Man/Habitat.

Dimension II was much easier to label as it was highly correlated with SMALL, therefore it was labelled Size. Dimension III was correlated with a group of scales all signifying positive affect, therefore the label Niceness was adopted.

More detailed labels for opposite poles of each dimension were:

Dimension I	-	Positive end:	Unrelated to man, water/air
	-	Negative end:	Related to man, land mammals
Dimension II	-	Positive end:	Small
	-	Negative end:	Large
Dimension III	-	Positive end:	Nice/noisy
	-	Negative end:	Nasty/quiet

These dimensions confirmed and enhanced existing findings. Although size and aspects of ferocity were found in this investigation, there was also evidence that semantic structure for animals terms is more complex than a 2-dimensional structure.

Dimension I was difficult to interpret except in terms of habitat and relatedness to man. Henley (1969) had interpreted one of her dimensions in a similar way, although in her case, relatedness was more biological than thematic. The Size dimension in this

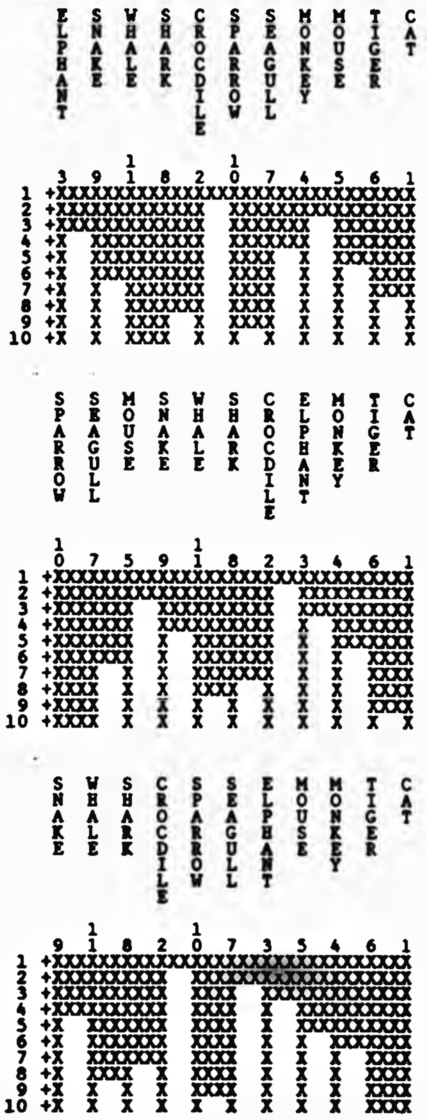


Figure 5.6 Hierarchical clustering solutions (average) linkage method) for mean dissimilarities from 7 year olds (n=10), 11 year olds (n=10) and adults (n=10)

study was consistent with previous studies, but the Niceness interpretation was only partly related to ferocity. It appeared to reflect a general evaluation of the stimuli along a dimension of pleasantness rather than ferocity or predativity. For example, TIGER had a moderate score on the Niceness dimension, probably because its similarity to other land mammals rendered it acceptable in the affective sense that the label of this dimension seemed to indicate.

5.10 Property fitting for Children and Adults

Although factor analysis of a subset of subjects with respect to the unidimensional scales justified combining the subject groups for property fitting, it also indicated that the children's unidimensional scale judgements were less consistent than those of the adults. The question of whether this would modify the interpretation of the solution was answered by repeating the property fitting exercise with scales from children and adults separately. For simplicity, the seven year olds were compared with the adults. The multiple regression results for the seven year olds and adults are shown in Table 5.4 and Table 5.5 respectively.

The multiple correlations were similar, but a minority of scales displayed a significant fit for one group and not the other. Scales which fitted the solution for children, but not for adults were NASTY, FRIENDLY, DANGEROUS and ATTRACTIVE. The pattern of simple correlations was consistent between groups and with the previous result for the combined groups in that these scales all

Table 5.4: Simple and multiple correlations between the 7 year olds' unidimensional scales and the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

Unidimensional Scale	------ Simple r ------ DI	DII	DIII	Multiple R
Far from people	.196	-.747	-.572	.841*
Water	.006	-.515	-.605	.891**
Trees	.549	.298	.229	.657
Abroad	.007	-.860	-.315	.861*
Wild	.207	-.471	-.691	.754
Nasty	-.074	-.430	-.833	.865*
Friendly	-.044	.407	.854	.867*
Dangerous	.050	-.544	-.837	.887**
Swims	.000	-.520	-.657	.915**
Walks	-.502	.477	.808	.943**
Noisy	.198	-.098	.696	.825*
Eats animals	.087	-.304	-.637	.646
Furry	-.509	.585	.542	.840*
Attractive	-.130	.453	.857	.879*
Large	-.051	-.966	-.480	.985***
Fat	.078	-.588	.241	.753
Man-like	-.275	.313	.610	.659
Animal-like	.041	-.095	.721	.809*
Know thoughts	-.773	-.108	.115	.783
Seen lots	-.098	.675	.581	.780

* p<.05
** p<.01
*** p<.001

Table 5.5: Simple and multiple correlations between the adults' unidimensional scales and the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

Unidimensional Scale	------ Simple r DI	------ DII	DIII	Multiple R
Far from people	.021	-.790	-.574	.861*
Water	.564	-.540	-.684	.917**
Trees	.246	.332	.118	.415
Abroad	-.247	-.853	.305	.889**
Wild	.412	.403	-.455	.650
Nasty	.359	.171	-.676	.736
Friendly	-.171	.234	.630	.640
Dangerous	-.125	-.604	-.588	.755
Swims	.363	-.527	-.792	.892**
Walks	-.672	.349	.703	.937**
Noisy	.179	.020	.791	.872*
Eats animals	.024	.094	-.461	.528
Furry	-.765	.438	.420	.907**
Attractive	-.537	.255	.553	.740
Large	-.005	-.980	-.370	.982***
Fat	.068	-.654	.263	.830*
Man-like	-.351	-.038	.381	.517
Animal-like	-.871	.184	.505	.966***
Know thoughts	-.476	-.417	.082	.656
Seen lots	-.085	.875	.594	.938**

* p<.05
** p<.01
*** p<.001

correlated with dimension III.

Scales which fitted the configuration for adults, but not for seven year olds were FAT and SEEN LOTS. For both scales, the patterns of simple correlations for adults, seven year olds and all subjects were comparable and therefore consistent with the initial interpretation.

Two scales which displayed significant multiple correlations for both seven year olds and adults were sufficiently different in their simple correlations when the age groups were compared to merit attention. SWIMS was equally correlated with all three dimensions for the seven year olds, but for the adults, its correlations increased markedly from dimension I through to III. ANIMAL-LIKE was exclusively related to dimension III for the seven year olds, but was split between dimensions I and III for the adults, being primarily related to dimension I.

Evidence of age differences in factul knowledge were not surprising, especially as the SWIMS scale invited comparisons of swimming ability on which subjects may be expected to differ. With respect to ANIMAL-LIKE, observations by subjects at the time indicated that this was one of the more difficult judgements. The property fitting results indicated that for adults, being animal-like is associated with Relatedness to Man/Habitat (Dimension I) and Niceness (Dimension III), while for seven year olds, it is exclusively related to Niceness.

Some investigators have reported combining correlated unidimensional scales in order to improve the fit with the scaling solution and, hence, the interpretability of individual dimensions (for example, Stockdale, Wittman, Jones and Greaves, 1979). In this study it was felt that the dimensions were sufficiently interpretable from their correlations with single scales. In addition, the study sought to identify, as far as possible, the fit of individual semantic features to the space, an aim which may be contrasted the often more general one of producing an overall description of structure.

5.11 Individual Differences in the SINDSCAL Model

Given that the solution corresponded with earlier findings in this field, the next question was whether it would display individual differences in semantic structure with respect to age. As initial issue was whether the model provided a better fit for some groups than others. SINDSCAL provides correlations between each subject's data and the solution. These showed that the fit between the SINDSCAL solution and subject's data increased with age. The solution explained 50.7% of the variance for the seven year olds, 56.1% for the eleven year olds and 65.6% for the adults.

Poor fit can result from 'noise' in the data or from inappropriateness of the model. As all subjects were consistent and the structure was interpretable, 'noise' was not a plausible explanation. The possibility that the children's data did not meet the assumptions necessary for metric scaling was not a serious problem because metric and non-metric MDS models are thought to perform equally well with non-metric data (Weeks and Bentler, 1979). It was more likely that the younger subjects' data were less suited to a dimensional model. The presence of the categorical structure in Dimension I supported this view.

If SINDSCAL provides an adequate fit and an interpretable solution as in this case, individual differences in structure are reflected in the salience weights which show how important each dimension is to each subject. A subject's salience weights in SINDSCAL indicate both the relative importance of the dimensions and the overall fit between the subject's data and the model. If the weights are used to draw a subject vector passing through the origin, the angles between the vector and the dimensions reflect the importance of the dimensions, while the length of the vector represents overall fit. Therefore, in order to measure dimensional importance without confounding by fit, it is appropriate to use a method which removes variability attributable to fit.

One approach is to compare the ratios of the weights rather than their absolute values (Davison, 1983). Alternatively, directional statistics which treat every vector a being of unit length and analyse their directions only may be used (Schiffman et al, 1981). However, Howard (1977) cautioned that the analysis of weight ratios or vector directions could foster a tendancy to overlook differences in fit, which could be a legitimate individual difference, therefore any group differences in the importance of dimensions should be interpreted with attention to the fit for each group. In this study, fit increased with age, but the difference of 15% VAF between the seven year olds and adults was small enough to render comparison of weights meaningful.

For ease of computation, the weight ratio method was the the adopted in this study. Each weight ratio was the proportion of the raw weight to the total of the subject's raw weights. That is, for weights W1, W2 and W3 the weight ratios are:

$$WR1 = W1 / (W1 + W2 + W3)$$

$$WR2 = W2 / (W1 + W2 + W3)$$

$$WR3 = W3 / (W1 + W2 + W3)$$

Weight ratios for subject groups may then be compared using analysis of variance and related techniques. Mean weight ratios for the three age groups are shown in Table 5.6.

In this study the 3-dimensional solution will represent age differences in semantic structure if the three age groups of subjects are reflected in the pattern of the weights. Discriminant analysis assesses the extent to which groups can be predicted from individual's scores on a combination of independent variables and indicates what proportion of cases would be classified correctly. With age as the grouping variable and the dimensional weight ratios as independent variables,

Table 5.6: The relative importance of dimensions in the 3-dimensional SINDSCAL solution for 11 animals, represented by the mean relative weight for each group of subjects on each dimension (n=30)

	Dimensions		
	DI	DII	DIII
7 year olds	.364	.343	.292
11 year olds	.408	.313	.279
Adults	.343	.328	.328

discriminant analysis was used to examine the relation between age and the dimension weights.

Although this analysis extracted Dimension III as a discriminating variable, the result was not significant ($p=0.26$). Only 43.7% of cases were classified correctly, the main mis-classifications being for the younger children, only 30% of whom were correctly classified.

Examination of the univariate Fs indicated that Dimension I ($F2,27=1.25$, $p=.0.30$) behaved like Dimension III ($F2,27=1.42, p=0.26$), while Dimension II was atypical ($F2,27=0.24$, $p=0.78$). From pattern of weights it appeared that the dimensions from I to III were of decreasing importance to all subjects and that this effect was most marked for children. In particular, Dimension III was least important to the children, but was equivalent to Dimension II for the adults. However these effects were not sufficiently robust to attain statistical significance.

5.12 Comparable Studies and Conclusions

In the course of this investigation two other studies of semantic development in the animal domain were published. Howard and Howard (1977) used ten mammals selected from Henley's set with children aged six, nine and twelve and a sample of college students. In order to obtain pair-wise ratings, they used a 'cardboard zoo" with six cages. Children were asked to decide where an animal should go relative to a standard placed at one end, given that animals which were very much alike should be in cages next to each other.

INDSCAL analysis resulted in a 3-dimensional solution reflecting size, domesticity and predativity. The proportions of variance accounted for by solutions in different dimensionality were comparable to those in this study at 30%, 49%, 75% and 61% respectively for solutions in one to four dimensions, but INDSCAL accounted for only 37.0% of the variance for their six year olds, in comparison with SINDSCAL's 50.7% for seven year olds in this study. However, as age was associated with high weights on dimension I (size) and low weights on dimension II (predativity), the result was similar to the pattern of weights in the current study. This was an interesting finding, given that Howard and Howard had used INDSCAL without property fitting and had followed an apparently less rigourous approach in the selection of dimensionality and the interpretation of the dimensions.

Storm (1980) conducted a series of experiments designed to describe semantic structure for twelve mammals in five age groups ranging from kindergarten to college age and also in a group consisting of zoologists. As her pair rating task was only given to subjects in the thirteen year old age group and above, the MDS analyses of these data could not be compared directly with the current findings. In addition, she did not attempt to interpret the INDSCAL space, using this only for a comparison of subject weights, from which she concluded that 'the similarities across the different educational levels were more apparent than the differences' (p.397). This was in keeping with the minor differences found between adults and much younger children in the current study.

In order to interpret dimensions, Storm used MDSCAL, a 2-way procedure, on the average matrix for each group. In no case was a third dimension interpretable and the only differences lay between zoologists and the remaining groups. Dimension I for the zoologists distinguished herbivores from carnivores, while dimension II appeared to represent variation in reproductive rate, an interpretation obtained upon enquiry amongst zoologists.

The subjects aged thirteen, seventeen and twenty produced remarkably uniform 2-dimensional solutions in which dimension I reflected size and dimension II appeared to indicate a water/land discrimination. Storm acknowledged that the aquatic dimension was probably due exclusively to the inclusion of SEAL and that no ferocity dimension emerged because the items did not differ sufficiently in this manner. However, considering the prevalence of ferocity in other studies, its failure to appear was surprising. An alternative explanation was that the MDSCAL procedure did not provide a good fit for the data (as Storm noted on page 395) and therefore could not recover a more complex structure encompassing abstract dimensions.

In comparing the dimensions reported in these studies with those that emerged in the current research, it was evident that a size dimension was the most consistent finding. Evidence for ferocity, predativity or as in this study, the Niceness dimension, was also strong, since its failure to appear in Storm's configurations could be attributed to the stimulus set and, possibly, to a lack of fit between data and solution.

The most problematic finding concerned the third dimension, which, in this study, appeared to reflect relatedness to man and the categorical structure in the stimulus set, separating land mammals from birds, reptiles, fish and the water mammal, WHALE. It was interpreted as domesticity by Howard and Howard and less convincingly as water/land by Storm, with stimulus sets consisting of mammals only in both cases. These are all compatible interpretations since in the real world unrelatedness to man implies either a different taxonomic group - and the major possibilities are avian or aquatic - or unrelatedness in the sense of location and habitat, which may be construed as domesticity.

CHAPTER 6

CONCRETENESS AND TAXONOMIC STRUCTURE IN THE SEMANTIC SPACE

6.1 Experiment II: The Concrete-Abstract Progression

6.1.1 Concreteness and Linguistic Development

Although Experiment I did not conclusively demonstrate individual differences in semantic structure owing to age, the results were sufficiently encouraging to ask whether the MDS model displayed any evidence of the concrete-abstract progression described by Anglin (1970). In Chapter 2 a parallel was drawn between the concrete-abstract progression and the 'form or function' debate that has characterised studies on language acquisition, since, intuitively, 'form' comprises concrete attributes, while 'function' is described in terms of abstract attributes.

If early semantic organisation is primarily concrete and perceptual (Clark, 1973), then attributes such as furriness should be better represented in children's semantic spaces than a relatively abstract notion such as ferocity. If, however, semantic organisation is fundamentally functional (Nelson, 1973), representation of ferocity of should be acquired early in development.

6.1.2 Concreteness in the SINDSCAL Semantic Space

This issue was examined by assessing the extent to which concrete and abstract attributes were represented in the semantic space produced by SINDSCAL. Experiment I indicated that while Dimensions I, II and III were equally important to the adults, they might be of diminishing importance to the children. If a concrete-abstract progression occurs, it would be reasonable to expect concrete attributes to be strongly associated with Dimension I, but less so with Dimensions II and III; similarly, abstract attributes would be associated with Dimensions II and III, but not with I.

Experiment I provided a measure of fit between attributes and dimensions in two forms. The simple correlation between a unidimensional scale and a dimension indicates the extent to which the scale might underly that dimension, while the multiple correlation between unidimensional scales and the overall configuration signifies the fit of the scale to the whole structure. The interpretation of these indices of fit derives from the SINDSCAL model, which extracts dimensions reflecting inter-subject variation. Therefore when a scale has both a high multiple correlation and is uniquely correlated with a dimension, it is an attribute on which subjects place differential importance. Scales with high multiple correlation and high simple correlations with more than one dimension are characteristics which do not differentiate between subjects.

A concrete-abstract progression implies that concrete discrimination is acquired at an early stage and retained while the ability to handle abstract attributes develops. If so, concrete attributes should be represented by solutions of low dimensionality and should not uniquely define dimensions in solutions of higher dimensionality. Abstract attributes should require additional dimensions for their representation and should be uniquely associated with these. If furriness and ferocity are respectively concrete and abstract as intuition suggested, their correlations with the SINDSCAL solutions support this view: FURRY/NOT FURRY was correlated with all dimensions, while the group of scales associated with ferocity were all uniquely related to Dimension III.

This formulation of the concrete-abstract progression predicted that concreteness of unidimensional scales would be positively correlated with their fit on Dimension I. As additional dimensions are added, concrete scales could increase in fit, but the greater increase should occur for abstract scales. It also predicted that concreteness should be associated with non-uniqueness of scales with respect to the dimensions.

A straightforward measure of uniqueness of fit was not available, but the first prediction could be formulated in terms of the gain in percent variance accounted for (%VAF) by successive dimensions in the 3-dimensional solution when they are predictors of a unidimensional scale in stepwise multiple regression. Specifically, concreteness of unidimensional scales should correlate positively with %VAF by Dimension I and negatively with the gain in %VAF obtained by the addition of Dimension III. No specific prediction was possible with respect to Dimension II.

In order to explore these possibilities a measure of the concreteness of each unidimensional scale was obtained from a sample of adult subjects. In addition, where the unidimensional scale consisted of an individual word or short phrase, ratings of imageability were obtained from the MRC Psycholinguistics Database (Coltheart, 1981).

6.1.3 The Subjects and Task

The subjects were ten adult females, all aged over twenty years. They each completed a single-page questionnaire, circling rating scale numbers to indicate the concreteness of the unidimensional scales.

Spreen and Schulz (1966) defined concreteness as that which referred to objects, materials or persons as opposed to abstractness which was a quality of objects that cannot be experienced by the senses. Many investigators have used these or similar words when measuring concreteness (for example, Gilhooley and Logie, 1980). Pilot testing of various forms of wording indicated that subjects had difficulty with the Spreen and Schulz definition in the context of attributes of animals. It was considered important to include this context in the instructions so that subjects rating concreteness and the MDS subjects both interpreted the scales in the same way. Subjects were told that the scales were all ways of describing animals and asked to use a 7-point scale to indicate the extent to which each scale referred to a physical object, attribute or action. The poles of the scale were labelled "Physical" and "Non-Physical".

6.1.4 Analysis: Concreteness and the Dimension Labels

Data from the ten subjects was averaged to produce a concreteness rating for each scale. Mean concreteness was then correlated with the scales' gain in %VAF when the dimensions of the 3-dimensional solution were used as predictors in stepwise multiple regressions. Gain in %VAF is the R square obtained with Dimension I in the equation; thereafter, it is the R square change associated with each dimension.

These results did not support the prediction. The correlations between concreteness and the gain in %VAF with Dimensions I, II and III were 0.144, 0.188 and -0.077 respectively. Mean rated concreteness and the R square change for each scale are shown in Table 6.1.

As some subjects had found the rating task difficult, it was useful to have normative ratings for Imagebility from the MRC Psycholinguistics Database (Coltheart, 1981) for a limited set of scales. These were six scales which had been defined by a single word and could, therefore, be rated easily, in contrast to phrases for which norms do no exist. Imageability ratings range from 100 to 700, with 100 indicating low Imageability.

In many cases polar opposites differ in Imageability. While some differences are minimal, others are substantial. For example, LARGE and SMALL were rated 449 and 447 respectively, but for NOISY and QUIET, the ratings were 215 and 426. This casts doubt on the validity of treating bi-polar scales as unidimensional for

Table 6.1: Rated concreteness of unidimensional scales and their increase in fit (= R sq change) across the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

Unidimensional Scale	Rated Conc	- R square change - DI	DII	DIII
Far from people	3.5	.014	.636	.080
Water	5.3	.338	.292	.189
Trees	4.8	.208	.090	.018
Abroad	4.1	.021	.717	.004
Wild	4.7	.078	.191	.182
Nasty	2.0	.013	.133	.583
Friendly	3.4	.013	.130	.503
Dangerous	3.9	.005	.375	.347
Swims	5.2	.218	.332	.280
Walks	5.2	.374	.120	.361
Noisy	3.2	.041	.015	.742
Eats animals	5.9	.003	.004	.358
Furry	5.5	.483	.293	.054
Attractive	3.0	.142	.324	.383
Large	4.8	.000	.951	.011
Fat	3.8	.012	.478	.237
Man-like	3.6	.152	.009	.121
Animal-like	3.6	.452	.035	.364
Know thoughts	2.5	.607	.010	.029
Seen lots	3.0	.010	.620	.168

semantic purposes, since some scales will have an inherent Imageability dimension. In the case of NOISY and QUIET, the difference probably reflects the contrast in sensory experience: NOISY may refer to a variety of sensations, while QUIET is a unitary experience. Mean Imageability was calculated and correlated with gain in %VAF obtained from additional dimensions as before.

Within this subset of scales, it was clear that concrete scales achieved the greater proportion of their fit from Dimensions I and II, while for the more abstract scales, the major increase in R square change occurred upon the addition of Dimension III. Correlations between mean Imageability of six scales and the gain in %VAF attribute to dimensions I, II and III were 0.639, 0.375 and -0.807 respectively. The correlation between mean Imageability and rated concreteness over the six scales was 0.722, indicating a reasonable overlap between these measures of concreteness. Mean Imageability, mean concreteness and the R square change data for these scales are shown in Table 6.2.

6.1.5 Discussion and Conclusions

Although the prediction was not demonstrated for all scales, it was evident for the subset of six scales for which normative data existed. Both rated concreteness and normative Imageability of these scales was related to their pattern of fit across the dimensions, such that concrete scales were represented by lower dimensions. The moderate positive correlations between concreteness of scales and the explanatory power of dimension II,

Table 6.2: Rated concreteness and normative Imageability of a subset of 6 unidimensional scales and their increase in fit (= R sq change) across the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

Unidimensional Scale	Rated Conc	Norm Imag	- R square change -		
			DI	DII	DIII
Nasty	2.0	381.0	.013	.133	.583
Friendly	3.4	439.0	.013	.130	.503
Noisy	3.2	320.5	.041	.015	.742
Furry	5.5	588.0	.483	.293	.054
Large	4.8	448.0	.000	.951	.011
Fat	3.8	538.0	.012	.478	.237

suggested that concreteness is important at this level of dimensionality.

The correlation between normative Imageability and the solution was probably due to the ease with which single words as opposed to phrases can be rated on a concrete-abstract scale. It may also be the case that although subjects rating concreteness were told about the animals, this alone was not sufficient to ensure that they interpreted the scales in the same way. For example, in relating the unidimensional scales to the SINDSCAL configuration, the high correlation of the scale YOU KNOW WHAT ITS THINKING with Dimension I had been surprising in the context of the concrete-abstract progression. While it appeared intuitively to be an abstract concept - and this was borne out by its subsequent mean rating of 2.5 - it nevertheless appeared to underly discrimination along Dimension I. A possible explanation for this is that subjects in the scaling task 'concretised' the scale, using it as a metaphor for relatedness to man. Knowing what an animal is thinking would imply familiarity and overall similarity to ourselves in form and habitat. These relatively concrete features could have formed the basis of subjects judgements on the scale.

6.2 Experiment III: Thematic to Taxonomic Structure

6.2.1 Taxonomy in Linguistic Development

Adults are adept at perceiving taxonomies and at categorising objects accordingly. They are particularly sensitive to the biological classification of animal terms and will readily distinguish between the various taxonomic classes of animals (Caramazza et al, 1976). In contrast, children tend to use thematic spatial and temporal relations when faced with a categorisation task (Anglin, 1970).

With material that may be organised either thematically or taxonomically, it may be possible to detect a developmental shift from one form of structure to the other. A corrollary of this is that children are less likely than adults to be uniform in their perceptions. This is because thematic organisation can result from any contiguous association and is therefore potentially idiosyncratic. However, children who have shared common classroom and play experiences may be expected to display a degree of consensus in their thematic associations.

6.2.2 Measures of Taxonomic Structure

Taxonomic structure in a semantic space is represented by clustering of items within taxonomically defined groups. The strength of clustering within any group of items may be measured by the inter-item distances within that group; and the degree of separation between groups would be the distances between pairs of items when each item in a pair belongs to a different group.

These somewhat intuitive measures were reported by King, Gruenewald and Lockhead (1978) in a classification study using animal terms: twelve subjects sorted forty animal terms into categories labelled NICE-NASTY, BIG-LITTLE and all combinations of these. Eight different subjects then performed dissimilarity ratings on a 10-point scale and the date were analysed using MDS. The average within group distance (W) and the average between group distance (B), calculated from the scaling solution were found to predict reaction times in the sorting task.

Homa, Rhoads and Chambliss (1979) also used B and W in a study examining the evolution of concepts with experience. Their subjects rated the dissimilarity of pairs of dot patterns following three levels of training and the extent of category structure in the scaling solutions was assessed by the ratio of W to B. As category structure increased, W/B approached zero; that is, as items clustered tightly and distances between clusters increased, the within measure, W, became very small relative to the between measure, B. They concluded that, using MDS, "the changing structure of an entire conceptual space may be tracked, as variables underlying category abstraction are systematically manipulated." (p.11).

6.2.3 Taxonomic Cluster Strength in the SINDSCAL Solution

Since mammals out-numbered other taxonomic classes in the stimulus set, it was decided that the strength of the mammal class relative to others would be assessed. Therefore three values for each subject were required. These were the mean within-mammal distance, averaged over fifteen pairs; the mean mammal-bird distance, averaged over twelve pairs; and the mean mammal-reptile distance, averaged over twelve pairs.

If age is associated with increasing awareness of taxonomy, within-mammal distances should decrease with age and, at the same time, mammal-bird and mammal-reptile distances should increase with age. Differences should be apparent between seven year olds and eleven year olds; and also between eleven year olds and adults. There was no reason to expect a systematic effect of age across the three types of measure.

Before inter-item distances for individual subjects may be calculated, stimulus spaces for individuals must be obtained from the SINDSCAL solution. An individual's stimulus space is the result of applying that person's dimension weights to the overall stimulus space. Coordinates of stimuli are multiplied by the square root of the subject's weight on that dimension to produce weighted coordinates (Carroll and Wish, 1974). These are then used to calculate the Euclidean distances between stimuli for the individual.

As there were three dimensions in this case, the Euclidean distance between two animals, a and b, was obtained from the formula

$$\sqrt{[(a1-b1)^2 + (a2-b2)^2 + (a3-b3)^2]}$$

where a1 and b1 were the coordinates on Dimension I; a2 and b2 were coordinates on Dimension II; and a3 and b3 were coordinates on Dimension III.

6.2.4 Analysis: Taxonomic Structure and Age

The mean within-mammal, mammal-bird and mammal-reptile distances for each group of subjects are shown in Table 6.3. Analysis of variance was applied to each measure in turn. Age did not affect within-mammal distances ($F_{2,27}=2.027$, $p=0.151$) and mammal-bird distances ($F_{2,27}=2.742$, $p=.082$), but a significant effect was found for mammal-reptile distances ($F_{2,27}=4.574$, $p=0.019$).

Further comparisons showed this to be due to the difference between adults and the two groups of children. A significant difference was found between children and adults ($F_{1,27}=4.446$, $p=0.044$), but not between the two groups of children ($F_{1,27}=0.678$, $p=0.417$).

6.2.5 Themes in the Semantic Space

Thematic relations differ from taxonomic links in being based on spatial and temporal contiguity rather than criteria for class membership. In the pilot studies and during informal discussions throughout the research childrens' ideas about relations between animals constantly emerged and these indicated some dominant themes which might be represented in the SINDSCAL solution.

The main themes related to locations where animals are commonly found or to some combination of location and function. Location themes identified animals found in water, in other countries and in trees. Location/function themes identified separate groups of domestic animals, pets and those found in zoos. Some children were particularly creative in their identification of themes. For example, one seven year old viewed ELEPHANT, MONKEY and CAT as a domestic group - in India.

Table 6.3: Mean distances within mammals (WM), between mammals and birds (MB) and between mammals and reptiles (MR) derived from the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults (n=30)

	WM	MB	MR
7 year olds	.391	.653	.532
11 year olds	.404	.694	.547
Adults	.418	.695	.586

In many cases animal themes mirror taxonomy, but there are critical animals for which one or more thematic links will contradict the taxonomic one. Although taxonomic structure was evident in the configuration, it was interesting to examine the locations of critical animals in more detail.

First, some critical animals were identified in relation to the dominant themes. WHALE was the most obvious since this could belong to the water theme or to the mammal cluster. SEAGULL was chosen because it could belong to the water theme or to the bird cluster. Finally, within the mammal cluster, CAT was selected as a critical animal because it is thematically domestic but also taxonomically related to TIGER.

Ideally a critical item should be compared with all items in its potential thematic grouping and with all items in its taxonomic cluster. However this was not feasible because while the composition of taxonomic clusters was known, that of thematic groupings had not been systematically determined. Therefore an alternative method based on triads was devised. For each critical item, one taxonomic and one thematic associate were selected to form a triad. For WHALE, the thematic associate was SHARK and the taxonomic associate was ELEPHANT. For SEAGULL the associates were SHARK and SPARROW respectively. For CAT, the associates were MOUSE and TIGER respectively.

Distances between the critical item and its two associates were then obtained for each subject from individually weighted semantic spaces as in the previous analysis of taxonomic cluster strength. The ratio of thematic to taxonomic distances was taken as an index of thematic/taxonomic preference. Values less than 1 indicated a preference for the thematic associate, while values greater than 1 showed that thematic distances were the greater, thereby indicating a preference for the taxonomic associate. Mean values for each of the three age groups on each triad are shown in Table 6.4.

Table 6.4 Mean ratios of thematic to taxonomic distance within triads derived from the 3-dimensional SINDSCAL solution for 11 animals rated by children and adults

	WHALE	SEAGULL	CAT
7 year olds	0.306	3.826	1.162
11 year olds	0.298	3.731	1.197
adults	0.310	4.035	1.132

WHALE: whale-shark-elephant
SEAGULL: seagull-shark-sparrow
CAT: cat-mouse-tiger

Within each triad the means across age groups were very similar and analysis of variance for each triad did not reveal any significant differences. All means and individual values for WHALE were substantially less than 1 indicating a strong thematic preference at all ages. All means and individual values for SEAGULL were greater than 1 indicating that for this item the taxonomic relation was three or four times stronger than the thematic link. For CAT, again all means exceeded 1, but only by a fraction, therefore, although CAT was seen as more similar to TIGER than to MOUSE, this taxonomic preference was marginal. Unlike the other triads, it was not consistent across all subjects: two of the seven year olds had values less than 1, indicating that their CAT-MOUSE distances were slightly smaller than their CAT-TIGER distances. However, in general, these analyses revealed no developmental differences.

6.2.5 Discussion and Conclusions

Experiment III provided limited evidence of developmental differences in taxonomic structure. In the first analysis strength of clustering amongst mammal items was assessed and found to be equivalent at all ages. However, a developmental difference emerged in the distances between mammals and other taxa. Although only the mammal-reptile distances achieved a significant affect, an age difference was also apparent for the mammal-bird measure. The overall pattern of differences suggested that differentiation of animal taxonomy might be ordered such that mammals and birds are differentiated before mammals and reptiles.

Adults differed from the two groups of children on the mammal-reptile measure, indicating that this differentiation takes place late in childhood. Although distances between groups on the mammal-bird measure were not significant, they suggested that mammal-bird differentiation occurs early in childhood. The eleven year olds and adults both had long mammal-bird distances relative to the seven year olds.

As birds are a more prominent feature of everyday life than reptiles, it is likely that they would be differentiated taxonomically at an earlier stage. It is also possible that the power of flight is a very salient feature which enables birds to be readily distinguished from land-based animals. For both reasons young children might be expected to perceive mammals and birds as belonging to separate classes in advance of differentiating between mammals and reptiles.

In a second analysis, some dominant themes reflecting location and location/function of animals were identified, largely by informal methods. Some of these were the same as the general distinctions concerning habitat, such as water-land discrimination, which emerged from the scaling analysis, but themes could be more specific. For example, MONKEY was a land animal in the scaling solution, but further themes located it in zoos and as a pet as well as in other countries. The presence of these themes suggested another measure of taxonomic structure in terms of the extent to which they affect certain animals, creating a conflict between thematic and taxonomic organisation.

This effect was assessed for some critical items - WHALE, SEAGULL and CAT - which all have strong thematic connotations, the first two with water and the last with domesticity, in conflict with their taxonomic classification. Distances between the critical item and a prominant thematic and taxonomic associate were compared. Although these were arbitrary measures, this was irrelevant to their role in tapping developmental differences in thematic versus taxonomic structure. However, subjects were remarkably consistent on these measures, all agreeing that thematic connotation predominated for WHALE and that the taxonomic relation predominated for SEAGULL. Most agreed that taxonomy marginally outweighted the theme for CAT.

This result showed that for the most obvious forms of thematic organisation, the scaling solution did not represent developmental differences. There was a minor indication that young children display a preference for the CAT-MOUSE theme over the CAT-TIGER relation, as this was evident in the distances of two seven year olds.

It is possible that other tasks might be more sensitive to potential developmental differences of this nature. In this study there was no attempt to classify themes or quantify their relative strength, but there is no reason why methods commonly applied to measuring within-category structure should not be used to assess the structure of themes. For example, Barsalou (1983) has shown that ad hoc categories, such as 'ways to make friends' and 'things that could fall on your head' display graded membership in the same manner as the natural categories investigated by Rosch (1975). Exemplars are not as consistently chosen for ad hoc categories as they are for natural categories, but this appears to reflect stronger established associations in natural categories. The process by which the two types of category are constructed appears to be the same. Thematic groupings are ad hoc categories which may be less ad hoc for children than for adults, therefore it would be instructive to compare themes with ad hoc and natural categories in a developmental context.

CHAPTER 7

THE MDS MODEL AND SENTENCE VERIFICATION LATENCIES

7.1 Introduction and Rationale

If the MDS solution is a valid model of conceptual relations between animal terms, inter-item distances in the model and the projections of items on to the dimensions should predict processing times when subjects make judgements concerning the animals. The model should also reflect individual differences in processing times. A sentence verification task in which the sentences were descriptions of the properties of animals, for example, A CAT IS FURRY, was chosen to be comparable with the unidimensional rating scale task used in Experiment I. Predictions relating latencies to the salience of properties and to the relative locations of animals on the properties were devised and tested. Finally, the fit between property verification latencies and the multidimensional space was assessed. These analyses were the most representative of research in semantic memory in that they tested the model's ability to predict processing times.

7.2 Comparing Child and Adult Latencies

Not surprisingly, children are slower than adults in speeded categorisation tasks (Rosch, 1973; Loftus and Grober, 1973). However, the initial large differences of, for example, 1585 msec reported by Rosch, have been shown to be partly due to the developmental difference in motor reaction time and in encoding the stimuli. Landis and Hermann (1980) compared latencies for oral categorisation and reading, predicting that the developmental difference for categorisation would exceed that for reading by a small amount. Subjects in the oral categorisation task were asked to say YES or NO according to whether a word belonged to a pre-defined category, while subjects in the reading condition merely named the word.

It was found that eight and seventeen year old subjects differed by 1282 msec on the categorisation task, but that reading also produced a substantial difference of 832 msec. This suggested that the difference for categorisation was of the order of 450 msec, assuming additivity of the two processes.

Landis and Hermann reported means of 2155 msec and 2219 msec for eight year olds in the category and non-category conditions of the word categorisation task. These compared with 886 msec and 923 for the seventeen year olds. Schvaneveldt, Ackerman and Semlear (1977) found mean latencies ranging from 1117 msec to 1421 msec across the conditions of a word recognition task performed by seven year olds. As similar tasks in adults generally yield latencies of less than 1000 msec, substantial developmental differences may be expected in the overall magnitude of latencies in a sentence verification task.

7.3 Choosing Attributes for the Reaction Time Task

Owing to the importance of testing the same subjects in both the scaling and reaction time tasks, properties for the reaction time study had to be selected before INDSCAL and later, SINDSCAL, analyses could be carried out on a substantial amount of data. Therefore properties were chosen to reflect the expected dimensions of the stimulus space. LARGE and FRIENDLY were chosen because size and ferocity dimensions are commonly found in studies with adults. The third property, FURRY, was included because pilot studies indicated that it was a salient feature for children. In the event, LARGE and FURRY were represented in the final SINDSCAL solution, but FRIENDLY just failed to reach significance at the 5% level.

7.4 Verification Latencies and the Semantic Space

It was hypothesised that the reaction times would reflect two aspects of the SINDSCAL space. First, latencies would reflect the salience of properties to the individual, where salience is represented by the fit between properties and dimensions, when individuals are allowed to weight dimensions differentially. Second, latencies would reflect the salience of properties with respect to the animals. Here salience refers to the importance of a property in defining an animal. For example, FLIES might be more salient for SPARROW than HAS SKIN. In the case of dimensional properties, a salient property would be one on which an animal has an extreme value. For example, size might be more salient to ELEPHANT and MOUSE than to MONKEY.

7.5 Latencies and the Salience of Properties to Subjects

Looking first at the salience of properties to the individual, the pattern of subject weights in the SINDSCAL solution indicated that Dimensions II and III might be less important to the children than they were to the adults. Although this effect was not statistically significant, it could mean that properties associated with these dimensions are relatively less accessible to children and would therefore, take longer to process in sentence verification.

The property fitting excercise showed that FURRY was associated almost equally with all three dimensions; that LARGE emerged with Dimension II; and that FRIENDLY, although displaying a small correlation with Dimension II, was primarily related to Dimension III. Since adults appeared to use all dimensions equally, there should be little difference between their latencies for judgements involving FURRY, LARGE and FRIENDLY, but children's latencies should increase in that order. It has already been shown that properties associated with Dimension III are more abstract. If children process concrete attributes more readily than abstract attributes, this would result in different processing times for FURRY, LARGE and FRIENDLY.

7.6 Latencies and the Salience of Properties in Animals

7.6.1 The Symbolic Distance Effect

Looking next at the salience of properties with respect to animals, the second hypothesis related the pattern of latencies to the locations of animals on the unidimensional scales and, by implication, on the dimensions of the scaling solution. It was thought that the position of an animal on a scale, could determine the speed with which it is judged on that scale. For example, when people agree with sentences such as ELEPHANT IS LARGE, they are asserting that this fact about elephants is true relative to some standard or reference point. The proximity of ELEPHANT to this reference point might influence reaction time. Then, the question of 'how large is large?' arises.

A possible source of individual differences in such tasks is that people may vary in their choice of reference points. Individual reference points for size may be determined by the sizes of objects already encountered and perhaps, primarily, by one's own size, particularly in the case of children. The smaller one is, the smaller might be the size that one is prepared to describe as 'large' and age, being correlated with size initially, might be related to reference points for size.

A pertinent finding is the symbolic distance effect which occurs when two items are compared on a given dimension. The time taken to compare two items varies inversely with the difference between their referents on the judged dimension (Moyer and Bayer, 1976). Paivio and Marschark (1980) showed that the effect holds for abstract dimensions like INTELLIGENCE and PLEASANTNESS.

Moyer and Dumais (1978) reviewed explanations for the symbolic distance effect, including reference point models which assume that objects are represented on a continuum relative to an ideal or reference point. If, for example, both ELEPHANT and the criterion for LARGE versus NOT LARGE have referents on an internal dimension of size, the speed with which subjects can decide whether to agree or disagree with the statement, ELEPHANTS ARE LARGE, will depend on the distance between the two referents. This means that subjects verifying size should be fast for unequivocally 'true' or 'false' statements, but slow for statements involving medium-sized animals, whose referents are close to the reference point. This would give rise to an inverted U-shaped function relating reaction times to rated size, whether large or small. If, in addition, the criterion for judgement varies between subjects, the positions of subjects' functions should vary in relation to a given ordering of the animals along the dimension of size.

7.6.2 The Symbolic Distance Effect and PREFMAP

The presence of an inverted U-shaped function and of individual differences in its location may be tested using PREFMAP, a hierarchy of MDS models developed for representing preference functions, but in practice, suitable for handling any judgements of objects on attributes. PREFMAP relates unidimensional property judgements for a set of objects to a multidimensional configuration of those objects, obtained, for example, from SINDSCAL. The two forms of relationship of interest here are the vector model and the ideal point model as these are the relationships predicted for unidimensional rating scales and reaction times respectively.

7.6.3 The Vector and Ideal Point Models in PREFMAP

The vector model assumes that the more of a property an object possesses the higher will be the numerical value of the judgement it receives. In the case of preference ratings, this means 'the more, the better', while for size ratings of animals it would represent the fact that the bigger the animal is perceived to be, the higher its rated size should be. This is analogous to the multiple regression procedure used to fit unidimensional properties to the MDS space in Experiment I. However, the vector relation does not hold for all preferences judgements in that too much of a property may be as undesirable as too little. For example, tea that is too hot is as unpleasant as tea that is too cold. In these cases, the most preferred amount of the property is an optimal value best represented by an ideal point in the multidimensional space (Carroll, 1972).

The ideal point is the same as a reference point on a semantic dimension because, according to the symbolic distance effect, preference and latency data behave in the same way. Latencies to judge size will relate to the size dimension in the same way that preference for tea relates to a heat dimension, an inverted U-shaped function describing both types of data. Therefore the hypothesis was that if both rated size and reaction time for the verification of size are fitted to the stimulus configuration using PREFMAP, rated size will be represented by the vector model, while for reaction time, the ideal point model will be more appropriate. PREFMAP provides significance tests which enable the ideal point and vector models to be compared for each subject. Furthermore, if the reaction time data are represented by the ideal point model, it would be appropriate to examine the dispersion of ideal points in relation to age.

Similar arguments may be applied to the other properties, FURRY and FRIENDLY. That is, the vector model should provide a good fit for rating data, while the ideal point model should represent systematic differences in latencies. However, unlike the size dimension, where judgements may relate to the individual's size, there were no obvious reasons for predicting the ideal point pattern for FURRY and FRIENDLY.

As an initial test of the predictions, latencies for LARGE were submitted to a PREFMAP analysis with a 2-dimensional solution from one of the early INDSCAL runs. The dimensions, labelled Mammals/Others and Size, were similar to Dimension I and II in the later SINDSCAL solution.

7.7 The Subjects and Task

Latency data were obtained from nine adults and nine children. The children's ages ranged from seven years and one month to twelve years and eight months, with a mean of nine years and eleven months. All of these data were analysed to examine differences between latencies for the three properties. In examining the data for a symbolic distance effect, data from seven adults and seven children were analysed. These subjects had all produced unidimensional scale ratings and reliable dissimilarities.

Sentences typed in lower case were presented using a tachistoscope and subjects responded by pressing a YES or NO key. The sentences were of the form A CAT IS LARGE and there were thirty-three sentences in all, the eleven animals being paired with three properties, LARGE, FRIENDLY AND FURRY. Alternate subjects used their preferred hand for the YES button. There were eight practice trials.

7.8 Analysis

7.8.1 Relating Latencies to Dimension Weights

The first analysis tested the hypothesis that children's latencies across the properties would differ in accordance with the properties' importances in the children's semantic space, while adults would be able to retrieve all property information with equal speed in keeping with the equivalence of dimensions in the adults' space. Table 7.1 shows the mean latencies for the three properties for each group of subjects.

	FURRY	LARGE	FRIENDLY
CHILDREN (n=9)	2641.44	2814.78	2967.78
ADULTS (n=9)	1148.44	1273.67	1220.33

Table 7.1: Mean latencies in msecs for judgements of 11 animals on 3 properties

The structure of the SINDSCAL space in terms of the dimension weights was apparent in the data. Children's weights on the dimensions had appeared to diminish from Dimension I to III; and latencies for FURRY were fastest, followed by those for LARGE and FRIENDLY in that order. FURRY had correlated with all dimensions; the correlation between LARGE and Dimension II was 0.975; and the correlation between FRIENDLY and Dimension III was 0.794. Adults latencies for the properties were less different: FURRY was again fastest, but this time followed FRIENDLY and then LARGE.

Analysis of variance confirmed that adults were faster than children ($F_{1,16}=16.74$, $p<0.001$) and that latencies for the three properties differed ($F_{2,32}=4.32$, $p<0.05$), but the interaction of age with property was not significant. Planned comparisons based on the SINDSCAL structure showed that reaction times for FURRY were faster than those for LARGE ($F_{1,32}=4.49$, $p<0.05$) and FRIENDLY ($F_{1,32}=7.98$, $p<0.001$), but the latencies for LARGE and FRIENDLY did not differ.

7.8.2 The Symbolic Distance Effect

The data for each property were first examined visually for the presence of the predicted inverted U-shaped relationship between reaction time and rated judgements. Each animal's mean latency was plotted against that animal's mean scale rating for children and adults separately.

The plots for LARGE are shown in Figure 7.1. This shows that while animals with extreme scale ratings tended to have short latencies, animals with moderate ratings had both long and short latencies. Therefore the predicted effect was not obviously present.

The plots for FURRY and FRIENDLY are shown in Figure 7.2 and Figure 7.3 respectively. Neither property showed evidence of a U-shaped function. If anything, the childrens' latencies appeared to increase with ratings for FRIENDLY.

These plots also showed that the data were extremely variable in some cases and that perhaps more observations would be required for specific patterns to emerge. In particular, examination of the raw data showed very long latencies, which in studies with more observations would be excluded from the analysis. However, as the data set was already small, it was appropriate to retain these data and examine the hypothesis further using a non-parametric test.

As the data for LARGE were the only set to show any variability approaching a U-shaped function, a test of trend was applied to these. This examined the latencies of animals ranked according to size in the scaling solution. First, the size ratings predicted by the SINDSCAL solution were calculated from the regression equation obtained when the three dimensions were used as predictors of the LARGE rating scale judgements. These are equivalent to the projections of animals on the LARGE vector which had been located in the solution in Experiment I. For ease

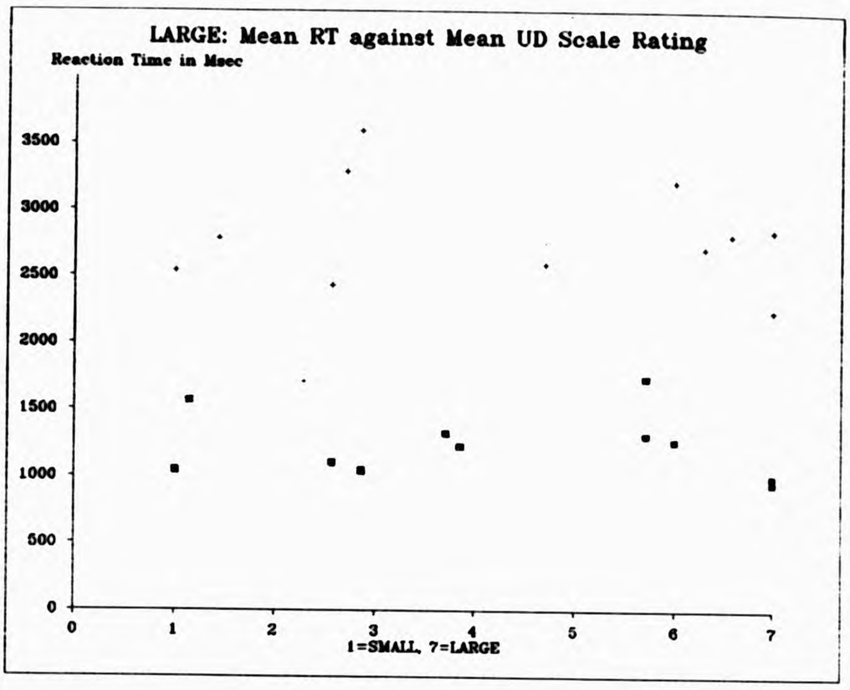


Figure 7.1: Mean latency against mean unidimensional scale rating for each of 11 animals judged by children and adults on size (n=14)

+ children
■ adults

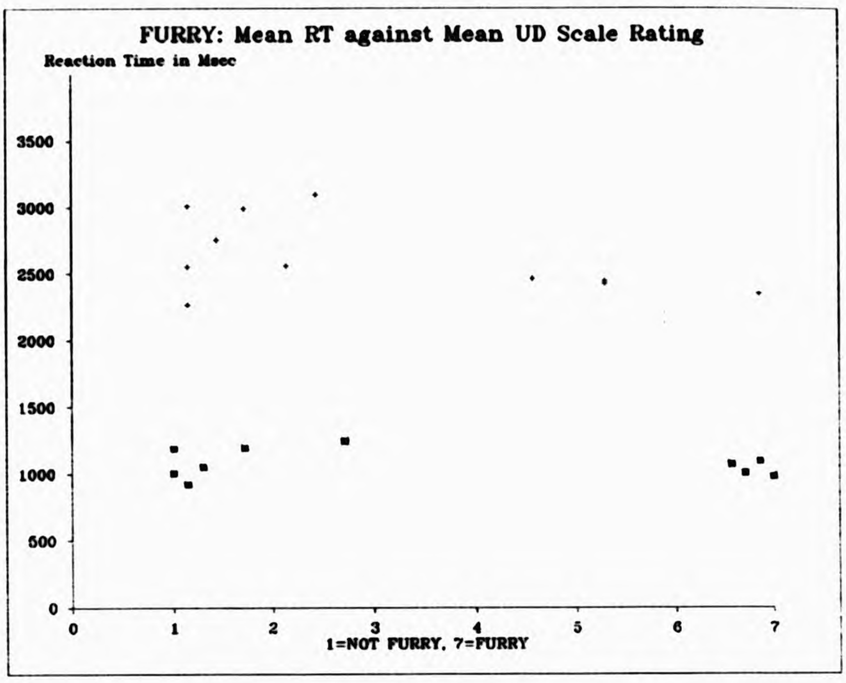


Figure 7.2: Mean latency against mean unidimensional scale rating for each of 11 animals judged by children and adults on furriness (n=14)

- \+ children
- ■ adults

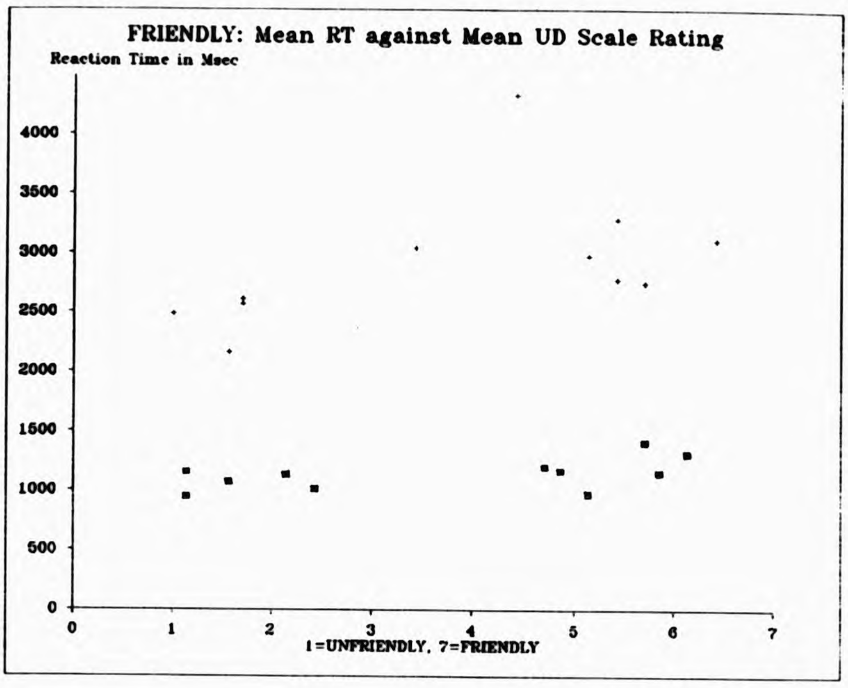


Figure 7.3: Mean latency against mean unidimensional scale rating for each of 11 animals judged by children and adults on friendliness (n=14)

- children
- adults

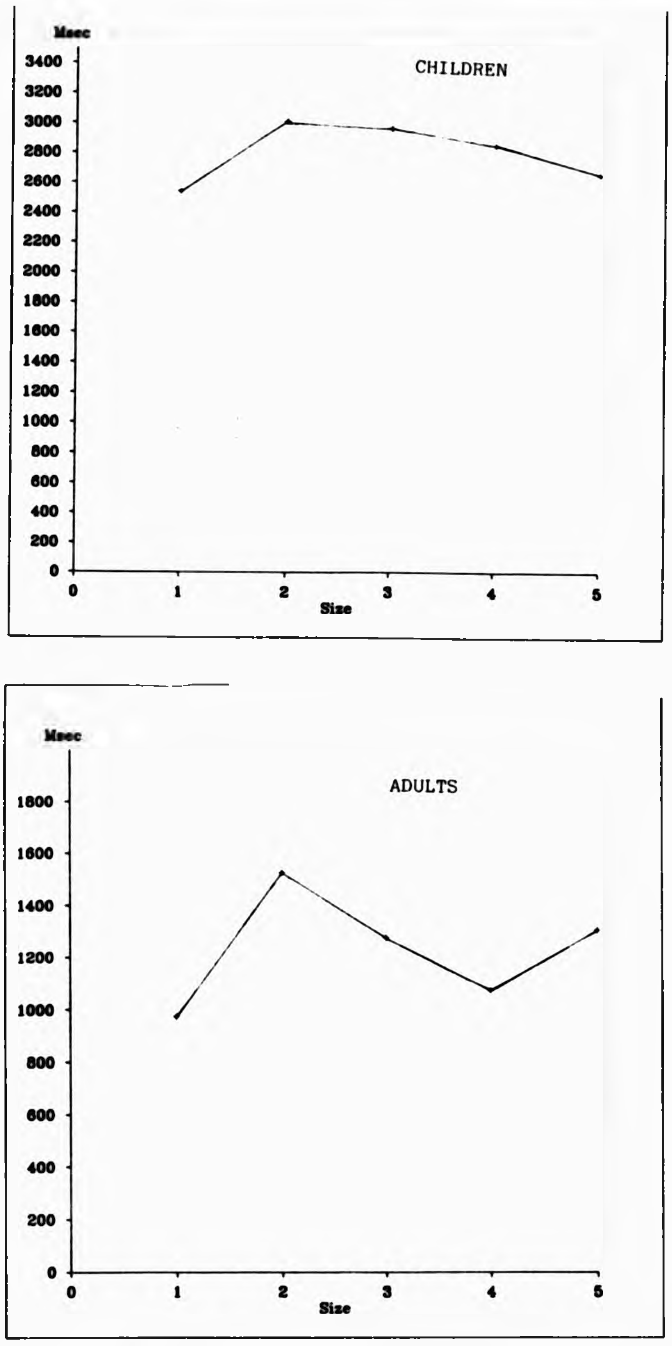


Figure 7.4 Mean latency per rank for children's and adult's judgements on size (n=14)

of testing, the ranks of eleven animals on this predicted vector were collapsed to five ranks by aggregating two sets of adjacent ranks from either end and the three central ranks. Mean latencies for each subject were calculated for each rank and a non-parametric test for monotonic and bitonic trend was then applied to adults and children separately to compare trend in the latencies with the rankings on size. (Ferguson, 1965).

In the childrens' data neither monotonic ($p=.203$) nor bitonic ($p=.108$) trend was discernible. However, for the adults, while the monotonic trend was not significant ($p=.179$), there was evidence of a bitonic relation ($p=.003$). Therefore only the adults' data showed the pattern predicted by the symbolic effect. A plot of mean latency against ranks is shown in Figure 7.4.

7.8.3 The PREFMAP Analysis

Although the data did not strongly display the predicted U-shaped effect, the pilot PREFMAP results were nevertheless examined to establish how well the LARGE latency data could be fitted to a configuration as a whole. The result confirmed the prediction that the vector model would provide a better fit than the ideal point model for the rating data. The vector model predicted all the rating data at $p<.01$ for twelve subjects and $p<.05$ for the remaining two subjects. When the increase in fit to be gained from the ideal point model was examined, only one subject's ratings were found to be better fit by the ideal point model ($p<.05$). Each fitted vector may be plotted using the direction cosine of the vector and the origin as the coordinates. Figure

7.5 shows the close fit between LARGE vectors and Dimension II in the stimulus configuration.

However, the prediction that the ideal point model would provide a better fit for the latency data was not upheld. First, only two subjects' latency data were predicted by the ideal point model, but not by the vector model. The ideal point model accounted for the latency data of another four subjects, but these data had also been predicted by the vector model; and for only one of these four subjects did the ideal point model provide a significant gain in fit. Therefore, to summarise, the ideal point model was superior to the vector model in accounting for reaction time data from only three of the fourteen subjects; and one of these subjects had been significantly fit by the vector model, although the ideal point model did improve the fit.

7.9 A Discussion of the Reaction Time Findings

Latencies for the three properties were generally in keeping with the patterns of dimensional weights for children and adults, even though the predicted interaction was not significant. Both groups were fast to judge furriness, an attribute which was marginally more correlated with Dimension I than with Dimension II and III. The slower latencies for LARGE and FRIENDLY reflected their correlations with Dimensions II and III respectively and this effect was most evident in the children's data.

When the relationship between latency and property rating was

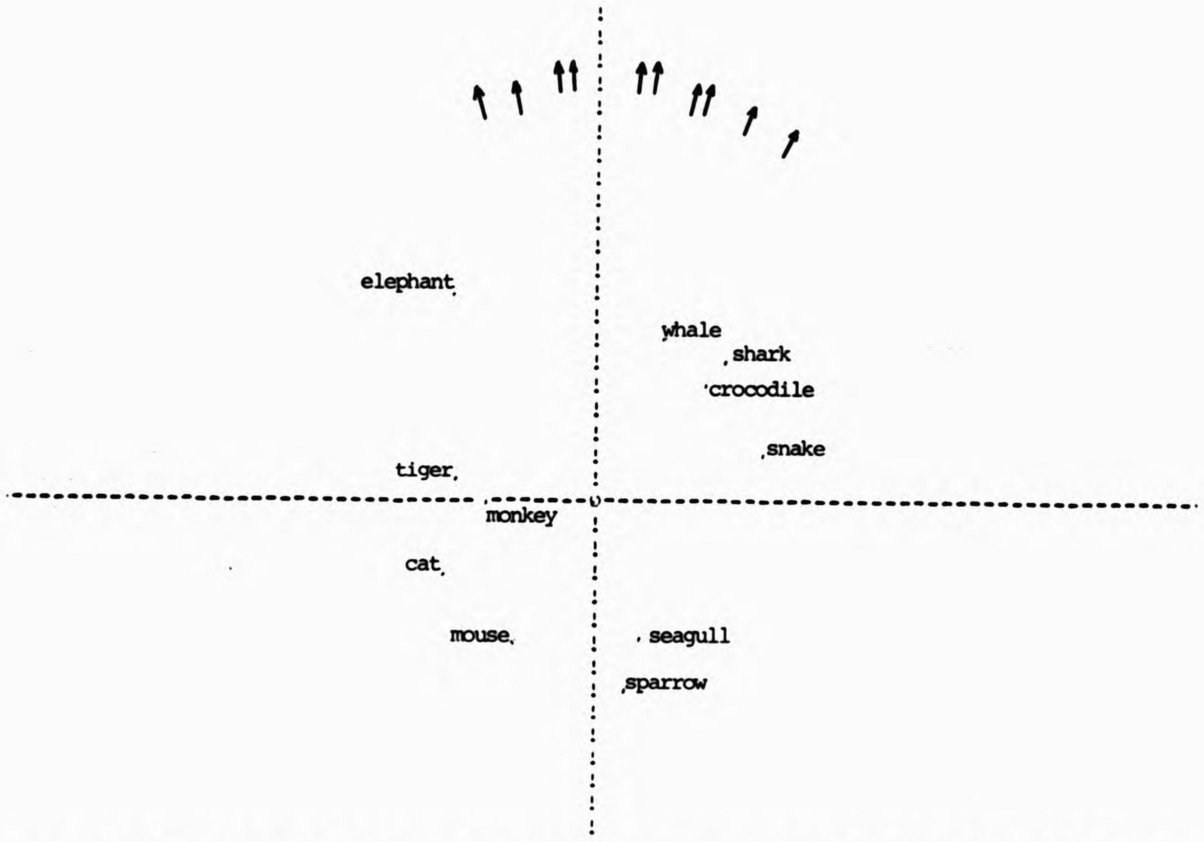


Figure 7.5: Vectors for Large/Small unidimensional scale judgements fitted to a 2-dimensional INDSCAL solution by the vector model in PREFMAP (n=14 children and adults)

visually examined, only the LARGE data displayed evidence of a symbolic distance effect. This was partly confirmed, in that a significant bitonic trend was obtained for the adults' data.

When the fit of LARGE latencies to an MDS model was analysed using PREFMAP, it appeared that although the model predicted some subjects' latency data, it did not do so in the expected manner. However, the failure of the ideal point model was not necessarily attributable to the failure of the MDS model, but to the lack of consistency in the reaction times, which did not strongly display the expected U-shaped function. Therefore, first, the symbolic distance effect predicting that animals with extreme ratings on a property would be judged faster than animals with moderate ratings was modest; and, second, it was not shown by the ideal point model, which should detect such effects.

Another way to look for a U-shaped function is to calculate the mean latency for each unidimensional scale point. The disadvantage of this approach is that unless all subjects use all scale points evenly, some scale points will be associated with means based on very few observations, or in the extreme case, will have no latencies at all. Coltheart and Evans (1981) used this method in their analysis of property latencies against ratings for birds and obtained a significant inverted U-shaped function. They collected data on twenty birds and twenty properties from each of sixteen subjects.

Given the small sample and smaller numbers of judgements obtained in the present study, it was clear that collapsing the data in this way would result in variation in the reliability of means and in missing data, which would render the data unsuitable for analysis. Examination of these plots indicated, as before, the general absence of an inverted U-shaped function, but in one case, a pattern emerged that was similar to the curve analysed by Coltheart and Evans. The plot of their data on a six point scale showed a strong curve, but with a suggestion of bimodality. A similar shape described adults' LARGE latencies in the current study, as shown in Figure 7.6.

If bimodality in this kind of data is not a purely random effect, it may have interesting implications for the way in which people handle dimensional properties. Strictly, people should take longer to assign medium sized animals to appropriate ends of a scale defined as size. If the function then decreases as it approaches the middle of the scale from both directions, this would appear to contradict the theory.

If, however, the behaviour is viewed in terms of the property's salience to each animal, a bimodal curve may indicate that people handle a size dimension by splitting it into categories of large, medium and small. Size could be salient for typical animals in each category. This could mean that information about size is quickly retrieved and processed. In contrast, animals with sizes that are peripheral to each size category might less readily generate a category label for size and

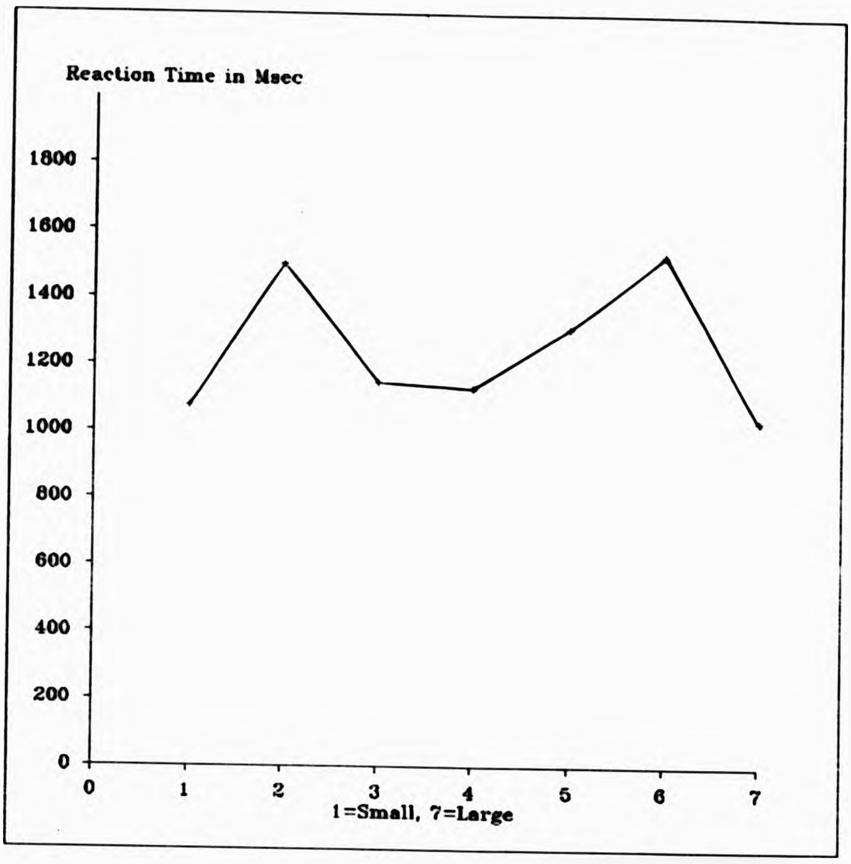


Figure 7.6:Mean latency per unidimensional scale point for adults' judgements on size

would, consequently, be harder to process in this respect. This could mean that local symbolic distance effects apply within regions on an internal dimension and that a simple dimensional model is inadequate even for characteristics which readily lend themselves to dimensional measurement.

A similar view appeared in Smith and Medin's (1981) description of a probabilistic feature model. This describes birds, each containing one of three size features, 'small', 'medium' and 'large'. Each size feature is mapped on to a dimension of size, consisting of features which are 'just noticeable differences'. The implication of this model is that dimensions are handled by aggregating sections of them into categories, which although ordered, are then treated as features in the same way as 'has wings' or 'can fly'.

In general, the results of the property reaction time experiment were disappointing. First, there was insufficient data for any patterns to emerge strongly. Although many reported reaction time studies use small numbers of subjects, they often compensate with large numbers of trials, but this would not have been feasible owing to the fatigue it would have engendered in the children. Second, the latencies may have been determined by a variety of extraneous variables such as frequeny and word length, but it was not possible to control for these because the stimulus materials were chosen in the context of the MDS analysis and were therefore predetermined. Finally, doubts arose about the reliability of the PREFMAP model (from MDS(X), University of

Edinburgh Program Library Unit), which currently carries a warning that errors may occur. As PREFMAP provided interpretable results for the fit between LARGE unidimensional scale ratings and the configuration, it was unlikely to be unsound, but the complexity and inaccessibility of the software meant that the issue could not be explored.

CHAPTER 8

THE EFFECT OF EXPERTISE

8.1 Experiment V: MDS Analysis of Expert Judgements

The results of Experiments I, II, III and IV indicated that the MDS model was of limited utility in representing differences between yound and mature semantic structures. Although the SINDSCAL analysis indicated the possibility of individual differences owing to age, these did not attain statistical significance.

This result may be valid and it may be that the knowledge-base for animal terms is fully developed by the age of seven years. Alternatively, it is possible that certain aspects of the material or the task, or both, ensured that the data did not adequately represent subjects' knowledge of the stimulus domain. The stimuli were constrained to represent the major taxonomic classes in the animal kingdom and may therefore have encouraged subjects to use broad rather than detailed bases for their discriminations. Equally, the judgemental process itself may be a constraint in that a dissimilarity judgement may be unable to encompass more than two or three dimensions of the stimuli being judged. Both of these possibilities imply that MDS may be too insensitive to represent detailed semantic structure.

It was impossible to decide whether the result was valid or attributable to task insensitivity because no independent measure of knowledge had been attempted. Although it is plausible that semantic structures change between the age of seven and adulthood, the differences may be minimal and unlikely to appear in a broadly based task such as scaling. It was therefore felt that a final test of the MDS model would be whether it could represent differences between experts and novices.

8.2 Characteristics of Expert Knowledge

There was comparatively little research on the cognitive differences between experts and novices. Indeed, research in cognitive psychology seemed based on the implicit assumption that expert knowledge is a suitable model for all knowledge in that most experiments examine the behaviour of college students whose education and cognitive abilities make them expert relative to the general population before they become truly expert in any particular field.

Experts clearly possess more knowledge than novices, but little was known about the characteristics of expert knowledge. Chase and Simon (1973) suggested that experts have more 'chunks' of knowledge and that each chunk contains more concepts. Experts' concepts are also thought to be more organised, for example, in terms of class-inclusion relations and to have more links between concepts.

Some research on expertise focused on problem solving and thinking, rather than on conceptual structure and was therefore less relevant. For example, Fiske, Kinder and Larter (1983) argued for a distinction between content knowledge, often described as schemata or prototypes, and process knowledge, which may be thought of as knowledge of strategies. Political experts in their study were able to handle inconsistent information more efficiently than novices. This was attributed to superior processing capacity as a direct result of a tighter organisation of concepts. Another source of similar information about expertise is the research on novice and expert physicists (Chi, Glaser and Rees, 1982). As expertise itself was not of central concern at this stage, expert-novice differences are considered in more detail in the concluding chapter.

Russell and Lambert (1980) used MDS to compare students perceptions of characters from "Othello" before and after a nine hour course of lectures. TORSCA-9, a 2-way scaling programme indicated 4-dimensional solutions for both the 'before' and 'after' conditions, with three dimensions, EVALUATIVE, DEMEANOUR and SOPHISTICATION in common. A major difference between the two configurations was the location of OTHELLO, whose rating on the EVALUATIVE dimension increased in accordance with the "emphasis in the lectures upon a sympathetic understanding of the universality of Othello's dilemma" (p284).

In order to examine experts' views of character in a different type of literature, Russell and Lambert carried out a scaling study of characters from the Sherlock Holmes adventures in which judgements were made by members of the Sherlock Holmes Society of London. The emergence of EVALUATIVE, SOPHISTICATION and ACTIVITY dimensions supported their vew that there is a common framework for the perception of characters.

The most relevant use of MDS in the study of expertise was O'Hare's (1976) work on the perception of paintings by art students and psychology students. INDSCAL analysis of dissimilarity judgements of twelve paintings indicated a 2-dimensional solution in which Dimension I was REPRESENTATIONAL/NON-REPRESENTATIONAL and Dimension II was CLEAR/INDEFINITE.

There were significant differences in the relative weights attached by the two groups to the dimensions, with novices using primarily Dimension I, while experts made full use of both dimensions. This is comparable to the result of Experiment I which indicated that children attach less importance than adults to the higher dimensions. It therefore seemed that unless MDS is particularly insensitive to variation in judgements about meaning, it should be possible to represent differences in semantic structure between experts and novices.

8.3 The Subjects and Task

The experts were twelve medical doctors, all of whom were at least three years post-qualification. Medical training was considered a sufficient indication of expertise in the biological domain.

The scaling task was in many essentials the same as that administered to the subjects in Experiment I. Differences were that the experts worked from written instructions and they were not asked to complete unidimensional scales. In order to maximise the effect of expertise a condition of expert versus lay instructions was introduced: six subjects received standard instructions, while the remaining six were specifically asked to bring all their special expertise to bear on the task.

8.4 Analysis

8.4.1 SINDSCAL Analysis of Experts, Adults and Children

SINDSCAL analysis of the experts' data together with dissimilarities from the subjects in Experiment I - a total of forty-two subjects - resulted in the choice of a three-dimensional solution. Three dimensions accounted for 59.7% of the variance, which was very similar to the 58.1% obtained in the previous analysis of thirty subjects.

The increase in %VAF for solutions ranging from one to five dimensions is shown in Figure 8.1. Again this was similar to the previous analysis, indicating a minor 'elbow' at dimension II.

In order to interpret this solution, unidimensional scale ratings for the ten adults in Experiment I were fitted to the configuration using multiple regression as before. The decision to use data from the adults alone was taken so that property fitting results could be compared for SINDSCAL solutions containing children and adults; children, adults and experts; and experts alone: the adults' unidimensional scale data could be treated as proxies for a set of independent judgements with

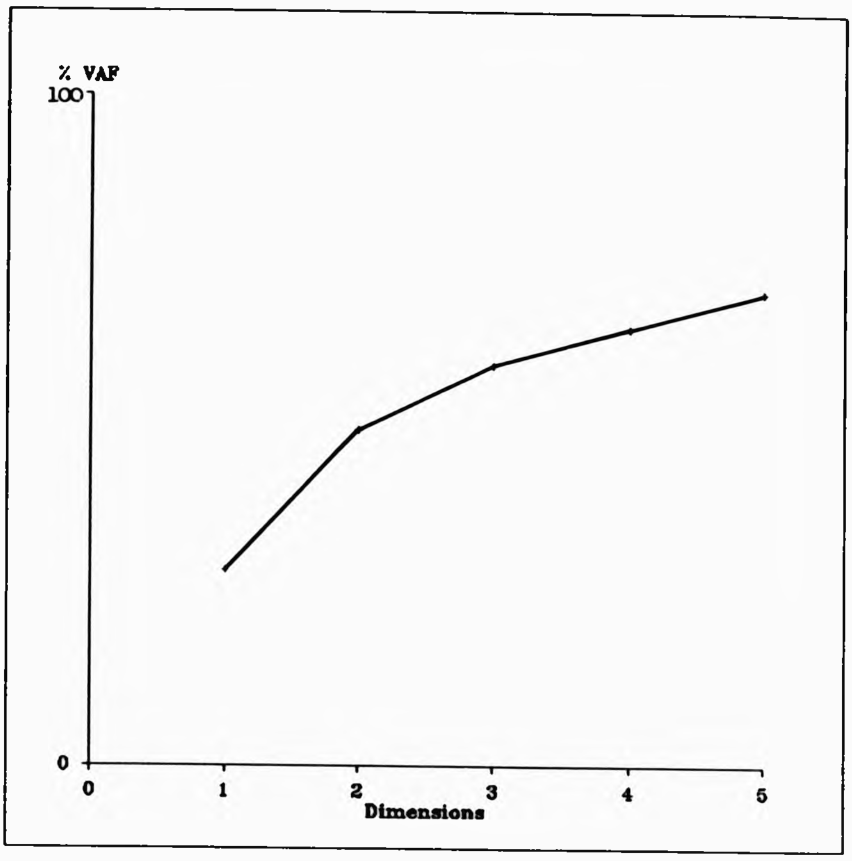


Figure 8.1: Percent variance accounted for (%VAF) by SINDSCAL solutions of different dimensionality for children, adults and experts (n=42)

respect to all three solutions.

The simple and multiple correlations between scales and the solution are shown in Table 8.1. Property fitting indicated essentially the same dimensions as before, but Dimensions I and II were in reverse order. Whereas before Dimension I was categorical and difficult to interpret dimensionally except in terms of Relatedness to Man/Habitat, this structure was now evident, if less clearly, in Dimension II. The clear Size dimension of which had been Dimension II for children and adults was now Dimension I. The Niceness scales remained associated with Dimension III. The I x II plane and I x III plane are shown in Figure 8.2 and Figure 8.3 respectively.

Although these dimensions were similar to those obtained for children and adults alone, there were some important differences. The fit of the Nice/Nasty vector in the I x III plane had deteriorated slightly, but more important, the categorical dimension appeared to be less easy to interpret than before. This was primarily because ELEPHANT had moved away from the mammals, past SNAKE and CROCODILE, towards SHARK and WHALE. This was thought to reflect the experts' appreciation of the similarity of ELEPHANT to WHALE, although it was surprising that ELEPHANT had moved away from the mammals, thus disrupting the taxonomic structure, when the structure could have been preserved by the movement of WHALE.

Apart from WHALE, ELEPHANT was the mammal most plausibly

Table 8.1: Simple and multiple correlations between the adults' unidimensional scales and the 3-dimensional SINDSCAL solution for 11 animals rated by children, adults and experts (n=42).

Unidimensional Scale	------ Simple r ------			Multiple
	DI	DII	DIII	R
Far from people	.825	-.078	-.350	.848*
Water	.688	.475	-.533	.898**
Trees	-.358	.213	.090	.429
Abroad	.807	-.284	-.004	.854*
Wild	.437	.317	-.520	.662
Nasty	.293	.192	-.804	.814*
Friendly	-.337	-.026	.686	.720
Dangerous	.652	-.239	-.453	.784
Swims	.696	.239	-.587	.844*
Walks	-.494	-.540	.735	.914**
Noisy	-.276	.328	.608	.776
Eats animals	.042	-.059	-.465	.499
Furry	-.463	-.683	.624	.924**
Attractive	-.335	-.387	.712	.776
Large	.953	-.015	-.075	.964***
Fat	.541	.212	.519	.935**
Man-like	-.100	-.312	.464	.512
Animal-like	-.272	-.755	.726	.965***
Know thoughts	.351	-.437	.386	.673
Seen lots	-.941	-.019	.278	.945**

* p<.05
** p<.01
*** p<.001

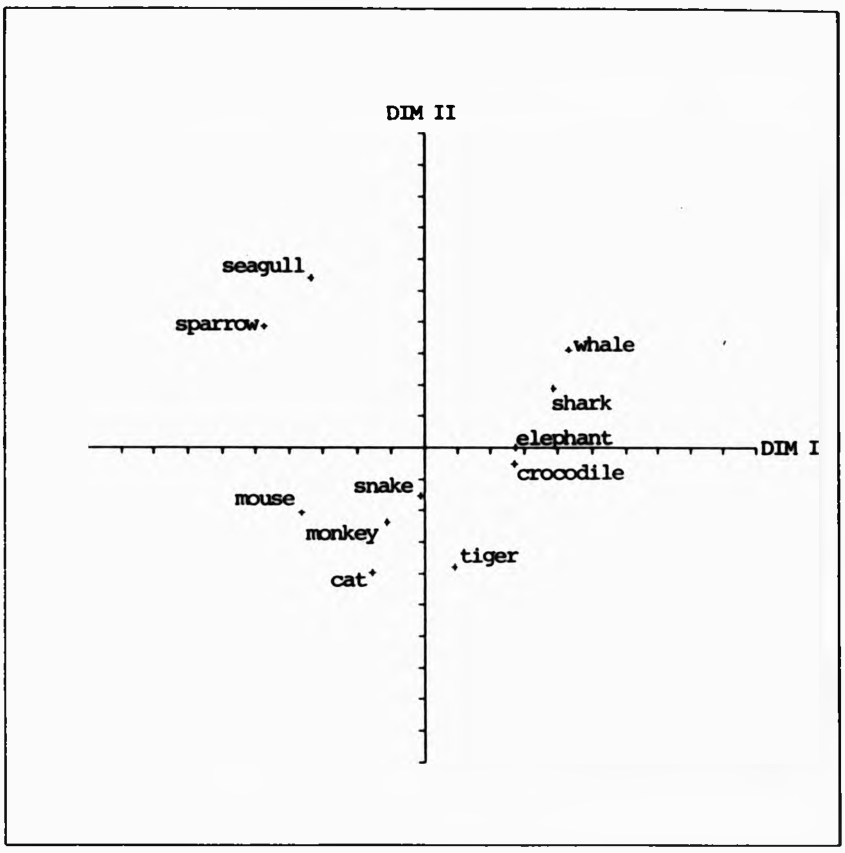


Figure 8.2: Dimensions I and II of the 3-dimensional SINDSCAL solution for 11 animals rated by children, adults and experts (n=42)

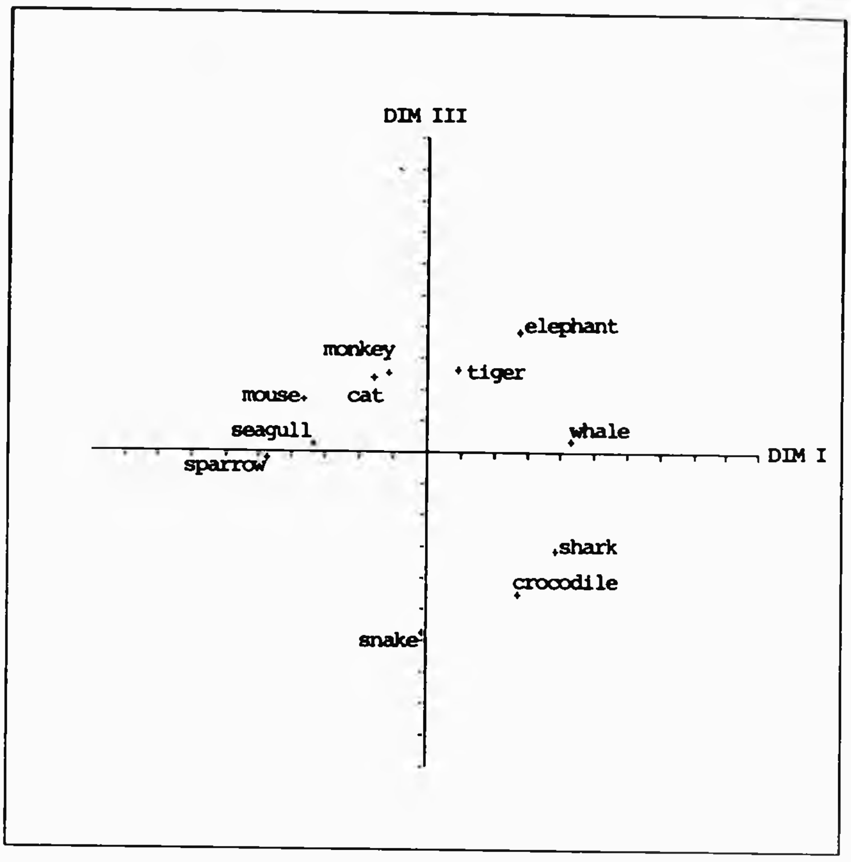


Figure 8.3: Dimensions I and III of the 3-dimensional SINDSCAL solution for 11 animals rated by children, adults and experts (n=42)

associated with a watery habitat, but the water/land scales did not uniquely define this dimension. It appeared that the aquatic features of WHALE were dominant for the majority of the subjects, if not for the experts themselves, and that this, in addition to the taxonomic relation with ELEPHANT, overcame ELEPHANT's links with the land mammal cluster.

Experts were combined to compare subject groups. Mean weight ratios for each group of subjects on each dimension are shown in Table 8.2. Discriminant analysis of the weight ratios extracted, first, Dimension II ($p=0.03$), followed by Dimension I ($p=0.05$). Examination of the weight ratio means and univariate Fs indicated that subjects differentially weighted Dimension II ($F_{3,38}=3.19$, $p=0.03$) and Dimension III ($F_{3,38}=3.06$, $p=0.04$), rather than Dimension I ($F_{3,38}=0.46$, $p=0.71$). Dimension II, the categorical structure, differentiated children from lay and expert adults, with the children giving it higher weights. Dimension III, Niceness, was important to the experts in comparison with the remaining subjects.

When the pattern of weights within subject groups was compared with the previous SINDSCAL solution, some similarities and differences emerged. The seven year olds displayed a consistent pattern, weighting Dimensions I (Size) and II (Relatedness to Man/Habitat) almost equally and more importantly than Dimension III (Niceness) in both cases. In the previous solution the eleven year olds had placed more importance on Dimension I - the

Table 8.2 The relative importance of dimensions in the 3-dimensional SINDSCAL solution for 11 animals (rated by children, adults and experts), represented by the mean relative weight for each group of subjects on each dimension (n=42)

	Dimensions		
	DI	DII	DIII
7 year olds	.375	.361	.264
11 year olds	.360	.387	.253
Adults	.391	.313	.296
Experts	.337	.285	.378

categorical structure - relative to II and III, but in this analysis, Dimensions I and II received similar weights and were more important than Dimension III.

Lay adults also displayed a different pattern: while in the previous analysis, they treated all three dimensions with approximately equal importance, this solution indicated that Dimension I, Size, outweighted the others in importance. These differences did not dramatically change the previous interpretation of age differences in semantic structure because the major finding - that the Niceness dimension was more important to adults than it was to children - was also apparent in this solution. However, they indicated that the subject weights were affected by context and that analysis of subsets of data might give rise to inconsistent interpretations of individual differences.

For experts, the most important dimension was Niceness (III), after which came Size (I) and Relatedness to Man/Habitat structure (II). This was consistent with the view that semantic development would involde increasing reliance on abstract concepts.

8.4.2 The Expert Semantic Space

Data from the twelve experts were analysed using SINDSCAL to explore contextual effects further and to examine the role of expert versus lay instructions. This time SINDSCAL performed better than on either of the previous occasions, explaining 68.8%

of the variance with three dimensions. A 3-dimensional solution had explained 58.1% of the variance for children and lay adults (n=30), while for children, lay adults and experts (n=42), the figure was 59.7%.

Examination of the increase in variance accounted for by solutions for one to five dimensions indicated a discontinuity after the 3-dimensional solution. Shown in Figure 8.4 this provided a clearer indication of dimensionality than comparable plots for the two previous solutions.

The dimensions were less correlated than in previous analyses. Between dimensions I and II, r=.185; between dimensions I and III, r=-.174; and between dimensions II and III, r=.100.

Examination of the 4-dimensional solution showed that the fourth dimension was the most important to three subjects and accounted for 11.4% of the variance, therefore an attempt was made to interpret four dimensions.

Unidimensional judgements from the ten adults were fitted to configurations in three and four dimensions. The simple and multiple correlations for the 4-dimensional solution are shown in Table 8.3 and for the 3-dimensional solution in Table 8.4.

The 4-dimensional solution was very difficult to interpret. Dimension I was Niceness, but Size was split between Dimension II and IV. Dimension II appeared to be the categorical structure, while Dimension III was similar, but without the aquatic

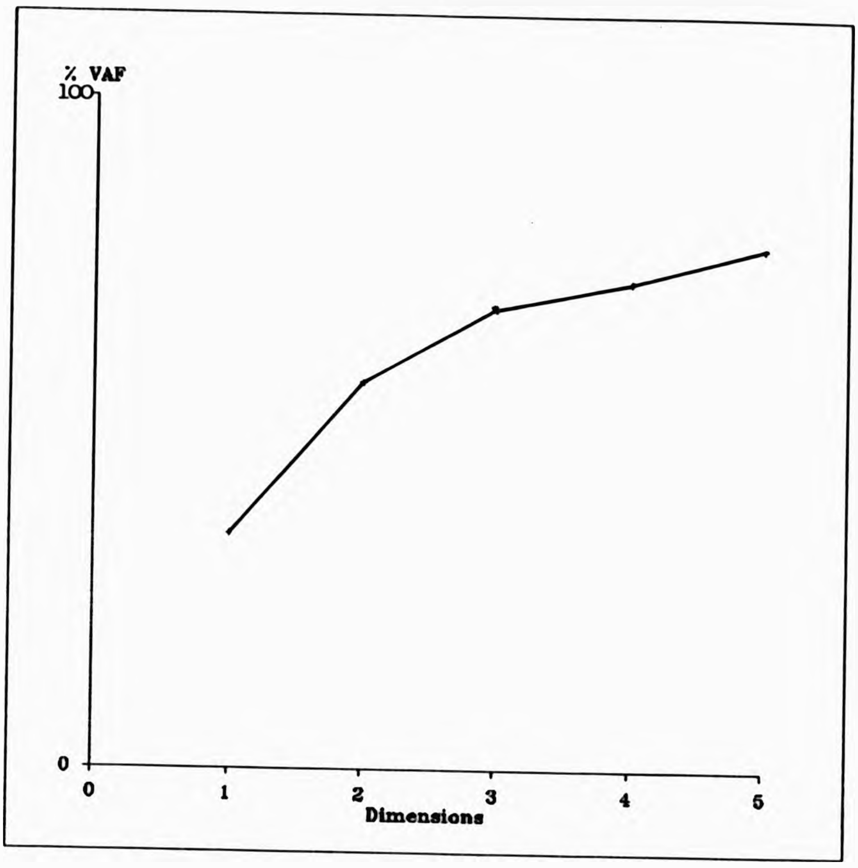


Figure 8.4: Percent variance accounted for (%VAF) by SINDSCAL solutions in different dimensionality for experts with ordinary instructions and experts with 'expert' instructions (n=12)

Table 8.3: Simple and multiple correlations between the adults' unidimensional scales and the 4-dimensional SINDSCAL solution for 11 animals rated by experts (n=12)

Unidimensional Scale	Simple r DI	DII	DIII	DIV	Multiple R
Far from people	.252	.700	.300	.598	.822
Water	.419	.792	-.236	.448	.865
Trees	.105	-.226	-.258	.338	.442
Abroad	-.046	.520	.473	.660	.810
Wild	.543	.421	-.201	.342	.691
Nasty	.860	.349	-.089	.201	.890*
Friendly	-.751	-.263	-.033	-.336	.847
Dangerous	.490	.404	.346	.639	.860
Swims	.489	.676	-.019	.513	.800
Walks	-.616	-.746	.343	-.226	.888*
Noisy	-.490	-.412	-.442	-.110	.764
Eats animals	.571	-.164	-.056	.331	.897*
Furry	-.598	-.768	.511	-.167	.961**
Attractive	-.724	-.570	.159	.000	.818
Large	.024	.692	.181	.870	.951**
Fat	-.579	.410	-.120	.488	.932**
Man-like	-.488	-.312	.319	-.081	.562
Animal-like	-.678	-.592	.601	-.025	.940**
Know thoughts	-.358	-.036	.495	.438	.668
Seen lots	-.155	-.836	-.236	-.699	.937**

* p<.05
** p<.01
*** p<.001

Table 8.4: Simple and multiple correlations between the adults' unidimensional scales and the 3-dimensional SINDSCAL solution for 11 animals rated by experts (n=12)

Unidimensional Scale	------ Simple r ------ DI	DII	DIII	Multiple R
Far from people	.298	.790	.217	.823*
Water	.486	.712	-.316	.862*
Trees	.089	-.387	-.204	.442
Abroad	-.016	.742	.402	.817*
Wild	.545	.406	-.210	.648
Nasty	.842	.261	-.073	.851*
Friendly	-.715	-.292	-.068	.754
Dangerous	.470	-.592	.348	.788
Swims	.526	.677	-.067	.792
Walks	-.684	-.535	.417	.879*
Noisy	-.514	-.332	-.384	.725
Eats animals	.486	-.003	.080	.527
Furry	-.672	-.514	.603	.957***
Attractive	-.762	-.322	.230	.794
Large	.066	.926	.115	.932**
Fat	-.512	.567	-.203	.932**
Man-like	-.473	-.187	.335	.556
Animal-like	-.731	-.305	.649	.937**
Know thoughts	-.376	.272	.507	.653
Seen lots	-.222	-.937	-.126	.939**

* p<.05
** p<.01
*** p<.001

component. This may have signified the emergence of separate aquatic and relatedness to man dimensions, but an unclear pattern of unidimensional scale correlations and doubts concerning the validity of higher dimensions, given the small number of stimuli, led to the selection of the 3-dimensional solution. The I x II plane and the I x III plane are shown in Figure 8.5 and Figure 8.6 respectively.

Property fitting in three dimensions showed that the same three dimensions as in the previous analysis emerged again, but with yet another change of position. Dimension I was Niceness; Dimension II was Size; and Dimension III was Relatedness to Man/Habitat, on this occasion, much disrupted by the displacement of ELEPHANT away from the land mammals. Had it not been for the previous solutions with different subjects, Dimension III may have been regarded as uninterpretable, but in the context of the other solutions, its development could be traced. It was possible that interpretation of the experts' space using experts' unidimensional scales and unconstrained by previous findings would lead to different conclusions, but this was thought unlikely and it contradicted the aim of representing individual differences in a single model of semantic space.

The persistent emergence of similar dimensions in all three analyses was supported by correlations between dimensions from the configurations relating to children and lay adults (n=30); children, lay adults and experts (n=42); and experts alone (n=12). These correlations are shown in Table 8.5 together with

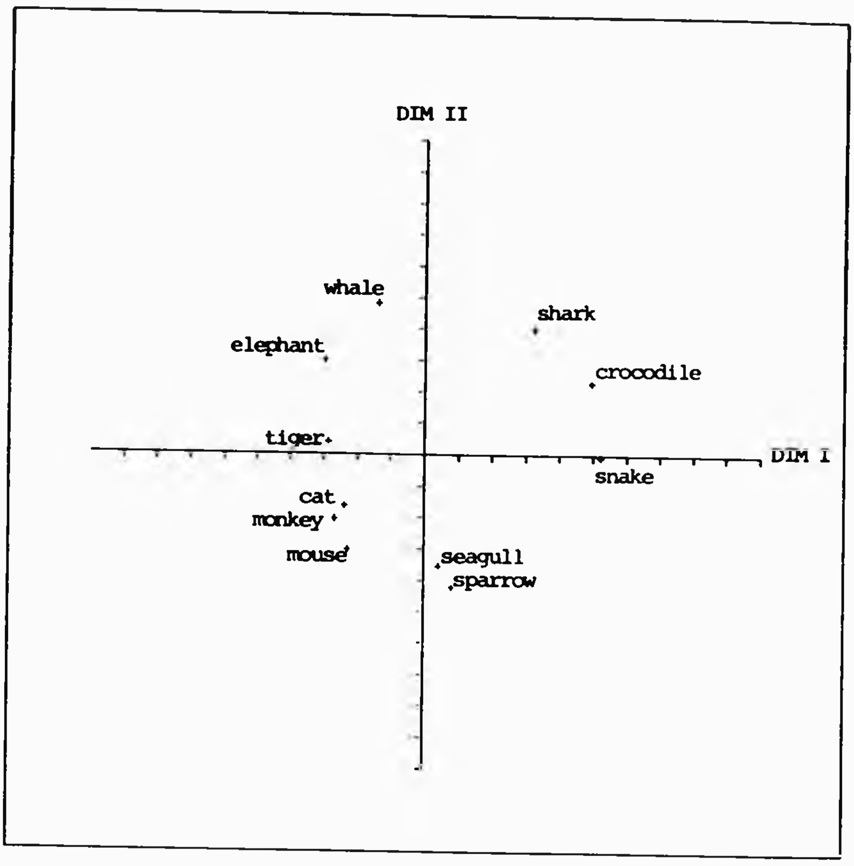


Figure 8.5: Dimensions I and II of the 3-dimensional SINDSCAL solution for 11 animals rated by experts with ordinary instructions and experts with 'expert' instructions (n=12)

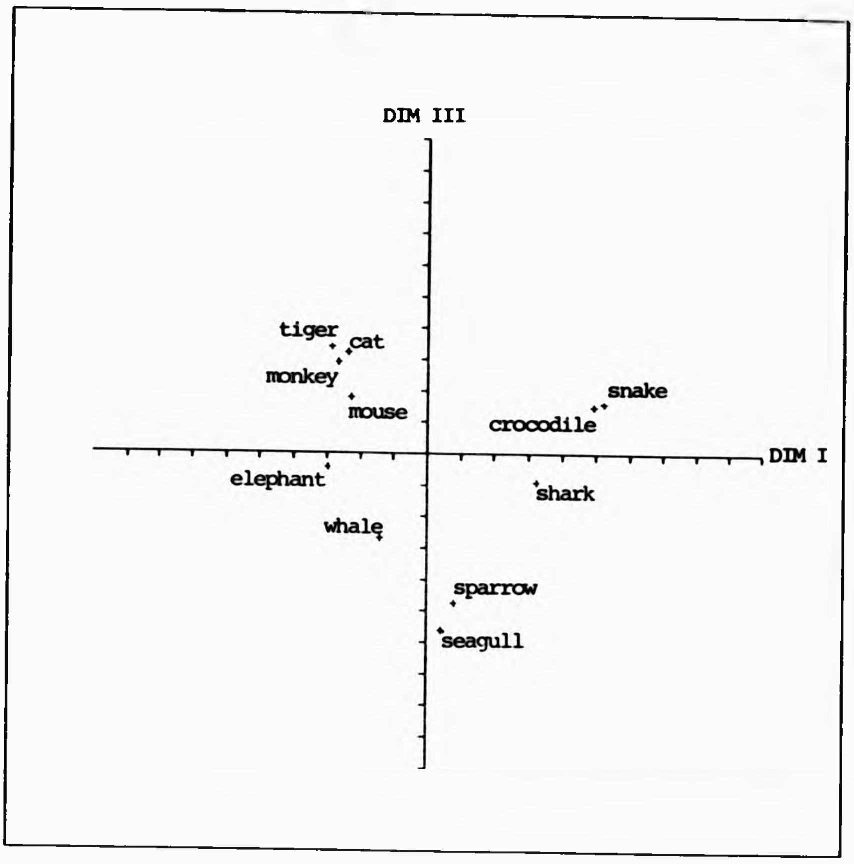


Figure 8.6: Dimensions I and III of the 3-dimensional SINDSCAL solution for 11 animals rated by experts with ordinary instructions and experts with 'expert' instructions (n=12)

Table 8.5: Correlations between the dimensions of three 3-dimensional SINDSCAL solutions for 11 animals:-
Solution A - children and adults
Solution B - children. adults and experts
Solution C - experts

		A			B			C	
		I	II	III	I	II	III	I	II
A	I								
	II	-.002	.						
	III	-.107	.318	.					
B	I	.008	-.956	-.558	.				
	II	.972	.004	.094	-.039				
	III	-.433	.026	.840	-.218	-.217			
C	I	.482	-.024	-.757	.182	.278	-.983		
	II	.062	-.933	-.556	.988	.021	-.225	.184	
	III	-.919	-.111	-.239	.170	.979	.103	-.173	.100

Relation between the Dimensions of the 3 Solutions

	A	B	C
Relatedness to man	I	II	III
Size	II	I	II
Niceness	III	III	I

a summary of the dimensions identified in each of the three solutions and how they related to one another. It appeared that when the data contained children's judgements, the Niceness dimension was in third place, but when experts alone were analysed, Niceness became the primary dimension. This was consistent with the experts' pattern of weights for the solution containing all forty-two subjects.

At this stage it was feasible to comment on the reproducibility of the configuration by examining the correlations in Table 8.5. Comparison of the configuration from children and adults (Solution A) with that from experts (Solution C) provided the best test because these solutions had no subjects in common. In both solutions Dimension II was Size and the correlation between these two dimensions was -.933. For children and adults Dimension I was Relatedness to Man/Habitat. This appeared as Dimension III for the experts, as indicated by the correlation of -.919 between these two structures. The least reproducible dimension was Niceness, which appeared as Dimension III for children and adults and as Dimension I for experts, but the correlation (r=-.757) was still substantial. The emergence of comparable dimensions from independent groups of subjects, coupled with a developmental rationale for the differential importance attached to dimensions, was evidence that the configuration represented a real and stable structure.

8.4.3 Expert versus Lay Instructions

Inspection of the mean subject weights for the two groups of expert subjects indicated that experts under instructions to behave expertly weighted the dimensions in decreasing importance from I to III, while their colleagues under lay instructions gave approximately equal weights to the dimensions. Mean weight ratios are shown in Table 8.6. As the data was rather small for discriminant analysis, separate tests were performed on subjects' weight ratios for each of the three dimensions. There were no significant differences between subjects on Dimensions I ($T10=1.220$, $p=0.250$) and II ($T10=0.457$, $p=0.658$). However on Dimension III, the difference between groups approached significance ($T10=2.006$, $p=0.073$). Since expertise had disrupted the categorical structure, which had been important to children in the previous solutions, it was feasible that encouraging expert modes of thought might further reduce its importance.

8.5 Taxonomic Structure in Experts, Novices and Children

8.5.1 Restating The Hypothesis

Final questions were whether the hypothesis that taxonomic structure increases with semantic development would be upheld by a comparison of expert structures with the rest; and whether expert instructions would increase taxonomic clustering within experts.

Using the method described in section 5.11, individual subject spaces were calculated for the solution containing all subjects (n=42) and for the solution with experts alone (n=12). Within-

Table 8.6: The relative importance of dimensions in the 3-dimensional SINDSCAL solution for 11 animals (rated by experts), represented by the mean relative weight for each group of subjects on each dimension (n=12)

	Dimensions		
	DI	DII	DIII
Experts (lay instruct)	.326	.352	.322
Experts (expert instruct)	.445	.319	.235

mammal, mammal-bird and mammal-reptile distances were calculated for subjects in each case. Semantic developments and expertise should lead to a decrease in within-mammal distances and to increases in both of the 'between' category distances.

8.5.2 Analysis of All Subjects

These analyses compared seven year olds, eleven year olds, lay adults, experts under expert instructions and experts under lay instructions. Mean within-mammal, mammal-bird and mammal-reptile distances for each of these groups are shown in Table 8.7. Analysis of variance of the within-mammals measure showed no differences between groups ($F_{4,37}=1.941$, $p=0.124$). However, groups were different on the mammal-bird measure ($F_{4,37}=2.794$, $p=0.040$) and on the mammal-reptile measure ($F_{4,37}=5.221$, $p=0.002$.

Inspection of the means revealed a surprising pattern. For the mammal-bird measure the mean distance for experts under expert instruction was totally at variance with the prediction. Instead of being the largest distance indicating the strongest differentiation between mammals and birds, it was the smallest. The experts under lay instructions displayed a smaller mammal-bird distance than adults and eleven year olds, which was also surprising.

This was in direct contrast to the means for mammal-reptile distances which were all in the predicted direction, increasing from seven year olds through to the experts under expert instruction, with the biggest difference between means

Table 8.7: Mean distances within mammals (WM), between mammals and birds (MB) and between mammals and reptiles (MR) derived from the 3-dimensional SINDSCAL solution for 11 animals rated by children, adults and experts (n=42)

	WM	MB	MR
7 year olds	.337	.649	.563
11 year olds	.349	.692	.589
Adults	.362	.699	.654
Experts (lay instruct)	.351	.686	.664
Experts (expert instruct)	.321	.620	.706

differentiating the children from the three groups of adults.

When the within-mammal, mammal-bird and mammal-reptile distances were inspected within subject groups, all but one group displayed a consistent pattern. For the majority, within-mammal means were the smallest, followed by mammal-reptile means, with mammal-bird means as the largest. The exception was the expert group under expert instruction, whose mammal-bird mean was smaller than their mammal-reptile mean. It was possible that for these subjects, the 'within' and 'between' cluster distances reflected their understanding of evolution and salient aspects of animal taxonomy. Birds evolved from reptiles and share warm blood and insulating covering with mammals and therefore, in an evolutionary sense, mammal-bird distances should be smaller than mammal-reptile distances. Other subjects may have displayed the reverse effect because they were influenced by the outstanding characteristic of flight, which distinguishes birds from land-based mammals and reptiles.

8.5.3 Analysis II: Taxonomic Structure amongst Experts

Within-mammal, mammal-bird and mammal-reptile distances were computed for the twelve expert subjects from the SINDSCAL solution containing experts only. Means for the two groups of subjects are shown in Table 8.8. The pattern of means was consistent with that obtained from the solution for all subjects, but there were no significant differences. Expert instructions appeared to have produced a tighter mammalian cluster and a configuration in which mammals were closer to birds than

Table 8.8: Mean distances within mammals (WM), between mammals and birds (MB) and between mammals and reptiles (MR) derived from the 3-dimensional SINDSCAL solution for 11 animals rated by experts (n=12)

	WM	MB	MR
Experts (lay instruct)	.328	.719	.672
Experts (expert instruct)	.295	.651	.726

reptiles. Experts with lay instructions displayed the alternative pattern in which mammals were closer to reptiles than birds.

CHAPTER 9
ANALYSIS OF FEATURES:
AN ALTERNATIVE TO MDS

9.1 Limitations of MDS Models

9.1.1 MDS Models of Semantic Development

When the results of Experiments I to V were evaluated and compared with other similar studies that had appeared in the course of this research, a concensus emerged. Although MDS models showed certain differences between semantic memory structures in children, adults and experts, they did not support the plausible hypothesis that maturation and education would create substantial differences between these groups in their representations of the meanings of common terms. Rather, all findings suggested that by the age of seven, the lexicon for common terms is almost as developed as it is ever going to be.

9.1.2 Constraints from the MDS Task

In Chapter 8, the possibility that the original sample of children and adults did not differ sufficiently in knowledge was explored by testing a group for whom superior knowledge of biological terms could be safely assumed. Although differences in the relative importance of dimensions were confirmed, children, adults and experts appeared to use the same limited set of three dimensions in discriminating amongst animals. Since dimensions like these have emerged from other studies which used different sets of animals, it was concluded that the choice of stimuli was not a major constraint on subjects' judgements.

The remaining possibility was that dissimilarity scaling is an inadequate tool for measuring individual differences in semantic structure. It may be that dissimilarity judgements impose a performance limit which cannot reflect the underlying competence of individuals. Furthermore, it is possible that a spatial model is not the best way to represent dissimilarity (or similarity) data. These arguments suggested that rather than changing the stimulus set, alternative methods of eliciting and representing data should be sought. The most straightforward approach was to elicit features which might be used to assess dissimilarity and to explore clustering techniques for representing data. Support for this strategy came from Tversky's (1977) comparison of distance and featural models of similarity, which concluded that featural approaches were superior.

9.1.3 The Validity of Distance Models of Dissimilarity

Tversky suggested that distance models are inappropriate for representing similarity. He argued that the fundamental requirements of a distance model impose severe constraints which are not always met by measures of perceived similarity. This may be seen with respect to the three axioms that underlie any distance model: the triangle inequality, minimality and symmetry.

Of these, the easiest to violate is symmetry because of the context in which similarity judgements are often made. Tversky suggested that a similarity judgement is an extension of a statement such as 'a is like b'. Such statements are directional in that the subject 'a' is judged with respect to the referent 'b'; and the choice of subject and referent is determined by the relative saliences of the objects. For example, we are more likely to state that 'an ellipse is like a circle' than 'a circle is like an ellipse'. In finding that an ellipse is more similar to a circle than a circle is to an ellipse, Tversky demonstrated that a variant is more similar to the referent or prototype than vice-versa.

In the context of animal terms, certain mammals may be prototypical, but the choice of prototype may differ with age. This could result in systematic biases in similarity judgements, depending on how the subjects are instructed and on how they interpret the task. For example, children may consider CAT to be the animal prototype, while adults may prefer TIGER. Therefore, unless otherwise instructed, a child would be asking 'how similar is a TIGER to a CAT' while an adult would say "how similar is a CAT to a TIGER". The extent to which the actual presentation of items on a page as one above the other or as a horizontal pair determines the directionality of judgements is unclear. Differences in distances resulting from these effects could not, in a standard scaling task, be distinguished from those attributable to other factors which might be expected to determine similarity.

The remaining axioms are not so problematic in the context of scaling animal terms. Minimality refers to the distance between an object and itself which must by definition be zero. In some contexts, for example, recognition experiments involving artificial stimuli, identical objects are not always identified as such; an object can be identified as another object more frequently than it is identified as itself. However, this is unlikely to affect stimuli such as animal terms.

Tversky did not refute the triangle inequality, but queried its validity using the example of perceived distances between Cuba, Jamaica and Russia. Cuba is similar to Jamaica in geographical location and similar to Russia in political terms, but Jamaica is not at all similar to Russia. It was suggested that the distance between Jamaica and Russia could exceed the sum of the distances between Jamaica and Cuba and between Cuba and Russia - a situation which violates the triangle inequality. Although the axiom is not refuted by such examples, Tversky argued that it should not be accepted as 'a cornerstone of similarity models'.

Tversky suggested an alternative model of similarity based on feature matching. The similarity between objects is expressed as a linear combination of the 'measures' of their common and distinctive features. The 'measure' or salience of a feature is determined by intensity relative to other stimuli and by relevence in the context of the objects being judged. Intensity is described in psychophysical terms by references to signal-to-noise, loudness, size and other similar factors. Of more importance in semantic similarity is relevance, here viewed as a 'diagnostic factor' resulting from the 'classificatory significance of features'. For example, if a group of objects do not differ in size, then size has a low diagnostic value.

Tversky concluded that models which use clusters and trees rather than distance can be used to discover the common and distinctive features of objects from similarity data. Alternatively, one can ask subjects to generate features of objects and identify common and distinctive features by examining the extent to which the feature lists overlap.

9.2 Experiment VI: Feature Generation

9.2.1 Introduction and Objectives

Tversky's formulation was of interest because it suggested a method of collecting data about semantic structures which overcame potential objections to the use of dissimilarity ratings. This study was designed primarily to elicit and compare the features used by children and experts in describing the eleven animals. A secondary objective was to compare a proximity measure derived from feature overlap between animals with the dissimilarity data obtained earlier.

Feature generation was investigated Gaden and Stradling (1980) who asked subjects aged seven, eleven and sixteen to imagine that they had to explain concepts to a 'man from Mars'. The concepts were unrelated (CAT, GAME, TRIANGLE, MARRIAGE and CHEATING). To classify features, the investigators used a general framework which identified utterances as one three types: the 'Material

Element' covered statements giving features or instances such as "Football, thats a game"; the 'Logical Content' consisted of statements relating the concept to other concepts, for example "Cats are animals"; and the 'Contextual Requirement' covered statements referring to the context of human activities and needs within which the concept has a function.

Some statements from each age group could not be classified, the seven year olds producing the highest proportion of unclassifiable ideas. However, the five concepts differed in their proportions of the three types of statements. Although there was a developmental trend away from the use of the Material Element and towards the Logical and Contextual, there was no age difference in the statements produced for CAT.

Although relevant, Gaden and Stradling's method of classifying features was not considered appropriate for the current study. First, as it was so general, it might fail to provide a detailed picture of differences between the subject groups. Second, it could not classify a proportion of statements, particularly those from the youngest children. Third, it was designed for investigating conceptual development in general terms, rather than with respect to a specific domain. Therefore, rather than impose pre-decided categories for features, the decision was taken to let the range of features elicited determine the classification strategy.

Turning to the secondary objective , this aimed to replicate

Tversky's finding that similarities computed from feature overlap have high correlations with rated dissimilarities. If this could be shown for both experts and adults, it would imply that a dissimilarities rating task can reflect individual differences in conceptual structure obtained by the more detailed method.

9.2.2 The Subjects and Task

Interviews were conducted with two new groups of subjects. These were eight children aged nine to ten years old and eight doctors aged between twenty-five and thirty-four. The children's ages were all within a year of each other around their tenth birthday; their mean age was ten years and three months.

Subjects were asked to imagine either that they were helping to design a robot or that they had met an intelligent creature from outer space who knew nothing about life on earth. What would the robot/spaceman have to know in order to recognise things on earth? For example, how would it recognise a cup?

Following this example, subjects listed features for each of the eleven animal terms and for a furtherfive terms: MAMMAL, BIRD, REPTILE, FISH and ANIMAL. The children were interviewed and their responses written down, while adults provided written lists. Given the variability in performance of the primary school children in earlier tasks, it was considered important that they be encouraged and prompted throughout the task.

9.2.3 Analysis of Features

Fortunately, the data set was small and could be handled manually, with assistance from a spreadsheet package, LOTUS 1-2-3, which was used to hold and sort the individual features.

Separate analyses were carried out for animal and taxonomic terms. The procedure was as follows. First, a spreadsheet was created for each group of subjects containing the raw features per animal and per subject. An alphabetic sort of features within animals was examined in order to develop a coding system. When this had been completed for both groups of subjects, a joint coding scheme was applied to each set of raw data.

Inevitably, there was a major subjective component in coding features, which could only be counteracted by asking independent judges to perform the coding procedure. As this was not feasible, subjectivity was minimised by repeating parts of the task after a substantial interval and comparing conclusions. In addition, the coding scheme was deliberately detailed, so that it did not represent the raw data in extremely reduced form.

The frequencies of coded features were obtained for each group of subjects. These represented the number of times a feature was elicited across the eleven animals within the subject group. Following the frequency count, features were coded a second time to form Feature Categories so that the two groups of subjects could be compared. The number of features in each Feature Category for each group of subjects is shown in Table 9.1. The features within each Feature Category and their frequencies are shown in Appendix 3.

Since experts produced more features than children, the number of features per Feature Category was also expressed as a percentage of each groups' total number of features. Chi-square tests were performed using the percentages for Feature Categories with large frequencies. Only the Superordinate category produced a significant result (Chi-square=5.35, p<0.05). Superordinate features formed 16.3% of experts' data, but only 5.5% for the children. The data also indicated that Movement might be less important to the experts (3.0%) than it was to the children (9.5%).

This procedure was repeated for the five taxonomic terms. Although the coding scheme produced for the eleven animals inevitably influenced the treatment of features produced for the taxonomic terms, it was sufficiently detailed to avoid masking different characteristics in the taxonomic data set. With one exception, the same coding scheme was found to be appropriate for both data sets. The difference between animal and taxonomic features was that the children produced animal instances amongst their features for taxonomic terms, particularly in response to REPTILE, while the experts, without exception, gave abstract descriptions of taxonomic terms. Features for taxonomic terms are listed by Feature Category with frequencies in Appendix 3.

Table 9.1: Occurences of features in children's and experts' descriptions of 11 animals, summarised by Feature Categories

	Feature Category	10 year olds No. Features	%	Experts No. Features	%
1.	SIZE & SHAPE	53	13.2	82	16.7
2.	COLOURING	27	6.7	44	8.9
3.	SKIN TYPE	23	5.7	16	3.2
4.	MOVEMENT	38	9.5	15	3.0
5.	HABITAT	43	10.7	46	9.3
6.	ANATOMY	107	26.7	117	23.8
7.	PHYSIOLOGY	1	0.2	4	0.0
8.	COMPLEX PHYSIOLOGY/ BEHAVIOUR	21	5.2	18	3.7
9.	NOISE	14	3.5	13	2.6
10.	DIET	10	2.5	25	5.1
11.	FEROCITY	35	8.7	24	4.9
12.	REPRODUCTION	-	-	5	1.0
13.	USES	7	1.7	3	0.6
14.	SUPERORDINATES	22	5.5	80	16.3
	TOTAL	401		492	

Table 9.2: Occurences of features in children's and experts descriptions of 5 taxonomic terms. summarised by Feature Categories

	Feature Category	10 year olds No. Features	%	Experts No. Features	%
1.	SIZE & SHAPE	7	6.7	4	2.2
2.	COLOURING	5	4.8	-	-
3.	SKIN TYPE	5	4.8	14	7.8
4.	MOVEMENT	16	15.2	20	11.2
5.	HABITAT	16	15.2	12	6.7
6.	ANATOMY	21	20.0	45	25.1
7.	PHYSIOLOGY	6	5.7	30	16.6
8.	COMPLEX PHYSIOLOGY/ BEHAVIOUR	9	8.6	4	2.2
9.	NOISE	5	4.8	-	-
10.	DIET	1	0.9	1	0.5
11.	FEROCITY	6	5.7	5	2.8
12.	REPRODUCTION	2	1.9	21	11.7
13.	USES	4	3.8	1	0.5
14.	SUPERORDINATES	2	1.9	22	12.3
	TOTAL	105		179	

The number of taxonomic features in each Feature Category for each group of subjects is shown in Table 9.2. Again the percentage of each groups' total number of features was calculated. The only significant chi-square in this instance was for the category, Physiology (chi-square=5.328, $p<0.05$), but tests were unnecessary to conclude that experts produced more features on Reproduction (11.7%) than children (1.9%) and that the experts' preference for Superordinate features (12.3%) again exceeded the childrens' (1.9%).

The most obvious difference between the two groups of subjects was the experts' greater use of Superordinate features and scientific language. In some cases experts and children expressed similar concepts using difference terminology. For example, experts produced 'four-limbed' in a similar context to the children's 'four-legged'; and experts used 'respires', for which the children's equivalent was 'breathe'. It must be stressed, however, that these features are similar, not identical: undoubtedly, the experts' connotations for 'respires' were considerably more complex than the children's understanding of 'breathe'.

9.2.4 Similarity from Feature Overlap

Tversky reported a study in which similarity between twelve vehicles (for example, BUS, CAR, TRUCK) computed from feature overlap was correlated with rated similarity. Using the simplest measure of feature overlap, defined as the number of features in common, the correlation between overlap similarity and rated similarity was 0.68. The complementary simple overlap measure is the number of distinctive features. This measures lack of overlap, being the sum of features that a pair of objects do not share. The correlation between this and rated similarity was -0.36. The multiple correlation between rated dissimilarity and common and distinctive features was 0.72. When the frequencies of features were taken into account in a more complex model, the multiple correlation improved to 0.87.

As the simple measure of feature overlap, the number of common features, correlated well with rated similarity, it was considered adequate in this context. It had the added advantage of being simple to compute. Within each group of subjects who generated features, a list of coded features was compiled for each of the eleven animals and the number of common features held by each pair of animals was counted. This resulted in a similarity matrix for each group. These were correlated with mean rated dissimilarity matrices for each of the subject groups studied earlier. These were seven year olds, eleven year olds, adults, experts under lay instruction and experts under expert instructions. The correlations are shown in Table 9.3.

Looking first at the correlations between children's feature overlap similarites and the dissimilarity matrices, the highest correlations were with adults (-0.756), with experts under expert instructions (-0.742) and with experts under lay instructions (-.630). The lowest correlation was with the age group closest

Table 9.3: Correlations between dissimilarity ratings for 11 animals and similarities computed from feature overlap

		DISSIM Age 7	DISSIM Age 11	DISSIM Adult	DISSIM Expt (lay)	DISSIM Expt (exp)	SIM Age 10
DISSIM	Age 11 (n=10)	.709					
DISSIM	Adult (n=10)	.784	.719				
DISSIM	Expert lay (n=6)	.505	.558	.802			
DISSIM	Expert exp (n=6)	.676	.650	.913	.871		
SIM	Age 10 (n=8)	-.609	-.537	-.756	-.630	-.742	
SIM	Expert (n=8)	-.410	-.492	-.707	-.604	-.673	.786

to themselves, the eleven year olds (-0.537). Therefore, the children's feature data correlated more with the three sets of adult dissimilarities than with either of the children's dissimilarities and these correlations were of the same order as obtained by Tversky (0.68) when he compared feature-overlap similarity with rated similarity.

A similar pattern was obtained for the experts' feature data, but the correlations were slightly lower. In this case, the highest correlation was again with adults (-0.707), but the lowest was with the seven year olds (-0.410). This was a predictable result and the correlations with comparable subject groups were all of the order obtained by Tversky.

It was noticeable, however, that a high correlation in Table 9.3 was between child and expert feature matrices - more correlated with each other than with any of the dissimilarity matrices (0.786). Correlations amongst the dissimilarities subject groups are also shown in Table 9.3. These indicated consistency in that correlations amongst the adults groups were greater than those between adults and children, but the two groups of children were not more correlated with each other than with other groups.

The high correlation between the children's and experts' feature matrices may have resulted from bias in the coding of features, but as overall impression from the task was that feature lists produced by the two groups were similar, this was thought unlikely. The main conclusion from this study was that similarities for children and experts derived from the feature overlap method produced matrices that were more similar to each other than they were to rated dissimilarity data from comparable subjects. This effect appeared to be stronger for children than for experts, indicating that the change of task from dissimilarity rating to feature generation enabled children to display a different and probably more complex semantic structure.

9.3 ADDTREE: An Alternative to SINDSCAL

9.3.1 Spatial Versus Tree Representations

The possibility that the SINDSCAL analysis provided an inadequate representation of the subjects' dissimilarity judgements was given credibility by the work of Sattath and Tversky (1977) in which spatial and tree representations were compared. In a subsequent paper, (Pruzansky, Tversky and Carroll, 1982), it was suggested that spatial representation may be more suited to perceptual or psychophysical data, while tree structures may be more appropriate for representing conceptual data.

9.3.2 MDS and Categorical Structures

MDS has often been compared with categorical techniques such as Johnson's Hierarchical Clustering. In most instances, they are applied jointly in order to describe structures in which both dimensions and clusters are interpretable (for example, Shepard (1974); Kruskal (1977); Stockdale and Hale, 1981). Shepard thought that the clustering of animals in a 2-dimensional space provided a more 'compelling' interpretation than their projections on to the dimensions, but Kruskal stressed the complementary nature of the two techniques when data contains error, suggesting that with error-free data either clustering or MDS will be appropriate, but that with error, both techniques can provide an interpretable model because they are sensitive to different aspects of the data.

Sattath and Tversky (1977) described ADDTREE, a clustering technique which is less restrictive on data than the better known Johnson's method. Traditional clustering imposes a condition known as the ultrametric inequality which means that any three points must form an equilateral triangle or an isosceles triangle with a narrow base (two distances must be equal and equal to or greater than the third). For two clusters this implies that all intra-cluster are smaller than inter-cluster distances and that all inter-cluster distances are equal. Data do not normally satisfy this requirement, therefore a less restrictive method is likely to provide a better representation of real data.

Sattath and Tversky compared ADDTREE with a spatial method, Smallest Space Analysis, as alternative representations of Henley's similarity data for thirty mammals. The ADDTREE result was most impressive and highly interpretable in comparison with the typically unclear dimensional structure derived from SMALLEST SPACE ANALYSIS; and, in addition ADDTREE provided the better fit in terms of the product-moment correlation and stress. These findings held not only for Henley's data, but also for two other data sets involving similarity ratings of Swedish letters of the alphabet and occupations respectively.

Henley's data displayed clear groups and there was also the suggestion that the vertical ordering of the ADDTREE solution represented size, the major dimension of scaling solutions with animal terms. Of major interest is the claim that ADDTREE arcs (the lines joining items) represent features and that the length of the arc indicates the importance of the feature. For example, in Henley's data the mammals divided first into herbivores, carnivores, apes and rodents. Of these groups, the apes displayed the longest arc, while the remaining groups subdivided again. This indicated their distinctiveness as a group relative to the other mammals.

Pruzansky et al (1982) compared the performance of ADDTREE with that of KYST, an MDS procedure, on artificial data generated either by a plane or by a tree and also on twenty data sets covering a wide range of the kinds of stimuli to which MDS is commonly applied. The popular Henley mammal proximities were again included. The authors aimed to compare tree and spatial structures both on goodness-of-fit and on their handling of certain characteristics of similarity data. In particular, they identified skew and elongation as potential diagnostic characteristics which could indicate the best representation for particular data sets.

Skew was important because they noted that tree structures tend to generate many large distances and fewer small distances, while planes produce predominantly small distances and few large distances. Elongation refers to the shapes of triangles generated by the data and it was shown that tree data produce a higher proportion of elongated triangles than are obtained from planes. Applying these tests across the range of data sets, they concluded that a tree representation was more appropriate when stimuli were non-factorial, conceptual and selected as familiar and good examples of their category. Where stimulus sets were perceptual and structured factorially, the MDS representation was superior.

9.3.3 ADDTREE Analysis of the MDS Dissimilarities

The dissimilarity data collected in Experiment I appeared better suited to tree representation on two grounds. First, it fell into the category of conceptual data, concerning stimuli that were good examples of their category. Second, descriptive statistics for the mean dissimilarities from each group of subjects indicated that the data were skewed and this was most marked for the youngest children. Therefore it was possible that ADDTREE might provide a better representation of any differences between the five groups of subjects (seven year olds; eleven year olds; adults; experts under lay instructions and experts under expert instructions). Unlike SINDSCAL, ADDTREE carries out a separate analysis for each dissimilarity matrix. Therefore the mean dissimilarity matrix for each group of subjects was analysed to produce a series of trees, each representing an average subject.

In all cases the proportion of variance explained (%VAF) was high, ranging from 73.7% for the adults through to 88.6% for the seven year olds. This compared very favourably with the SINDSCAL figures, indicating clear superiority of ADDTREE, particularly for the children's data. The %VAF provided by ADDTREE for each group is shown with the comparable figures obtained from the three SINDSCAL analyses in Table 9.4. The SINDSCAL figures were the square of the mean correlation with the model for each subject group. It may be seen that the seven year olds' data were particularly badly fit by SINDSCAL when the analysis included experts as well, indicating that a single model could not adequately fit data deriving from extremes of expertise.

However the superior fit of ADDTREE had to be balanced against the dificulty of interpreting trees and comparing them with one another. The main differences across the groups of subjects were the number of main categories; the salience of these as indicated by the lengths of arcs; and the emergence of the biological taxonomy.

The seven year olds' tree is shown in Figure 9.1. The two main categories distinguished animals with fur or feathers from those with scales or bald skins. The longest arc joined SEAGULL to SPARROW, indicating that these items formed a very distinctive group. The arc is the horizontal line identifying birds as a

Table 9.4: Percent variance accounted for (%VAF) by SINDSCAL and ADDTREE analyses for children, adults and experts

	3-dimensional SINDSCAL n=30	n=42	n=12	ADDTREE
Age 7	50.7	47.7		88.6
Age 11	56.1	54.8		82.5
Adults	65.6	66.6		73.7
Experts (lay)		63.4	65.4	79.4
Experts (exp)		68.4	70.9	84.2

group.

The eleven year olds produced three main categories as shown in Figure 9.2, but these were less interpretable . They appeared to represent animals that were large, small and water-based/reptilian. The main difference between this and the seven year olds' tree seemed to be in the movement of ELEPHANT away from a water/reptile cluster to a group of mammals.

The first three divisions in the adults' tree, shown in Figure 9.3 represented furriness, baldness and birds. Again ELEPHANT had moved, this time back to the water/reptile group, thereby necessitating the baldness label. The notable characteristic of this tree was that it showed the first identification of birds as a distinctive category at the initial level of division.

The ADDTREE result for experts under lay instructions is shown in Figure 9.4. This displayed an almost perfect taxonomic structure. In addition to birds, which had been evident in the adults' tree, these subjects also identified land mammals at the primary level of division and they were the first to put reptiles into a single cluster. However, WHALE remained steadfastly in the 'fish' category, although according to the length of arc, WHALE and SHARK were not a distinctive pair. The experts under expert instruction produced a similar tree, shown in Figure 9.5 but both WHALE and ELEPHANT violated the taxonomic structure. ELEPHANT had moved out of the mammal group and back to the water/reptile category, but again this was not a distinctive group.

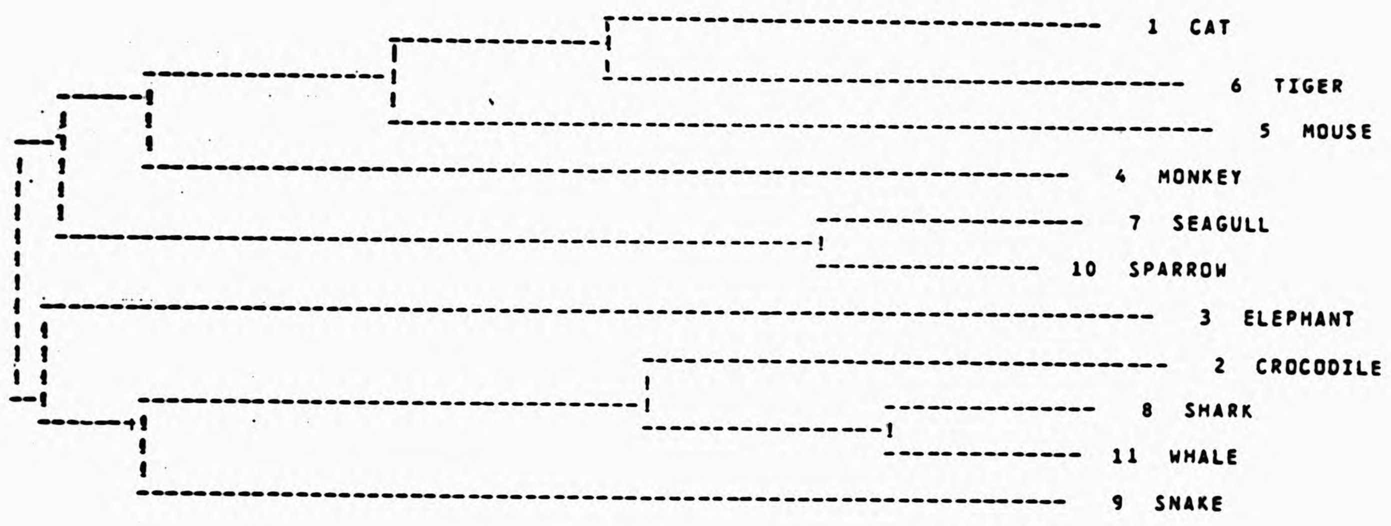


Figure 9.1: ADDTREE solution for 11 animals from dissimilarity ratings by 7 year olds

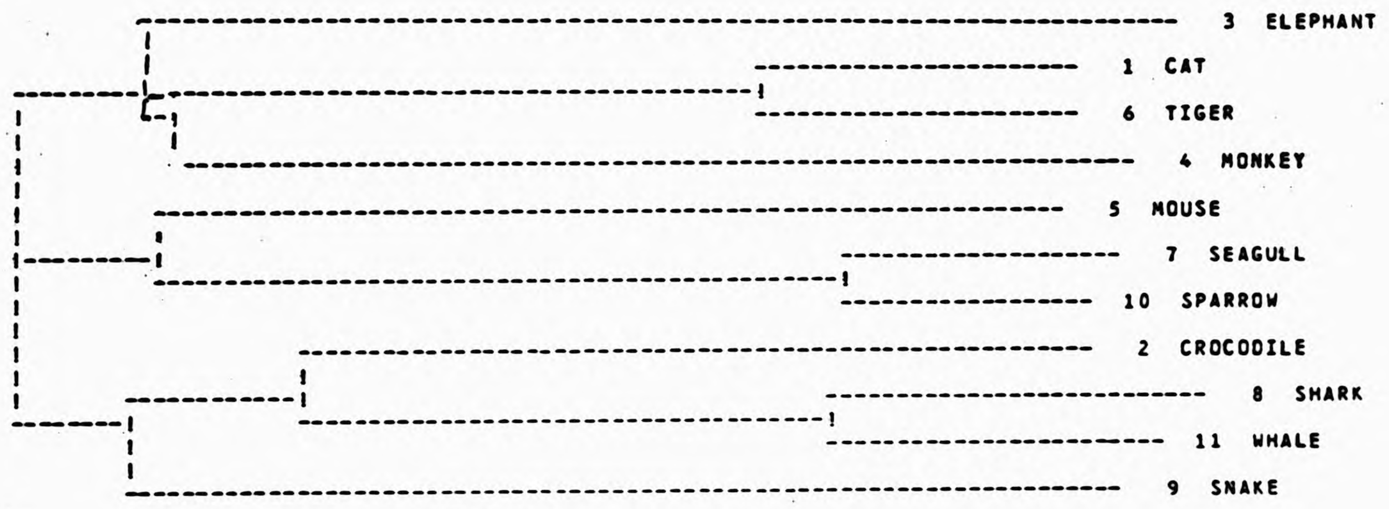


Figure 9.2: ADDTREE solution for 11 animals from dissimilarity ratings by 11 year olds

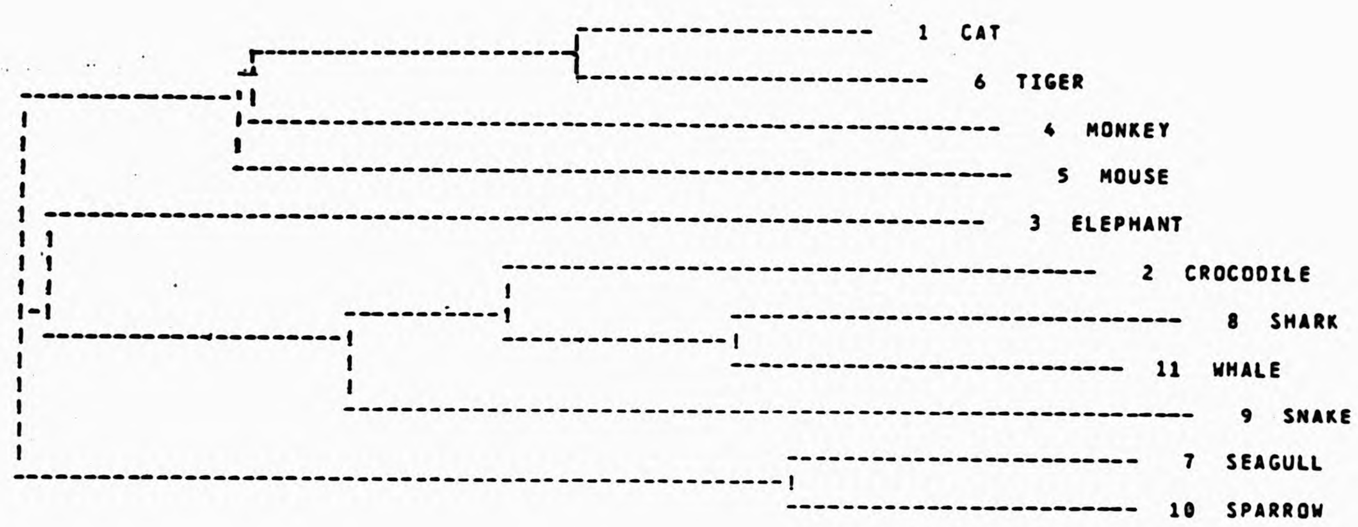


Figure 9.3: ADDTREE solution for 11 animals from dissimilarity ratings by adults

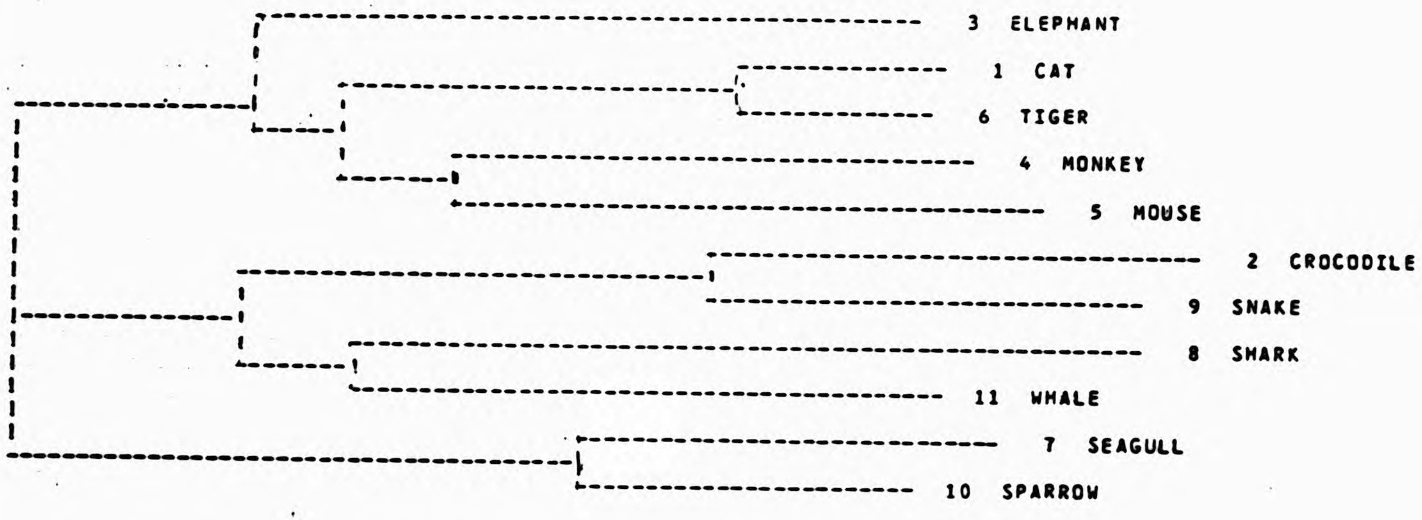


Figure 9.4: ADDTREE solution for 11 animals from dissimilarity ratings by experts with ordinary instructions

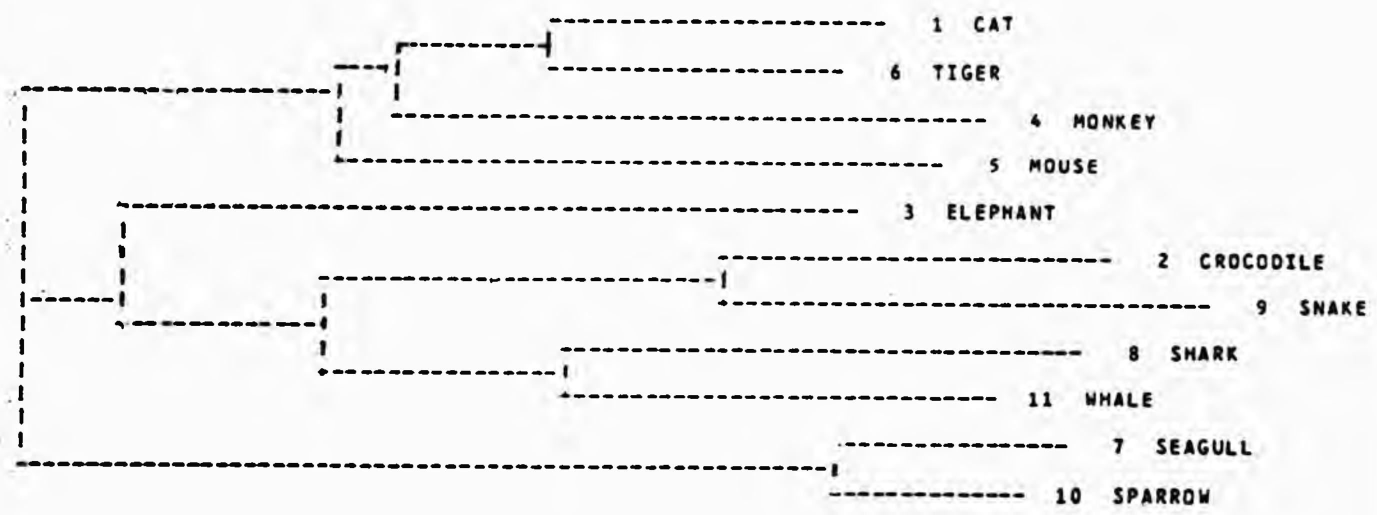


Figure 9.5: ADDTREE solution for 11 animals from dissimilarity ratings by experts with 'expert' instructions

9.3.4 ADDTREE Analysis of the Feature Overlap Similarities

Children's and experts' similarities between the eleven animals derived from the feature overlap measure were also submitted to ADDTREE, with interesting results. The children's tree had a clear taxonomic structure, dividing initially into reptiles, mammals/birds and the ubiquitous 'fish' group comprising SHARK and WHALE. This is shown in Figure 9.6. Data from subjects of a comparable age in the MDS task had not displayed any taxonomic groups when analysed by ADDTREE.

The experts' feature tree, shown in Figure 9.7, was paradoxically less clear, but was unique in being the only structure to group WHALE with mammals in preference to SHARK. In this tree WHALE and ELEPHANT formed a group at the initial level of division. However, reptiles were not a distinctive group and birds were not identified as a category. The three main divisions appeared to be large mammals, ferocity and fur/feathers.

9.3.5 A Discussion of the ADDTREE Findings

The much improved fit of the ADDTREE model in comparison with SINDSCAL suggested that trees provide superior representations for this type of data, particularly when the data are obtained from children. ADDTREE analyses indicated the emergence of taxonomic structure with age and expertise; and they confirmed the importance of a variety of features identified in the earlier SINDSCAL analyses and property-fitting. However, it was noticeable that the non-taxonomic ADDTREE groupings suggested

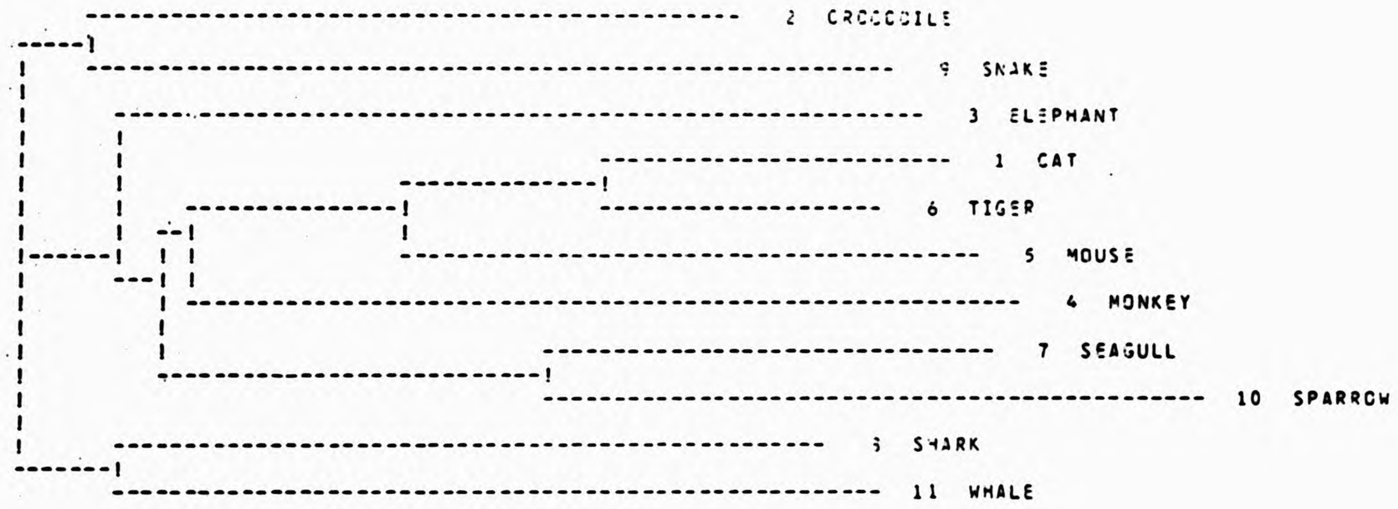


Figure 9.6: ADDTREE solution for 11 animals from feature overlap similarity in 10 year olds' feature lists

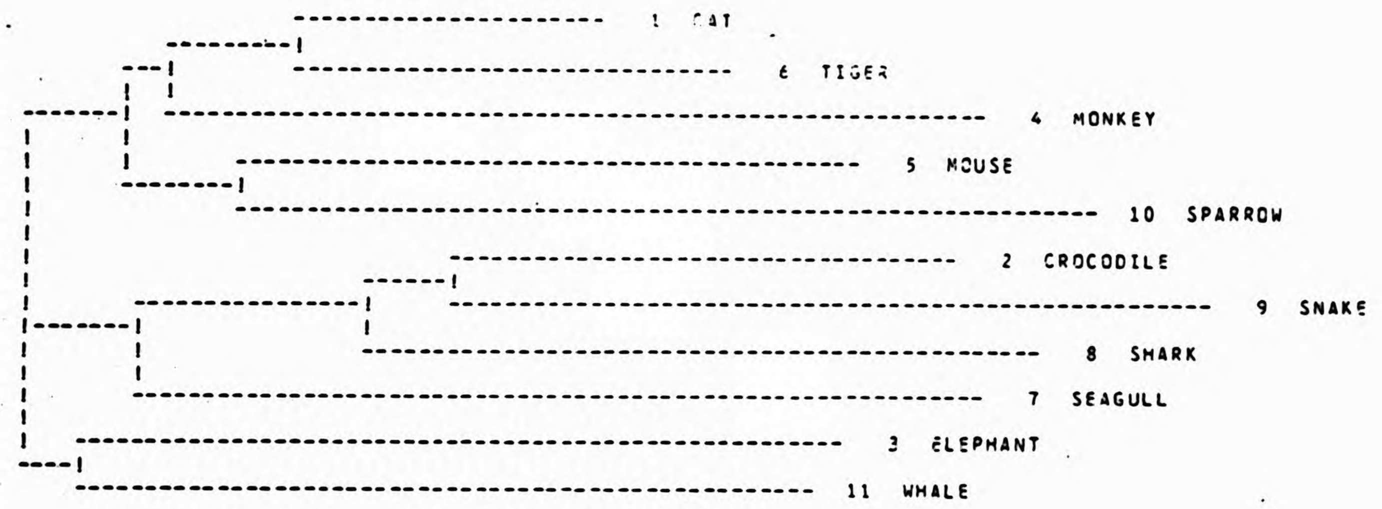


Figure 9.7: ADDTREE solution for 11 animals from feature overlap similarity in experts' feature lists

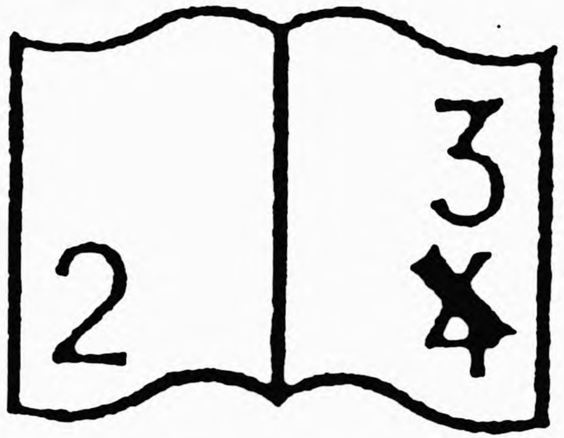
primarily concrete interpretations, rather than the Nice/Nasty distinction indicated by property fitting to the SINDSCAL space.

It was interesting to note that, as in the SINDSCAL analyses, ELEPHANT wandered between the land mammal and water/reptile groups. The separate ADDTREE analyses for each group showed these movements in detail. For eleven year olds and experts with ordinary instructions, ELEPHANT went with land mammals, while for the remaining groups it fell into the water-reptile category. Since the eleven year olds did not display a coherent mammal category - MOUSE was with the birds - their positioning of ELEPHANT with land mammals was probably based on size rather than taxonomy.

By a similar argument, although both seven year olds and experts behaving expertly put ELEPHANT into the water/reptile group, they are unlikely to have done so for the same reasons. The seven year olds' tree showed a primary division between animals with fur or feathers and animals with bald skins, while the experts displayed clear taxonomic groups with the exception of ELEPHANT and WHALE. These results suggested that amongst experts and possibly also adults, organisation based on similarity of features conflicted with awareness of formal taxonomy.

PAGINATION ERROR

After P 226.
Pages re-start at
P 222.



CHAPTER 10

THE DEVELOPMENT OF SEMANTIC MEMORY: CONCLUSIONS

10.1 Spatial Models, Semantic Development and Expertise

Conclusions from this research fell into three domains. First and most obvious were answers to the study's original question, concerning the extent to which a spatial model of semantic memory would represent predicted trends in semantic development. The second domain consisted of some findings about semantic development itself. In a very basic sense these were observations about what develops and, therefore, what phenomena theories need to explain. The third domain centred on the nature of expertise and how it relates to semantic development. Differences between novices and experts were not the original focus of this research, but with hindsight, acknowledging them highlighted that early models of semantic memory were limited by assuming expert structures. Expertise also provided a useful perspective for integrating research in semantic memory and cognitive development over the past 20 years.

10.2 Can a Spatial Model Represent Semantic Development?

Studies of language acquisition and development during the 1970s identified developmental trends indicating ways in which semantic structures in children and adults might differ. Children's concepts appeared to be primarily concrete and perceptual, while adult's representations contained both concrete and abstract information (Anglin, 1970). A consequence of this was that children's concepts were likely to be thematically organised in contrast with adult's more taxonomic structures (Smiley and Brown, 1979). In addition, studies using factor analysis to investigate the development of the Evaluative, Potency and Activity dimensions identified by Osgood et al (1957) indicated that age was associated with solutions of higher dimensionality representing greater discrimination between semantic attributes (for example, Saltz et al, 1975).

From these findings it was possible to predict how a spatial model of semantic relations between words, which had proved useful in representing adult semantic structure, might fare in modelling developmental differences between seven year old, eleven year olds and adults. The individual differences scaling model, SINDSCAL, was used to represent subjects' judgements of the dissimilarity amongst animal terms and the predictions were that age would be associated with greater dimensionality, greater reliance on abstract dimensions and with taxonomic rather than thematic structure.

The scaling model did not represent all of the predicted developmental trends. The dimensions were identified by property fitting as Relatedness to Man/Habitat, Size and Niceness. Adults appeared to attach greater weight to the higher dimensions, but this trend in the subject weights was not statistically significant. Correlations between properties of animals and the dimensions showed that, for a subset of properties, the more abstract properties correlated better with the last dimension, Niceness, but in the absence of an association between age and dimensionality, this could not be taken as confirmation of a developmental trend from concreteness to abstractness.

The scaling model achieved limited success with predictions about taxonomic structure. The seven year olds demonstrated less differentiation between mammals and reptiles than evidenced by both eleven year olds and adults. However, age groups did not differ in terms of the strength of mammal clustering, the differentiation between mammals and birds or on some pairwise measures of thematic clustering.

The model was at its least successful in predicting property verification times. An abstract property took longer to verify than a concrete property, but concreteness did not interact with age. These weak effects may have been partly due to noisy data from the children who, unlike the adults, did not display the symbolic distance effect which has been shown elsewhere for dimensional judgements (Moyer and Bayer, 1976). This is the finding that it takes longer to judge that objects possess a property if they have it in intermediate amounts than if they possess it to an extreme degree. Being a U-shaped function, the symbolic distance effect should relate to a semantic space via an ideal point model such as provided by PREFMAP for fitting preferences to multidimensional spaces, but this was not supported. This particular analysis was carried out with a preliminary INDSCAL version of the model, but it is unlikely that

INDSCAL was grossly inadequate because it performed reasonably well in predicting rating data. A combination of poor data and the relative obscurity of the PREFMAP programme may have contributed to this more than the adequacy of the semantic space.

The scaling model may have failed to represent developmental change convincingly, but it was an adequate model in two respects. First, it predicted subjects' judgements on a number of animal properties like size, furriness and dangerousness. Second it replicated dimensions of relatedness to man, size and ferocity found in several other studies. Furthermore, of these studies, some of the non-developmental ones had demonstrated relationships between inter-item distances in their scaling models and adults' reaction times to make semantic judgements. Therefore the current model could not be taken as a poor reflection of all aspects of semantic structure. The options at this stage were that a distance model was theoretically inadequate, or that the judgemental task underlying it was too insensitive to represent developmental changes or that the predicted trends themselves did not exist.

To explore the last possibility a second model was created by adding data from experts; conceptual change from the age of seven to adulthood could be subtle or non-existent, but it should be evident in the transition to expertise. A comparable dimensional structure was obtained and this time individual differences between groups emerged, with children weighting the second dimension highly while experts attached importance to the third.

Developmental differences in taxonomic clustering were also apparent in this solution.

The dimensions identified in a subsequent model based on data from experts only were similar to those in the two ealier models; and, as predicted, the abstract Niceness dimension which had been in third place in the first model, moved to first place in the experts' model. The Relatedness to Man/Habitat dimension which had been prepotent for children and lay adults was the least important dimension for experts. Correlations between dimensions in the three models showed that the most robust dimension was Size. Relatedness to Man/Habitat and Niceness were both evident across all solutions, but were disrupted by changes in the relative locations of ELEPHANT, WHALE and the bird and reptile items. However, the experts' model performed better in terms of variance accounted for than the models containing children's judgements. This suggested that a single scaling model for all subjects might be inappropriate because differences between subject groups could represent substantial variation in fit rather than structure.

The overall conclusion was that a spatial model derived from MDS could represent broad developmental differences when these were maximised as exemplified by seven year olds and adult experts. However, it was possible that the dissimilarity judgement task used to collect data for scaling did not allow subjects' the opportunity to demonstrate the full range of their knowledge.

This idea was explored via a feature-generation task in which features of the same animals were elicited from a group of ten year olds and from experts.

The feature generation task produced considerable detail about the different ways in which children and experts choose to describe animals. As such it was a source of information about what changes occur in semantic development and how experts differ from novices. These aspects of the results are described later in this chapter. Of relevance at this point is how similarities based on features compared with the rated dissimilarities that had been analysed with SINDSCAL.

Following Tversky (1977), a measure of similarity between animals was computed, based on the extent to which their features overlapped. The ten year olds' and experts' feature-based similarities were then correlated with the dissimilarities obtained from the original groups of children, adults and experts. This simple exercise indicated that the two tasks were revealing different things: ten year olds' feature-based data correlated better with dissimilarities data from adults and experts (r ranged from-.630 to -.756) than with those from their peer group, the eleven year olds (r=-.537); and they correlated best of all with experts' feature-based data (r=.786). Correlations between experts' feature-based data and adults' and experts' dissimilariies ranged from -.604 to -.707. Therefore the dissimilarities task had imposed a performance ceiling, particularly on children's behaviour and to that extent, the traditional method of obtaining data for a scaling model was not appropriate for children.

10.3 Dimensional Maps or Feature Trees?

An unexpectedly important part of this research was the comparison of MDS (SINDSCAL) and clustering (ADDTREE) as methods for representing individual differences in semantic structure. Hierarchical clustering had been employed as an aid to interpreting the original SINDSCAL solution, but was not at that stage viewed as offering a potentially superior model. However, Pruzansky et al (1982) showed that cluster or tree structures can provide a better fit when data reflect conceptual judgements and have particular distributional characteristics. Therefore both dissimilarities and feature-based similarites were modelled using ADDTREE, a non-restrictive clustering technique (Sattath and Tversky, 1977).

For all data sets ADDTREE produced the better fit in terms of variance accounted for, particularly for the children's data. Although SINDSCAL, aided by property fitting, produced three interpretable dimensions, the ADDTREE analyses provided a richer, if potentially less systematic, description of the differences attributable to age and expertise. The ADDTREE representations were superficially better than SINDSCAL in suggesting the variety of important features and the differences between groups of subjects in the salience of taxonomy. In these ways, ADDTREE appeared superior to MDS in recovering the underlying structure, thereby supporting the view that a tree structure is the more appropriate model for representing the relations between objects selected from natural semantic categories.

However, three aspects of ADDTREE made it less attractive than SINDSCAL. First, each group of subjects generated a separate ADDTREE solution, rendering comparisons between groups difficult in contrast with SINDSCAL which can represent all subjects within the same multidimensional space. Second, the application of property fitting to the SINDSCAL solutions indicated a Niceness dimension which was not readily suggested by the ADDTREE structures. Indeed, without something comparable to the property fitting procedure, which provided a direct and objective method of interpreting SINDSCAL's dimensions, it was difficult to interpret the ADDTREE results with confidence. Third, the SINDSCAL approach seemed the easier to test: it offered a systematic way of examining the concrete-abstract progression via the varying fit of unidimensional scales across dimensions of different weight to different subjects; and hypotheses relating property verification latencies to the locations of objects in continuous space are attractive, although they received limited support in this research. It is possible that ADDTREE could be treated in a comparable way, but this was beyond the scope of the current project.

To summarise, ADDTREE could not provide an overall summary of developmental changes and did not offer obvious ways of testing developmental hypotheses, but it highlighted important aspects of the data. First, it showed that tree representation provided a better fit for all subjects, but that the gain for children outstripped that of the other groups. Second, the emergence of taxonomic organisation was evident. Third, although it was apparent that ADDTREE was no better than SINDSCAL at coping with the peripatetic ELEPHANT, tree representations made this animal's movements around the domain clearer with respect to each group of subjects, so highlighting the conflict between similarity and taxonomy.

10.4 Development of Animal Concepts

10.4.1 Similarities between Adults and Children

One important conclusion to emerge from this study and others (Storm, 1980) was that the natural semantic category of animal terms is well formed by the age of seven and that subsequent changes probably owe more to formal education than to cognitive development. Evidence for this came first from the SINDSCAL model which showed that although children aged seven and adults appeared to weight semantic dimensions differently, this effect was not robust enough to attain statistical significance. When judgements from a group of experts were analysed with data from the children and adults, the same dimensions emerged and there was evidence of significant individual differences in dimension weights.

This conclusion was further supported by a comparison of the features generated by children and experts when asked to describe how they would identify animals in order to teach a 'spaceman' or a 'robot'. Although the children who generated these features were aged ten, the features were very similar to those generated by six and seven year olds in the pilot experiments. Ten year olds and experts were similar in their use of terms which described physical features, habitats and behaviours of animals, differing mainly in their use of superordinate terms and scientific language. For example, where experts and children diverged in their production of features, this was in the use of a generic name like MAMMAL and in the use of scientific terms, such as VIVIPAROUS.

The scientific terms appeared to represent a formalisation of ideas that the children produced in more everyday form. For example, the children would say FOUR LEGS and BREATHES instead of the expert version, FOUR LIMBED and RESPIRES. It is also likely that the experts produced some of these terms as defining features for the various generic classes of animals. The children had a grasp of taxonomy, but their knowledge of the use of the features for animal classification was sketchy. Throughout the research, it was apparent that age and expertise conferred fluency as they were associated with an increase in the unprompted number of features produced, but this was likely to be a metamemory phenomenon, indicating concentration and more efficient strategies for recalling information, rather than extreme variation in amount of knowledge held.

Carey (1985) examined the development of the ANIMAL concept in children aged between four and ten, concluding that six year olds are in transition, while ten year olds are behave much like adults. Carey's aim was descriptive. For example, in one study children aged four and ten and a group of adults were questioned to discover their beliefs about what things breathe, eat, sleep, get hurt, and have hearts. The objects were people, unfamiliar animals (aardvarks, hammerheads, stinkoos), some plants and inanimate objects. Four year olds under-attributed eating and breathing, that is, they did not know that all animals do these things. Attribution patterns were also explored using 'inductive projection': children were taught that people, or dogs or bees have a spleen and asked which other objects had spleens. Carey's main conclusion was that four year olds have a people-centred concept of animal, based on social and pyschological beliefs: when taught that people have a spleen, they attribute spleens to other animals, but when told that dogs or bees have a spleen, they are equally likely to attribute spleens to inanimate objects as to other animals. In contrast, ten year olds have an adult concept reflecting a biological understanding of the function and role of animal properties in ensuring survival. In projecting possession of a new property, ten year olds do not treat people more prototypically than dogs - that is, when told that dogs have omenta, their attribution pattern is similar to that obtained when they are taught that people have omenta.

The methods employed in the current research did not permit such detailed conclusions, but the emergence of a relatedness to man dimension, which decreased in importance with cognitive development, was similar to Carey's finding. Carey's transitional six year olds should be comparable to the seven year olds who made dissimilarity judgements in the current study and for whom relatedness to man seemed important. Also the similarity she reported between ten year olds and adults in inductive projection is paralleled by ten year olds and experts in feature generation here.

10.4.2 Differences between Adults and Children

Despite the commonalities in overall semantic structures and in the types of features generated, adults and children appeared to differ in their preference for abstract organisation and in their ability to handle abstract information. Children's preference for concrete organisation was indicated by the first SINDSCAL model in which the 'categorical' first dimension reflected Relatedness to Man/Habitat, followed by Size and Niceness dimensions in second and third place. There was a non-significant trend for age to be associated with the increasing importance of the third and intuitively abstract Niceness dimension. When experts were introduced to the model Dimension I was Size, Dimension II was Relatedness to Man/Habitat and Dimension III remained as Niceness; and in this model age and expertise were were associated negatively with Dimension II and positively wit Dimension III.

Examination of concreteness ratings for individual attributes of animals in relation to their fit with the dimensional structure indicated that concrete attributes like FURRY, FAT and LARGE tended to be represented by Dimension I and II, while more abstract attributes like NOISY, FRIENDLY and NASTY were not well represented until the emergence of Dimension III. A concrete-abstract progression has been widely noted and appears to be a general feature of cognitive development.

There were also indications that children are slower than adults in using abstract information. Reaction times to judge animals on concrete and abstract properties showed that abstract judgements took longer, but although this effect was greatest for children, the interaction between age and concreteness was not significant.

These results did not show that children were incapable of using abstract information. When unidimensional scale judgements from the seven year olds were fitted to the SINDSCAL solution, the overall pattern was similar to that obtained by fitting adults' data and to the result when the whole data set, including eleven year olds, was used. Indeed, for scales associated with Niceness, the seven year olds' data achieved a better fit than the adults', although the pattern remained the same. It seems that the organisation of children's semantic memories or their understanding of what is required by a similarity judgement or some combination of these factors favours discriminations based on concrete attributes; and that in speeded tasks concrete information is more readily available. Some version of abstract information exists, but is not preferred or accessed as quickly.

Another indicator of abstract organisation was the amount of taxonomic clustering in the semantic space, but again, adults and experts were not dramatically superior to children. When individual subjects' distances in the SINDSCAL solution were used to compute measures of clustering between and within taxonomic categories, seven year olds were found to have a shorter mammal-reptile distance than older children and adults, indicating that these groups of animals were least differentiated from one another for the youngest subjects.

Similar analyses of the SINDSCAL space when experts were included produced a more complicated result because experts with instructions to behave expertly showed an atypical ordering of the 'between cluster' distances. Their mammal-bird distance was shorter than their mammal-reptile distance, in contrast to all other groups who displayed the reverse order. It was thought that this might reflect expert awareness of the evolutionary sequence in which reptiles emerged first, followed by birds and then, independently, by mammals, while all other subjects may have been influenced by a ground-air distinction.

The fact that some animals have obvious thematic associations with animals from other taxonomic groups permitted a further exploration of the extent of taxonomic structure. The children informally expressed many themes, most concerning typical locations and functions of animals. It was feasible that the semantic spaces for younger children might display shorter distances between thematic associates like CAT and MOUSE than bewtween taxonomic associates like CAT and TIGER. However, comparison of taxonomic and thematic distances in such triads did not indicate developmental differences.

Children's weaker taxonomic clustering and their lack of preference for it was most evident in the ADDTREE representations of dissimilarity data, which showed a gradual emergence of taxonomic grouping with age and expertise. Birds were grouped at the second level of branches in trees for seven year olds and eleven year olds, but they emerged as a first level branch in the lay adult tree. Reptiles appeared as an exclusive group only in the expert dissimilarity trees. It must be said that no trees displayed perfect taxonomic groups. The conflict between similarity and taxonomy was reflected in the movements of ELEPHANT, which appeared either with land mammals or with fish and reptiles in all but the experts' feature-based similarity tree which was the only structure to group WHALE and ELEPHANT exclusively; however, these experts' features failed to group birds. For all other subjects WHALE remained with SHARK and in some cases was joined by ELEPHANT and reptiles.

10.5 Conclusions about Expertise and Conceptual Development

10.5.1 A Perspective for Progress in Cognitive Research

Expertise was not the original topic of this research. Experts were only included when it became apparent that the spatial model of semantic structure did not distinguish children from adults and the possibility that this might be a real effect rather than a result of the model's insensitivity had to be tested. However, the differences between experts and children proved more interesting than expected. They were initially interpreted in terms of the goals of experts, new theories about intensional and extensional aspects of meaning Schwarz (1979) and the feature model of semantic memory put forward by Smith et al (1974).

Speculations about the development of expertise then provided a framework for integrating progess in semantic memory during the 1970s with a new approach to concept acquisition and development which emerged in the 1980s, exemplified by the work of Markman (1989), Carey (1985) and Keil (1989).

The nature of expertise had not appeared to concern semantic memory theorists in the 1970s, but the assumption that concepts have expert definitions limited their approach and later underpinned the growing dilemma of how to represent both classical and probabilistic concepts. Meanwhile, with the exception of Keil (Keil and Batterman, 1984), the emerging developmentalists made minimal reference to semantic memory, but their claim that constraining theories are necessary to explain concept development points to the importance of understanding the role of expertise in conceptual structure. Theorising characterises experts, but the developmental studies show that it is not exclusive to those officially credited with expertise.

Finally, levels of expertise also parallel different types of similarity and help to explain context effects in various measures of similarity. Similarity defined as appearance matching cannot explain animal taxonomy, but similarity based also on relations between attributes provides a sensible account.

10.5.2 An Interpretation of the Study's Expertise Findings

Experts were included in this research to provide an extreme contrast with seven year olds in the evaluation of the scaling model as a way of representing conceptual development. It was initially surprising that experts' dissimilarity judgements were not dramatically different from children's, but once the feature generation data had been collected, it emerged that the main difference between experts and children might be in knowledge of formal animal taxonomy and the reasons for it. Where experts' dissimilarity judgements differed from children's, this was due to salient aspects of formal taxonomy overruling perceptual similarity as in the movement of ELEPHANT towards WHALE. The emphasis is on the 'formal' in taxonomy, because children possessed informal simplistic knowledge of animal hierarchies.

No attempt was made in this research to determine unequivocally that experts 'knew' taxonomy, while seven year olds did not, but it was a plausible assumption. The experts were medical doctors who had received formal training in biological taxonomy. In contrast pilot studies with children aged six to eight indicated that at this age there is a concept of 'animals', some of which are 'birds', 'fish' and 'insects'; that the prototypical 'animal' is a quadruped mammal; and that the word 'mammal' and the distinguishing features of the class are not explicitly known. However, a lack of formal taxonomy in children and doubtless many lay adults does not noticeably impede everyday communication between all groups, including experts. If anything, it is the expert knowledge that is unnecessary. A further observation was that children seemed less troubled by unclear taxonomy than experts, as if they know that unequivocal rules for classifying animals exist, but are comfortable either with not knowing them or with an approximate version. It is highly likely that lay adults are in a similar position.

The Feature Model of semantic memory described by Smith et al (1974) offered a suitable framework for representing this difference between everyday and expert knowledge because it deems some features to be defining and others to be characteristic. Smith et al suggested that the decision time to compare two words is the result of a 2-stage process in which words are first compared on characteristic features and then on defining features. If the subject and predicate in a false sentence are related, subjects in a speeded task tend to make errors owing to the large overlap in everyday terms between the features of the two objects; only by resorting to a stage two comparison can subjects decide if the apparent similarity is supported by a match on defining features as well. It is possible that stage one comparisons involve everyday knowledge, while the stage two operations make use of the expert knowledge.

Conceptual development could therefore be viewed as change in what constitutes defining features, with perceptual and possibly idiosyncratic defining features giving way to a commonly agreed set. Either features change status and become defining or new defining features are acquired. For animal terms the acquisition of new features would seem most likely because the basis of animal classification is often hidden from view or requires understanding of a process. In this way the acquisition of defining features could underly the concrete-abstract progression because defining features could be invisible or more abstract.

Developments in semantic theory indicated why this difference between novices and experts might be particularly large for animal concepts. Analyses of meaning often distinguish between the intension and extension of a noun. The intension is a list of defining properties and the extension is the class of objects to which the noun applies. However, Schwartz (1979) argued that there is a difference between 'nominal kind terms' and 'natural kind terms', suggesting that the intension-extension theory does not apply to both types of noun (p309-311).

While a term like BATCHELOR has a clear intension, terms like GOD or TIGER do not. Their intensions must be discovered empirically and will alter with the progress of scientific enquiry. For example, there are many of tests used by jewellers to establish that a substance is GOLD, but it is likely that the atomic number is the closest to an intension in the same sense as the intension of BATCHELOR is 'male and unmarried'. The point is that intensions of natural kind terms are generally not known, but this does not stop people using the terms effectively. A more practical proposition is that individuals have different 'working intensions' for natural kind terms, that these intensions can be modified with experience and that development involves the acquisition of adult intensions, followed possibly by expert intensions.

In addition, Schwartz's analysis and the experts' handling of the conflict between overall similarity and taxonomy in this research suggested that perhaps experts knowingly create 'working intensions' by converting some of their knowledge about natural kind terms into the explicit rule-based definitions that apply to nominal terms. Schwartz observed that natural kind terms refer to things which change state - grow, degenerate, display stages - but remain essentially the same kind of thing. Therefore a cognitive system dealing with natural kinds might be motivated to find the rules that pin down such variable entities. Children must also use rules of some kind to generate basic categories and informal superordinates, but these rules would be less complex and would not generally contradict perceptual similarity.

This view of expertise is supported by evidence that experts have nore detailed basic categories. Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) noted that an experienced airplane mechanic made more discriminations between types of planes than lay people and had a more differentiated set of basic categories. Coltheart and Walsh (1988) showed that expert bird-watchers use largely the same dimensions as lay people in discriminating amongst everyday birds in a scaling task; where experts differ is that they can scale more specified subsets of birds.

Similar ideas have been proposed elsewhere. Morton and Byrne (1975) reformulated some of Rips et al's (1973) reaction time findings to fit a distinction between 'semantic' and 'scientific' stores. In their view, information in the 'scientific store' is explicitly learned rather than naively observed and held as facts and rules which are used to verify propositions, especially those involving quantifyers. They proposed that the time to verify a scientific proposition should be a constant and, following some judicious tinkering with the Rips et al data, they were able to demonstrate some support for this. Miller and Johnson-Laird (1976) suggested that concepts consist of a definitional core surrounded by identification information, a view which parallels the characteristic-defining model, but implies a selective advantage of characteristic features in facilitating identification. Findings by Zaidel (1987) may reflect these ideas. She reported a hemispheric asymmetry in semantic judgements, with the right brain being more affected by semantic relatedness. The left brain, being the logical linguistic brain, is a likely repository for the information and mechanism of expertise, but if the left brain does not display semantic distance effects, theories of semantic memory must be revised to account for categorisation without relatedness effects.

To summarise, the defining characteristic of expertise does not appear to be in terms of quantity of knowledge, but in quality and organisation. Experts and lay people alike make empirical observations of the world, but experts are more likely to enquire closely, going beyond everyday information and uncovering details, such as atomic number, which are not available to everyday observers. Expert findings then enable the identification of new categories and classification rules, both of which can create revised perceptions of similarity.

10.5.3 Expertise, Semantic Memory and Concept Formation

Early research in semantic memory was dominated by an assumption of expert knowledge and strategy which may have constrained progress. A brief review of the history of semantic memory illustrates this. The Collins and Quillian (1969) model was based on the Teachable Language Comprehender (Quillian, 1968), a computer system which aimed to recognise text and answer questions. It contained heirarchical structures with explicit class-inclusion rules, for example, CANARY-ISA-BIRD-ISA-ANIMAL. Furthermore, with the rational assumption of cognitive economy, it allowed properties to be stored only where they had to be in order to ensure correct inferences: for example, WINGS were stored with BIRD and not with CANARY. The time it took to verify propositions was directly related to the number of steps separating subject from predicate in the hierarchy.

Almost immediately, studies showed the importance of factors other than the rational structure of the hierarchy in determining verification time. Conjoint frequency between instance and category (Wilkins, 1971), property frequency for categories (Conrad, 1972), typicality of instances (Rips et al, 1973; Rosch, 1973) and category size (Landauer and Meyer, 1972) all seemed better than the steps in the hierarchy at predicting verification times. Many of these variables boiled down to the notion of perceived semantic relatedness or distance (Rips et al, 1973) which was not represented in the hierarchical structure - for example, typicality ratings indicate that a robin is a bird in a greater sense than a chicken is a bird, but a hierarchical structure with ISA links, provides no justification for some ISA links being stronger than others. The effect of these findings was the abandonment of hierarchical structure and the revision of the model to include links of different strengths (Collins and Loftus, 1975). The aim of this summary of research history is to emphasise that hierarchies and all-or-none defintions of what is and is not a bird are outcomes of expert behaviour. Similarly, cognitive economy is an rational description of an expert's knowledge. Therefore it seems more sensible to view such hierarchies as potential products of a knowledge system, rather than as the fundamental structure of the system itself. Structure might be more accurately deduced by examining normative products, rather than the rarities generated by experts.

A similar assumption of expertise as generating rather than resulting from conceptual structure and activity can account for the history of research in concept formation. Early studies investigated what has since been termed the 'classical' view of conceptual structure (Medin and Smith, 1981). This holds that concepts are defined by necessary and sufficient features which all instances of the concept must possess. Such concepts are easily created using artificial combinations of colour and shape and they seem to describe certain states like 'bachelorhood', but philosophers , notably Wittgenstein, have shown that the necessary and sufficient features of some concepts are very had to identify. In cognitive psychology such concepts are described as having a prototypical or family resemblance structure which does not specify any features as necessary, but allows items to be instances if they have an adequate combination of features (Rosch and Mevis, 1975). The acceptance of this probabilistic view of is largely due to Rosch's descriptions of protoype effects. In a review of classical and probabilistic approaches Medin and Smith also looked at exemplar theory; this is not included here because it is not vital to understanding the role of expertise.

In discussing the new approach to understanding concepts Neisser (1987) referred to the 'Roschian revolution', while Lakoff (1987) described Rosch's papers as 'electrifying', reactions which perhaps indicate the strength of commitment to the earlier classical view. The devotion of academics to the classical view should not surprise because classical concepts are expert's concepts. That is, they imply a complete understanding of the components of a concept and they permit reasoning. Furthermore, even when expert's are fully aware that classical definitions do not fit observations, they allow exceptions to stand until the definition is improved.

Armstrong, Gleitman and Gleitman (1983) showed that apparently classical concepts like 'odd number' and 'even number' display prototype effects. Subjects were able to rate instances of odd and even numbers in terms of typicality, for example, '7' was thought odder than '23'. Armstrong et al distinguished rating category membership from typicality or exemplariness and found that subjects could, on one hand, agree that it made no sense to rate odd numbers for membership and, on the other, provide typicality ratings. It was interesting that substantial percentages of subjects also claimed that instances of categories like fruit (43%), sport (71%) and vehicle (74%), which are supposed to lack a classical definition, could not be rated for category membership either. This is reminiscent of the dilemma that subjects experience in making dissimilarity judgements for scaling studies when superordinates are included (Coltheart et al, 1985). Some subjects have no difficulty rating dissimilarity between pairs like ROBIN-BIRD and CHICKEN-BIRD in the context of instance-instance pairs. These subjects interpret the task as a typicality rating. Other subjects adopt the view that all instances are equally and maximally similar to the superordinate, presumably because the all-or-none class inclusion rule is prepotent.

Armstrong et al sensibly refused to accept that there can be anything 'fuzzy' about odd and even numbers and asked instead why graded responses keep showing up. Their interpretation was in terms of concepts consisting of cores plus identification procedures (Miller and Johnson-Laird, 1976), where the core allows reasoning and the identification element permits fast recognition.

Lakoff (1987) felt that the apparent contradiction of simultaneous classical and prototypical concepts is the result of wrongly interpreting effects as structure. It is wrong to say that graded ratings (the effects) only occur where categories are not all-or-none (the structure); instead, they can be a by-product of superimposed models of the world. There are many numerical concepts, such as the digits 0 t 9, multiples of 5 and powers of 2, which all operate at once and can make some numbers seem more privileged than others. Lakoff viewed category structures and proptotype effects as by-products of Idealised Cognitive Models. Assessment of Lakoff's ideas in detail would require a review of their psycholinguistic antecedents, which cannot be attempted here, but the central notion that concepts are models of the world and hence best understood in terms of their theoretical content has become important and is another indicator of the significance of expertise.

Murphy and Medin (1985) argued that classical and probabilistic theories of conceptual structure fail because they do not explain the selection either of features or of correlations between features. To do so requires an understanding of the context provided by the categorising system. Complex concepts such as 'ocean drive' illustrate the need for context, since they cannot be comprehended by combining their components, but require a host of other concepts making sense together. Context can be in terms of human interests, needs and goals, but Lakoff (1987) took it to a logical conclusion with the idea of embodiment: the selection of features and correlations is fundamentally tied to the perceptual system, which, for example, cannot summate red and green light wavelengths because these mutually inhibit neurones, but can combine blue and green wavelengths which stimulate different neurones.

Putting this into the context of expertise, an expert's concepts are likely to be based not just on matching features, but on an understanding of complex relationships between features which renders some features more important than others. Therefore the concept is not just the list of features but the principle that unifies them. Uncertainty about features, graded structure and unclear cases may imply either that the underlying theory is unclear or that it is a simple unstable theory based on superficial similarity. Therefore the child's or novices concepts may contain either poorer theories or less complex ones based on spatial and temporal contiguity. It may be no accident that classical definitions reflect expert concepts at their best and probabilistic structures describe everyday concepts if the two types of conceptual structure are derived from theories of varying quality.

Nelson (1984) showed that when learning is incidental, the resulting categories are probabilistic, This is consistent with poor theory because incidental learning presumably provides little opportunity for understanding the basis of a concept; the learner must resort to a simple organising principle. Keil (1987) speculated about the role of prototypes, suggesting that they might be the most pervasive way of acquiring and organising knowledge, particularly for the ignorant. Since everyday learning is often pervasive and incidental, it could feasibly create probabilistic concepts, resulting from behaviour that is commonly regarded as non-expert. However, the expert-novice dichotomy may be misleading because what appears to be non-expert behaviour may reflect the best 'hypothesis' available to the novice or child. Murphy and Medin (1985) noted that conceptual development could be a vital area for showing the role of theories.

10.5.4 The Role of Theories in Cognitive Development

The central problem of language acquisition is that on hearing a new word, the child must work out what aspect of the situation the word refers to. Since this could be almost anything, the problem is to work out what directs the child, namely, what constrains induction. This issue has been discussed by many researchers, for example, Carey (1986) and Markman (1989).

Studies of linguistic development in the 1970s demonstrated that young children display a strong preference for thematic over taxonomic associates when asked to chose an object that is the same as a standard. Children were assumed to be bound by perceptual information, but they were not incapable of perceiving taxonomic relations because they could justify a taxonomic associate if asked to do so (Smiley and Brown, 1979). However, a preference for themes would be unhelpful in learning words, as illustrated by underextensions which indicate that the child has mistakenly included a specific thematic context as part of a word's meaning.

As outlined in Chapter 2, Markman and Hutchinson (1984) identified one circumstance in which young children would choose a taxonomic associate of their own volition, overriding perceptual similarity. This is when children think they are learning a new word: if the standard picture in a matching to standard task is given a nonsense name like DAX, children as young as two and three years old will choose the taxonomic alternative when asked to find the picture which is 'the same'. If the standard is not named, children choose thematically.

There is also evidence that infants' words are less overextended that was previously thought. Hettenlocher and Smiley (1987) analysed the context of infant speech and found that many spatio-temporal overextensions indicate memory for an event and act as comments. There was no evidence of underextension.

Markman cites studies with Gelman identifying a second situation in which young children abandon their thematic preferences. Subjects were shown a picture of a fish and told that 'fish breathe underwater', followed by a picture of a dolphin accompanied by the information that 'dolphin come up to breathe'; they were then shown a picture of a shark, drawn to resemble a dolphin, and asked 'what does this fish do - breathe underwater or come up?' Children aged 4 deduced that the 'fish' must breathe underwater, despite its visual similarity to the 'dolphin'. Children did not make irrelevant inferences about properties like weight and speed, for example, they would generally not deduce that a legless lizard could run as fast as a four-legged lizard.

In both of these cases the child is following a rule. Markman and Hutchinson described it as a 'taxonomic constraint', but Carey (1985) has probably captured the overall principle in the idea of a theory. According to this view, the child hearing a new word has the theory that the new word must refer to the same 'kind' of thing rather than to any thematic associate. How the child identifies 'kind' is important and presupposes an ealier theory. In the second case the child's theory is that if two things have the same name they must be the same kind of thing with similar properties.

Carey (1985) examined the development of the 'animal' concept itself between the ages of four and ten. When asked if animals, flowers and inanimate objects eat, breath and have hearts etc., four year olds under-attributed these properties to animals and the pattern of attribution did not vary with type of animal. Some four year olds attributed animate properties to a mechanical monkey. In contrast, adults attributed properties sensibly and differentiated between animal classes, for example, identifying that all breathe, but only vertebrates have bones. Six and ten year olds displayed intermediate patterns approaching adult behaviour.

Carey then reasoned that if subjects were taught a novel property of people, the extent to which they project the property to other objects is a measure of similarity between people and objects and can act as a standard against which to compare other belief patterns. Young children were taught that people have spleens, while ten year olds and adults were taught that people have omenta. The projection pattern of four year olds matched their attibution patterns of properties like breathe and eat, implying that these properties had been projected from people. Ten year olds' and adults' projection patterns did not match their property attributions, for example, while adults projected omenta to some animals, they attributed eating to all animals. Carey concluded that four year olds were using 'comparison-to-exemplar' reasoning for both projection and attribution, the exemplar being people in all cases; older children and adults used this type of reasoning when they had no alternative, but their contrasting projection and attribution patterns showed that some judgements were based on biological knowledge.

The prototypicality of people for four year olds was evident in their willingness to project spleens from people to animals but not from dogs to animals. Adults' and ten year olds' projection patterns were similar, irrespective of whether they believed that people or dogs had omenta, indicating that people and dogs were equivalent as possessors of most animal properties. There was a significant interaction between age and the exemplar that subjects were taught on: four and six year olds projected more from people than from dogs, while adults and ten year olds projected slightly more from dogs than from people, indicating that the prototype effect of people with respect to other animals was replaced by the similarity of dogs to other animals. Elsewhere Carey showed that similarities based on projections from people to objects were different from similarities judged in a neutral context, for example, while adults did not attribute omenta to inanimate objects, they were able to judge such objects to be similar to people in varying degrees. (The use of 'similarity' to describe both judgements is somewhat confusing and is discussed in section 10.5.5.).

Interpreting her results in terms of theories, Carey argued that young children have well developed beliefs about social and psychological aspects of people, but relatively little notion of biology. As these aspects of people are the children's only basis for making inferences, they cannot make sensible judgements about biological matters. By the age of ten children have developed a theory of the role and relationships of biological properties. Carey (1986) believes that conceptual structure must be understood in terms of the theories held when concepts are formed and that this leads inexorably to infant studies in order to identify theories that the child is born with. As shown earlier, these views have gained credence in cognitive research, with Murphy and Medin's (1985) argument that category structures require theories to specify attributes and relationships and Lakoff's (1987) description of concepts as Idealised Cognitive Models.

Keil's (1989) work was of greatest relevance to this thesis, first, because it applied Smith et al's (1974) model of semantic memory based on characteristic and defining features to conceptual development, exploring the developmental shift indicated by the current research; and second, because it explored the implications of the status of objects, that is, whether they are natural kind, nominal or artifactual. Keil and Batterman (1984) showed that a characteristic-defining shift occurs between the ages of five and ten. They constructed alternative descriptions of concepts so that one text contained characteristic features and an incorrect definition, while the other provided atypical characteristics with a correct definition. The objects had to have lots of characteristic features plus completely unambiguous definitions. Nominal kind terms like ISLAND and UNCLE met this requirement, while natural kind terms like CAT clearly did not. For example, an island could be described as a place sticking out of the land with water on all sides but one and with palm trees, girls with flowers in their hair and a warm climate; or it could be a place with appartment buildings, snow, no trees and water on all sides. When asked if these could be islands, younger children assented for the characteristic but invalid descriptions, while older children displayed the reverse pattern.

Probing for justifications showed that the children were reasoning, albeit sometimes erroneously: for example, a five year old might deny that a defining text described an UNCLE because 'he's little and 2 years old', thereby implying that adulthood is a defining feature of the concept. The shift occurred at different ages within semantic domains - five year olds were more likely to have correct concepts of a moral terms than cooking terms. Futhermore, this is not, as suggested earlier, a concrete-abstract or perceptual-conceptual change. Keil argued that if it were so the visibility of features should predict the age at which the shift occurs. Where both characteristic and defining features are invisible, as in moral terms like LIE and ROBBER, the shift should be late, but in fact the opposite occurs, moral terms shifting earlier than cooking terms, which have visible features.

The problem for natural kind terms was, as always, their lack of necessary and sufficient defining features, but drawing heavily on philosophical arguments such as Schwarz (1979) on the absence of intension for natural kind terms, Keil developed ideas about what might act as a definition for natural kinds (Keil, 1989, Chapter 3). The crucial factor is Schwarz's speculation that natural kind terms exist to make sense of things which have many superficially different but related forms. These forms might be stages in an animal's life cycle like infancy and maturity or temperature-dependent states of chemical compounds like water and ice. Therefore the concept of a natural kind must contain an underlying unifying trait that allows the kind to be recognised in spite of its varied forms.

Keil explored the effect that hypothetical discoveries and transformations would have on children's identification of natural kinds. For example, if scientists discover that horses have cow insides, parents and babies, older children are likely to assert that the objects are cows that just happen to look like horses, while younger children continue to believe that the objects are horses. For the older child the genetic history of the animal is prepotent, while for the younger child characteristic features define identity.

Given an account of plausible cosmetic operations that alter a skunk to look like a racoon, again nine year olds will assert that it is still a skunk while five year olds believe that it has become a racoon. However, five year olds are not completely misled by characteristic features of natural kinds: if a porcupine is altered to look like a cactus, five year olds assert that it is still a porcupine, implying that no information has been provided to convince them that an animal has changed into a plant. This reflects the early significance of ontological categories, that is, those that define fundamental kinds of things like events, physical objects, animals and plants. They must result from very early theories and underly the theory that allows a new word to be efficiently learned. (Keil (1979) had originally put forward a theory about the development of ontological categories, which is not directly relevant here, but is outlined in Chapter 2).

Drawing again on philosophy for his rationale, Keil suggested that the definition of a natural kind is our understanding of the causal relations underlying its feature correlations (Keil, 1989, Chapters 3 and 13). This is much the same as Carey's (1985) emphasis on theories, but Keil was the more explicit in pointing out the significance of causal relations for models of semantic memory or conceptual structure. He argued that measures of typicality or similarity between objects might describe conceptual structures but cannot in themselves explain which features become defining. Only an understanding of the causal relations between features identifies which features are important and which are not. This implies that models must represent similarity relations and causality or that causality underpins similarity.

Keil described the basis of prototype models as 'atheoretical tabulations of correlations and feature frequencies', dubbing this 'original sim.' Children are thought to be bound to 'original sim.' only when their theories fail; and since studies of language acquisition and early concept formation have shown that very young children are guided by simple theories, it may be, Keil argued, that when we think 'original sim' is operating, we have merely failed to uncover the theory. However, incorporating Lakoff's (1987) argument for embodiment implies that original sim. is not neutral because initial perception of correlation and similarity in the world must surely reflect theories embodied in the physiology of the perceptual system.

10.5.5 The Status of Similarity

Similarity has historically been important in accounts of conceptual structure, but has been dismissed in favour of theories. Murphy and Medin (1985) argued that similarity and attribute matching are insufficient to explain why concepts are formed because they ignore the difficulty of defining and selecting attributes; only a constraining theory could overcome this problem. They acknowledged that it was possible to treat relations as attributes, but described this as a richer structure than an 'attribute list' which underestimates knowledge. In particular, Keil (1989) described similarity as depending on theory-neutral tabulations of salient features, which adults escape by understanding the causal relations between features - although it is difficult to see how features can be salient without some underlying rational that causes them to be so.

These authors all acknowledged the influence of the philosopher Quine (1977) who described a subjective 'animal sense of similarity' that becomes an objective similarity 'determined by scientific hypotheses and posits and constructs'; Quine felt that 'things are similar....to the degree that they are interchangeable parts of the cosmic machine revealed by science' (p.171) However, since adults' perception of similarity is based at least on an understanding of simple spatial and temporal causal relations and to varying degrees on science, it cannot be the atheoretical construct that Murphy and Medin and Keil dismiss. In addition, it is difficult to imagine an 'animal sense of similarity ' or 'attribute matching' or 'theory-neutral tabulation' that does not depend on perception of at least relations between spatial, temporal and motivational features, therefore attempts to dismiss similarity as a basis for concept formation may be due to a definition of similarity that is too narrow and untenable.

Types of similarity have been identified which illustrate what underlies a similarity judgement and what may determine a concept. Ortony (1979) distinguished literal from non-literal similarity. Going further still Gentner (1983) described a continuum separating appearance matching, based on attributes only, from analogy, based predominently on matching between relations. It is possible that the extent to which a similarity measure reflects appearances as opposed to relations depends on the context in which similarity is judged.

Ortony felt that Tversky's (1977) feature matching model, accounting for literal similarity, could not deal with the radical asymmetry of metaphors, but could be modified to reflect the essence of metaphor, which is that a highly salient feature of one object is used to highlight a matching but less salient feature of another object. Metaphors are asymmetrical because reversal alters their effect: 'sermons are like sleeping pills' is more metaphorical than 'sleeping pills are like sermons'. The difference between literal and non-literal similarity is illustrated in comparing a statement like 'billboards are like placards', which depends on literal similarity, with 'billboards are like warts', which emphasises the ugliness of billboards and is an example of non-literal similarity accounted for by the modified model.

Gentner (1893) suggested a continuum depending on varying combinations of attribute and relation matching. For example, 'sunflower' and 'sun' are appearance matches depending on attributes alone. For a strong sense of similarity, Gentner argued that both attributes and relations have to match and that this constitutes literal similarity. Next on the continuum was analogy in which few attributes but many relations match. Abstraction was reserved for the situation in which the objects possess few attributes so that matching can only be based on relations.

Tversky (1977) identified a way in which context determines similarity via the 'diagnosticity' of attributes: in the context of animals, for example, the attribute 'real' has no diagnostic value and would be of very low salience, but if mythical beasts are introduced, it could become a basis for discrimination. This suggests that context should determine the degree to which similarity is based on attributes or relations.

The assumption underlying similarity judgements for MDS is that subjects will use everything they know about the set of objects, and that the most salient characteristics, whether attributes or relations, will determine similarity. For a set of animal terms, children who know little beyond appearances might be restricted to attribute matching, while experts could make use of relational information, such as similar physiological systems. This kind of similarity is different from Carey's (1985) measure which was derived from projecting spleens (or omemnta depending on age) from people to animals and objects. Carey's task forced subjects to focus on biological relations, such as they understood them, because the spleen was described as a biological property of people. Therefore the task was not 'how similar are these objects in terms of what you know about them?', but 'in terms of biology, how similar are these objects to people and therefore, how likely are they to have a spleen?'

Finally, Keil (1989) suggested that Gentner's appearance matching and literal similarity might underly the characteristic-defining shift. Therefore perceived similarity can mean different things and is a product of theories; and the narrow notion of appearance matching similarity, according to which a whale is similar to a shark, would not predict the structure of a concept based on more complex relational theory.

10.5.6 Novice-Expert Studies

Research explicitly directed at differences between novices and experts did not relate to the current studies as closely as might be expected because it has focused more on problem solving in complex domains than on explicating conceptual structure and has employed tasks and materials that do not resemble those used in semantic studies. However, children have been shown capable of demonstrating expert knowledge in specific domains, such as dinosaurs (Chi and Koeske, 1983).

The importance of expert theories which permit mental simulations of problem solving was the focus of various projects brought together by Gentner and Stevens (1983) and Chi has presided over a series of studies on expertise in physics (Chi, Glaser and Rees, 1982). One relevant finding was that novice physicists group physics problems according to surface features, such as the objects involved, while experts use deep features reflecting physical laws (Chi, Feltovich and Glaser, 1981). When asked to combine problems successively into as many groups as seem reasonable, only experts produce multi-level hierarchies. Experts can also be inferior to novices on certain tasks: Adelson (1984) showed that experts had difficulty answering questions about the algorithm of a computer programme, but not about what the programme produced; novices were better at algorithmic details than the output specification. Adelson interpreted this in terms of experts' use of abstract representations lacking concrete procedural details, while novices' are restricted to the concrete level.

Murphy and Wright's (1984) study of clinical psychologists was comparable to semantic studies in that it obtained feature lists of three diagnostic categories. Numbers of features listed increased with expertise, but distinctiveness between categories decreased: novices' lists hardly overlapped in comparison with experts' who allowed features which did not distinguish between diagnostic categories to appear under more than one category. Clinical diagnosis is a notoriously contentious activity, therefore fuzzy categories could reflect imperfect theory, but the finding could also imply that novices and experts interpreted the task differently. Experts may have included non-distinguishing features because these partly explain the diagnosis, while novices interpreted the task as requiring distinctiveness in the way that people listing features of animals rarely include common attributes like 'breathes'. Alternativly, as Murphy and Wright suggested, novices might focus on distinctive prototypes because these are important in training. Keil (1989) described other evidence of fuzzy expert categories in medical diagnosis, speculating that these might be the manifestation of very complex causal relations or the use of prototypes to facilitate rapid thought.

Lack of clarity has characterised expert biological classification - the knowledge domain of this reseach - to a greater degree than is generally realised. Lakoff (1987) reviewed competing views of biological taxonomy: similarity of form or function and phylogenetic history offer alternative bases for classification. Lakoff also noted the difficulties involved in identifying species.

Looking at biological taxonomy from a philosophical point of view, Dupre (1981) made a 'constructive suggestion' to account for the relation between everyday and scientific concepts. He noted various instance in which ordinary and scientific concepts do not coincide. For example, the genus Opunta is divided into several species, but ordinary language identifies some of these as prickly pears and others as chollas. The point is that science recognises no basis for this discrimination. Similarly, science regards onions and garlic as lillies, while ordinary usage places them in a distinctive culinary category, far removed from flowers. Dupre felt that science attempts to avoid an antropocentric view, but ordinary usage reflects the functions of objects in relation to people.

The contribution of philosophers to an understanding of conceptual structure and development is greater than one might think from reading seminal papers in semantic memory. Wittgenstein is acknowledged to have influenced Rosch and no major account of development is now without reference to philosophical treatments of the nature of natural and nominal kinds in relation to conceptual structure (Carey, 1985; Markman, 1989; Keil, 1989).

10.5.7 Summary

Expert knowledge of animals appears to consist of everyday and abstruse facts, organised by classification rules, to produce taxonomies. When asked to make similarity judgements experts were primarily influenced by everyday knowledge, possibly because this, rather than their expertise, is more amenable to making graded judgements. There is no reason why knowledge obtained from expert enquiry should not affect similarity in the same way as knowledge available to everyone else, but its influence could sometimes be minimal in relation to superficial features of objects.

Subsequent developments in the field indicate that concepts can be thought of as theories about regularities and causal relations in the world. If this account is valid, it suggests that models of semantic memory should account for similarity or semantic relatedness by identifying specific relations between features. A comprehensive model would have to encompass the perceived relations and potential rules which both influence similarity and guide the acquisition of knowledge. Owing to the need to explicate theories, a model would also have to consider the physiology of perception because this determines the first relations that are perceived and the earliest theories.

10.6 The Future for Semantic Memory

Some final observations may be made about nature of research in semantic memory and its likely future direction. First, the view that memory consists of independent episodic and semantic systems has been modified. Second, the achievements of semantic memory as a paradigm have been challenged. Third, however, there is a new form of network modelling - neural networks - which has revived the initial promise of the field.

There is evidence that semantic information affects episodic tasks (McCloskey and Santee, 1981), therefore the systems are not functionally independent as first thought. Some authors now feel that the distinction has outgrown its utility and distracts from more important aspects of cognitive function (McKoon, Ratcliff and Dell, 1986). However, the episodic-semantic distinction was not a novel idea. Herrmann (1982) traced the history of memory typologies, finding that ideas mirroring the episodic-semantic distinction proposed by Tulving (1972) existed at the time of Aristotle and occur throughout history in various disciplines from philosophy to neurology. Much agreement on the nature of episodic memory was identified, but conceptions of semantic memory tended to refer either to the storage of facts (true semantics) or to skill (knowing how to do something). Given this historical ambivalence in notions about semantic memory, it is interesting that in recent history it has been investigated by psychologists and computer scientists as a factual store, but has now changed to a dynamic view, encompassing theories about the causes of co-occurence.

The early achievements of semantic memory research have been questioned. Chang (1986) noted that, over two decades, research activity in semantic memory had risen and declined in a dramatic manner and that some leading authors felt that it had taught us precisely nothing about the way people handle information in a long term store. Although this is an extreme claim, the confusion generated by detailed and often inconsistent findings was more apparent at the outset of the current research than any strong and unifying theory.

Connected with this initial sense of disappointment was the growing view that research in cognition is perhaps too experimental to progress with a sense of accomplishment and satsifaction. One effect of strict experimentation is to create highly specialised theories which cannot be compared, while another perhaps more serious trend is the tendency to develop elaborate theories about artificial or restricted aspects of behaviour. Claxton (1988) identified various criteria that cognitive research attempts to fulfil and demonstrated their disadvantages. For example, in meeting the experimental criterion, testable hypotheses are generated, but this may be too harsh a requirement because it rules out general investigations which can yield useful results. In the current context it can be seen that Rosch's work on the structure of categories and Carey's on age-related beliefs about animals were largely descriptive in origin and were both atypical styles in an experimental area. It seems that too strict an experimental approach is likely to generate trivial narrow models that explain a subset of artificial behaviour. Claxton suggested alternative criteria, including several that relate to development, growth and change, either in an evolutionary sense or within the individual. The final criterion was that theory be pragmatic and of everyday use.

Johnson-Laird (1984) criticised semantic memory networks and, by implication of their sameness, feature theories, for being too powerful to refute by evidence. More seriously, he argued that they display the 'symbolic fallacy' of translating one set of symbols into another without explaining anything because they do not connect the symbols to the real world. This is because they are alleged to represent only intensional relations which tie one word to another by propositions, rather than extensional relations which would attach words to the world. For example, they represent DOG ISA ANIMAL without specifying the extensions of either.

Furthermore, argued Johnson-Laird, if a model represented what dogs are and what animals are, it would not need to specify the intensional relation that 'dogs are animals'; this would be an automatic consequense of representing extensions. He proposed two levels of representation. First, there must be an intensional representation or proposition. This would be translated to an extensional representation or 'mental model', reflecting a realistic sample of events to which the expression might refer.

This criticism of networks is hard to fathom. Johnson-Laird acknowledged that a network of a different sort could be designed to overcome the problem. Also intensional relations in networks specify extensions and assume that words are lexical items that map onto non-linguistic representations at some point. For example, the attributes of DOG are its intension and these stand for the various mental states that correspond to the extension. It is possible that the criticism refers to the concentration of semantic theorists on words and their failure to identify fundamental attributes corresponding to events in the world, a lasting issue in cognition (Fodor, Garrett, Walker and Parkes, 1980).

Johnson-Laird provided many examples of the need for extensional representations in comprehension. One example is a sentence like ' the ham sandwich at Table 5 is hungry and getting impatient'. This kind of sentence has been considered by linguists to have a special status because literal meanings and context are of no help whatsoever in comprehending it. The only way to understand this sentence is to retrieve the referent, namely, the customer at Table 5. A developmental perspective facilitates the observation that this situation has much in comon with that of a child learning a new word: the child must identify the object - the customer - and apply the name - ham sandwich - and it would be understandable if a child in this situation learned something it should not. The child has an active conceptual system some of which is attached to lexical items and theoretically all of these co-occurences could be represented in a network model.

Mental models cannot be dismissed because it seems intuitively reasonable that some sample of referents or a specific referent is tied to every concept and, as Johnson-Laird showed, some aspects of spatial comprehension require a mental simulation. However, as Johnson-Laird acknowledged, they do not preclude networks as representations of conceptual structure.

Oden (1987) felt that network approaches would ultimately provide a good account, but that the measurement of a much wider range of factors would be necessary to quantify relations within network. Perhaps those working on the role of early theories should become involved in specifying mini-theories to be incorporated in networks. Some of the original network advocates now work on neural networks, otherwise known as connectionism or parallel distributed processing. Fodor and Pylyshyn (1988) examined the contribution of connectionism, concluding that it is a way of implementing the earlier propositional theories.

The central principle of connectionism is that rules do not need to be represented because the connections themselves embody rules, so it would appear that neural networks accomplish Johnson-Lairds notion of a model based on extensions that does not require intensional relations. However, Lachter and Beaver (1988) claimed that neural networks contain arbitrary devices and architectures which mimic rules and that, at best, they help in the understanding of complex associations. It is reasonable to suppose that some rules have to be wired in to the network if only to simulate human bodily constraints. Neural networks are attractive because they accomplish much and are plausibly analogous to what happens in real brains therefore their advocates should concentrate on early knowledge and embodiment to ensure that networks represent reality.

References

Adelson B (1984) When novices surpass experts: The difficulty of a task may increase with expertise. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 483-495

Anglin J M (1970) The growth of word meaning. Cambridge, Mass: MIT Press

Anglin J M (1977) Word, object, and conceptual development. New York: W W Norton & Co Inc

Anglin J M (1978) From reference to meaning. Child Development, 49, 969-976

Arabie P, Kosslyn S M & Nelson K E (1975) A multidimensional scaling study of visual memory of 5-year olds and adults. Journal of Experimental Child Psychology, 19, 327-345

Armstrong S L, Gleitman L R & Gleitman H (1983) What some concepts might not be. Cognition, 13, 263-308

Barsalou L W (1983) Ad hoc categories. Memory and Cognition, 11, 211-227

Battig W F & Montagu W E (1969) Category norms for verbal items in 56 categories: a replication and extension of the Connecticut Category Norms, Journal of Experimental Psychology Monograph, 80 (3, Pt2)

Berlin B (1978) Ethnobiological classification. In E Rosch & B B Lloyd (Eds) Cognition and categorisation. Lawrence Erlbaum Assoc.

Bisanz G Y, LaPorte R E, Vesonder G T & Voss J F (1978) On the representation of prose: New dimensions. Journal of Verbal Learning and Verbal Behaviour, 17, 337-357

Bowerman M F (1976) Semantic factors in the acquisition of rules for word use and sentence construction. In D M Morehead and A E Morehead (Eds) Normal and deficient child language. Baltimore: University Park Press

Carey S (1985) Conceptual change in childhood. MIT Press

Carey S (1986) Constraints on semantic development. In W Demopolous & A Marras (Eds) Language learnng and concept acquisition: Foundational issues. Ablex Publishing Corporation, New Jersey

Caramazza A, Hersh H & Torgerson W S (1976) Subjective structures and operations in semantic memory. Journal of Verbal Learning and Verbal Behaviour, 15, 103-117

Carroll J D (1972) Individual differences and multidimensional scaling. In R N Shepard, A J Romney & S B Nerlove (Eds) Multidimensional scaling: Theory and applications in the behavioural sciences. Vol 1. New York: Seminar Press

Carroll J D & Chang J J (1970) Analysis of individual differences in multidimensional scaling via an N-way generalisation of 'Eckart-Young' decomposition. Psychometrika, 35 (3) 283-319

Carroll J D & Wish M (1974) Models and methods for three-way multidimensional scaling. In D H Krantz, R C Atkinson, R D Luce, P Suppes (Eds) Contemporary developments in mathematical psychology. Vol 2. Measurement, psychophysics and neural information processing. San Francisco: W H Freeman and Co

Chang T M (1986) Semantic memory: facts and models. Psychological Bulletin, 80, 1-24

Chase W G & Simon H A (1973) The mind's eye in chess. In W G Chase (Ed) Visual information processing. New York: Academic Press

Chi M, Glaser R & Rees E (1982) Expertise in problem solving. In R Sternberg (Ed) Advances in the psychology of human intelligence. Vol 1. Hillsdale, N J: Lawrence Elrbaum

Chi M & Koeske R D (1983) Network representation of a child's dinosaur knowledge. Developmental Psychology, 19, 29-30

Claxton G (Ed) (1988) Growth points in cognition. Routledge

Collins A M & Quillian M R (1969) Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behaviour, 8, 240-247

Collins A M & Loftus E F (1975) A spreading-activation theory of semantic processing. Pyschological Review, 82 (6) 407-428

Coltheart M (1981) The MRC Psycholinguistic Database. Quarterly Journal of Experimental Psychology, 33A, 497-505

Coltheart V & Evans J St B T (1981) An investigation of semantic memory in individuals. Memory & Cognition, 9 (5) 524-532

Coltheart V, Hale D & Walsh P (1985) Semantic representation of taxonomic categories: multidimensional scaling and categorisation experiments. Annual Meeting of the Psychonomic Society, Boston

Coltheart V & Walsh P (1987) Expert Knowledge and Semantic Memory. In M M Gruneberg, P Morris & R N Sykes (Eds) Practical aspects of memory: Current research and issues. Chichester: John Wiley

Conrad C (1972) Cognitive economy in semantic memory. Journal of Experimental Psychology, 92, 149-154

Coxon A P M & Jones C L (1978) The images of occupational prestige. London: MacMillan

Coxon A P M & Jones C L (1979) Class and heirarchy. London: MacMillan

Coxon A P M & Jones C L (1979b) Multidimensional scaling: exploration to confirmation. Royal Statistical Society Spring Conference

Davison M L (1983) Multidimensional scaling. John Wiley & Sons, Inc.

Di Vesta F J (1966) A developmental study of the semantic structures of children. Journal of Verbal Learning and Verbal Behaviour, 5, 249-259

Di Vesta F J & Stauber K A (1971) Identification of verbal concepts by preschool children. Developmental Psychology, 5 (1) 81-85

Dupre J (1981) Natural kinds and biological taxa. The Philosophical Review, 90 (1), 66-90

Ervin S M & Foster G (1960) The development of meaning in children's descriptive terms. Journal of Abnormal and Social Psychology, 61, 271-275

Fiske S T, Kinder D R & Larter W M (1983) The novice and the expert: knowledge-based strategies in political cognition. Journal of Experimental Social Psychology, 19, 381-400

Ferguson G A (1965) Nonparametric trend analysis. McGill University Press

Fodor J A, Garrett M F, Walker E C T & Parkes C H (1980) Against definitions. Cognition, 8 (3), 1-105

Fodor J A & Pylyshyn Z W (1988) Connectionism and cognitive architecture: A critical analysis. Cognition, 28, 3-71

Gentner D (1983) Structure-mapping: a theoretical framework for analogy, Cognitive Science, 7, 155-170

Gentner D & Stevens A L (1983) Mental models. Hillsdale, NJ: Erlbaum

Gilhooley K J & Logie R H (1980) Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1,944 words. Behaviour Research Methods and Instrumentation, 12, 395-427

Glass A L, Holyoak K J & O'Dell C (1974) Production frequency and the verification of quantified statements. Journal of Verbal Learning and Verbal Behaviour, 13, 237-254

Hampton J A & Gardiner M M (1983) Measures of internal category structure: A correlational analysis of normative data. British Journal of Psychology, 74, 491-516

Hampton J A (1979) Polymorphous concepts in semantic memory. Journal of Verbal Learning and Verbal Behaviour, 18, 441-461

Hampton J A (1981) An investigation of the nature of abstract concepts. Memory & Cognition, 9 (2) 149-156

Hampton J A (1982) A demonstration of intransitivity in natural categories. Cognition, 12, 151-164

Henley N M (1969) A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behaviour, 8, 176-184

Herrmann D J (1982) The semantic-episodic distinction and the history of long-term memory typologies. Bulletin of the Psychonomic Society 20 (4) 207-210

Hollan J D (1975) Features and semantic memory: Set-theoretic or network model? Psychological Review, 82, 154-155

Homa D, Rhoads D & Chambliss D (1979) Evolution of conceptual structure. Journal of Experimental Psychology: Human Learning and Memory, 5 (1) 11-23

Howard D V & Howard J H (1977) A multidimensional scaling analysis of the development of animal names. Developmental Psychology, 13, 108-113

Howard J H (1977) More on the statistical treatment of individual differences in multidimensional scaling. Perception and Psychophysics, 22 (1) 106-107

Hutchinson J W & Lockhead G R (1977) Similarity as distance: A structural principle for semantic memory. Journal Experimental Psychology: Human Learning & Memory, 3 (6) 660-678

Huttenlocher J & Smiley P (1987) Early word meanings: the case of object names. Cognitive Psychology, 19, 63-89

Johnson-Laird P N (1983) Mental models. Cambridge University Press

Johnson-Laird P N, Hermann D J & Chaffin R (1984) Only connections: A critique of semantic networks. Psychological Bulletin, 96 (2) 292-315

Johnson-Laird P N (1987) The mental representation of the meaning of words. Cognition, 25, 189-211

Johnson S C (1967) Hierarchical clustering schemes. Pyschometrika, 2 (3), 241-254

Kendall D G (1971) Maps from marriages. An application of nonmetric multidimensional scaling to parish register data. In F R Hodson, D G Kendall & P Tatu (Eds) Mathematics in the archeological and historical sciences. Edinburgh: Edinburgh University

Keil F C (1979) Semantic and conceptual development. An ontological perspective. Harvard University Press

Keil F C (1987) Conceptual development and category structure. In U Neisser (Ed) Concepts and conceptual development: Ecological and intellectual factors in categorisation. Cambridge University Press

Keil F C (1989) Concept, kinds and cognitive development. MIT Press

Keil F C & Batterman N (1984) A characteristic to defining shift in the development of word meaning. Journal of Verbal Learning and Verbal Behaviour, 23, 231-236

King M C, Gruenewald P & Lockhead G R (1978) Classifying related stimuli. Journal of Experimental Psychology: Human Learning and Memory, 4 (5) 417-427

Kruskal J (1977) The relationship between multidimensional scaling and clustering. In van Ryzin (Ed) Classification and clustering. New York: Academic Press

Kruskal J B & Wish M (1978) Multidimensional scaling. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001. Beverley Hills and London: Sage Publications

Koslowski B (1980) Quantitative and qualitative changes in the development of seriation. Merrill-Palmer Quarterly, 26 (4), 391-405

Lachman J L & Lachman R (1979) Theories of memory organisation and human evolution. In C R Puff (Ed) Memory organisation and structure. Academic Press

Lachter J and Beaver T G (1988) The relation between linguistic structure and associative theories of language learning - a constructive critique of some connectionist learning models. Cognition , 28, 195-247

Lakoff G (1987) Women, fire and dangerous things: What categories reveal about the mind. University of Chicago Press

Landauer T K & Mayer D E (1972) Catergory size and semantic memory retrieval. Journal of Verbal Learning and Verbal Behaviour, 11, 539-549

Landis T Y & Herrmann D J (1980) Differences in encoding and categorising between children and adults. Child Development, 51, 896-898

LaPorte R E & Voss J F (1979) Prose representation: A multidimensional scaling approach. Multivariate Behavioural Research, 14, 39-56

Loftus E & Grober E (1973) Retrieval from semantic memory by young children. Developmental Pyschology, 8, 310

McCloskey M & Glucksberg S (1979) Decision processes in verifying category membership statements: Implications for models of semantic memory. Cognitive Psychology, 11, 1-37

McCloskey M (1980) The stimulus familiarity problem in semantic memory research. Journal of Verbal Learning and Verbal Behaviour, 19, 485-502

McCloskey M & Santee J (1981) Are semantic memory and episodic memory distinct systems? Journal of Experimental Psychology: Human Learning and Memory, 7 (1) 66-71

McKoon G, Ratcliff R & Dell G S (1986) A critical evaluation of the semantic-episodic distinction. Journal of Experimental Psychology: Learning, Memory and Cognition, 12, 295-306

Mansfield A F (1977) Semantic organisation in the young child; Evidence for the development of semantic feature systems. Journal of Experimental Child Psychology, 23, 57-77

Markman E M (1989) Categorisation and naming in chldren: Problems of induction. MIT Press

Markman E M & Hutchinson J E (1984) Children's sensitivity to constraints on word meaning: taxonomic versus thematic relations. Cognitive Psychology, 16, 1-27

Means B M & Rohwer W D Jr (1976) Attribute dominance in memory development. Developmental Psychology, 12 (5) 411-417

Melkman R, Tversky B & Baratz D (1981) Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering and retrieval. Journal of Experimental Child Psychology, 31, 470-486

Mervis C B & Rosch E (1981) Categorization of natural objects. Annual Review of Psychology, 32, 89-115

Meyer D E (1970) On the representation and retrieval of stored semantic information. Cognitive Psychology, 1, 242-300

Miller G A (1969) A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 6, 169-191

Miller G & Johnson-Laird P N (1976) Language and perception. Cambridge, Mass.: Harvard University Press

Morton J & Byrne R (1975) Organisation in the kitchen. Attention and Performance V, 517-529

Moyer R S & Bayer (1976) Mental comparisons and the symbolic distance effect. Cognitive Psychology, 8, 228-246

Moyer R S & Dumais S T (1978) Mental comparison. The Psychology of Learning & Motivation, 12, 117-154

Murphy G L & Medin D L (1985) The role of theories in conceptual coherence. Psychological Review, 92 (3), 298-316

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

Nelson D G K (1984) The effect of intention on what concepts are acquired. Journal of Verbal Learning and Verbal Behaviour, 23, 734-759

Nelson K (1973) Some evidence for the cognitive primacy of categorization and its functional basis. Merrill-Palmer Quarterly of Behaviour and Development, 19, 21-39

Nelson K (1974) Concept, word, and sentence: interrelations in acquisition and development. Psychological Review, 81, 267-285

Neisser U (1987) Concepts and conceptual development: Ecological and intellectual factors in categorisation. Cambridge University Press

Nygren T E & Jones L E (1977) Individual differences in perception and preferences for political candidates. Journal of Experimental Social Psychology, 13, 182-197

O'Hare D (1976) Individual differences in perceived similarity and preference for visual art: a multidimensional scaling analysis. Perception and Psychophysics, 20 (6), 445-452

Oden G C (1987) Concept, knowledge and thought. Annual Review of Psychology, 38, 203-227

Ornstein P A & Corsale K (1979) Organisational factors in children's memory. In C R Puff (Ed) Memory organisation and structure. Academic Press

Ortony A (1979) Beyond literal similarity. Psychological Review, 86 (3), 161-180

Osgood C E, Suci G & Tannebaum P H (1957) The measurement of meaning. Urbana: University of Illinois Press

Parker R K & Day M C (1971) The use of concrete, functional and abstract concepts in multiple classification. Developmental Psychology, 5, 312-319

Paivio A (1971) Imagery and Verbal Processes. New York: Holt, Rinehart & Winston

Paivio A & Marschark M (1980) Comparative judgements of animal intelligence and pleasantness. Memory & Cognition, 8 (1), 39-48

Pruzansky S, Tversky A & Carroll J D (1982) Spatial verses tree representations of proximity data. Psychometrika, 47 (1), 3-23

Pylyshyn Z W (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 80, 1-24

Quillian M R (1966) Semantic memory. In M Minsky (Ed) Semantic information processing. Cambridge, Mass: MIT Press

Quine W V (1977) Natural kinds. In S P Sshwartz (Ed) Naming, necessity and natural kinds. Cornell University Press

Ramsey D O (1978) Multiscale: Four Programs for multidimensional scaling by the method of maximum likelihood. National Educational Resources Inc, Chicago, Illinois

Richards M M (1977) Sorting out what's in a word from what's not: Evaluating Clark's semantic features acquisition theory. Interdisciplinary Linguisitics Conference, University of Louisville

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

Rosch E (1973) On the internal structure of perceptual and semantic categories. In T E Moore (Ed) Cognitive development and the acquisition of language. New York: Academic Press

Rosch E (1975) Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104 (3), 192-233

Rosch E, Mervis C B, Gray W D, Johnson D M & Boyes-Braem P (1976) Basic objects in natural categories. Cognitive Psychology, 8, 382-439

Ross R T (1939) Optimal orders in the method of paired comparisons. Journal of Experimental Psychology, 25, 417-421

Rumelhart D E & Abrahamson A A (1973) A model for analogical reasoning. Cognitive Psychology, 5, 1-28

Russell G W & Lambert W B (1980) Sherlock Holmes meets Othello: a MDS analysis of literary characters. British Journal of Educational Psychology, 50, 277-288

Saltz E & Dunin-Markiewicz A (1978) Development of natural language concepts. IV. The relationship between semantic organisation and learning. Journal of Experimental Child Psychology, 25, 447-458

Saltz E, Dunin-Markiewicz A & Rourke D (1975) The development of natural language concepts. II. Developmental changes in attribute structure. Child Development, 46, 913-921

Sattath S & Tversky A (1977) Additive similarity trees. Psychometrika, 42 (3), 319-345

Schaeffer B & Wallace R (1969) Semantic similarity and the comparison of word meanings. Journal of Experimental Psychology, 82 (2), 343-346

Schaeffer B & Wallace R (1970) The comparison of word meanings. Journal of Experimental Psychology, 86 (2), 144-152

Schiffman S S, Reynolds M L & Young F W (1981) Introduction to Multidimensional Scaling. Theory, Methods and Applications. Academic Press, Inc.

Schvaneveldt R, Ackerman B P & Semlear T (1977) The effect of semantic context on children's word recognition. Child Development, 48, 612-616

Schvaneveldt R, Durso F T & Mukherji B R (1982) Semantic distance effects in categorization tasks. Journal of Experimental Psychology: Learning, Memory and Cognition, 8 (1), 1-15

Schwartz S P (1979) Natural kind terms. Cognition, 7, 301-315

Shephard R N (1974) Representation of structure in similarity data: problems and prospects. Psychometrika, 39 (4), 373-420

Simard L M & Taylor D M (1977) Evolution de l'identite ethnique chez des enfants d'un village quebecois. Mimeograph, Universite de Montreal

Smiley S S & Brown A L (1979) Conceptual preference for thematic or taxonomic relations: A non-monotonic age trend from preschool to old age. Journal of Child Psychology, 28, 249-257

Smith E E & Medin D L (1981) Categories and concepts. Harvard University Press

Smith E E, Shoben E J & Rips L J (1974) Structure and process in semantic memory: A featural model for semantic dicisions. Psychological Review, 81 (3), 214-241

Snider J G & Osgood C E (1969) (Eds) Semantic differential technique: a sourcebook. Chicago: Aldine

Spence I & Graef J (1974) The determination of the underlying dimensionality of an empirically obtained matrix of proximity. Multivariate Behavioural Research, 9 (3), 331-341

Spence I & Domoney D W (1974) Single subject incomplete designs for nonmetric multidimensional scaling. Psychometrika, 39, 469-490

Spreen O & Schulz R W (1966) Parameters of abstraction, meaningfulness and pronounciability for 329 nouns. Journal of Verbal Learning and Verbal Behaviour, 5, 459-468

Stockdale J E, Wittman L S, Jones L E & Greaves D A (1979) A multidimensional analysis of subjective crowding. In H R Gurkaynak & W A LeCompte (Eds) Human consequences of crowding. New York: Plenum Press

Storm C (1978) Acquiring principles of semantic organisation. Journal of Experimental Child Psychology, 25, 208-223

Storm C (1980) The semantic structure of animal terms: a developmental study. International Journal of Behavioural Development, 3, 381-407

Tenny Y S (1975) The child's conception of organisation and recall. Journal of Experimental Child Pschology, 19, 100-114

Tomikawa S A & Dodd D H (1980) Early word meanings: Perceptually or functionally based? Child Development, 51, 1103-1109

Torgerson W S (1958) Theory and methods of scaling. New York: Wiley

Tulving E (1972) Episodic and semantic memory. In E Tulving & W Donaldson (Eds) Organisation and memory. New York: Academic Press

Tversky A (1977) Features of similarity. Psychological Review, 34 (4), 327-352

Weeks D G & Bentler P M (1979) A comparison of linear and monotone multidimensional scaling models. Psychological Bulletin, 86 (2), 349-354

Wish M, Deutsch M & Biener L (1972) Differences in perceived similarity of nations. In A J Romney, R N Shepard & S B Nerlove (Eds) Multidimensional scaling. Vol 2. New York: Seminar Press

Wilkins A J (1971) Conjoint frequency, category size, and categorization time. Journal of Verbal Learning and Verbal Behaviour, 10, 382-385

Young F W & Torgerson W S (1967) TORSCA, A FORTRAN IV program for Shepard-Kruskal Multidimensional Scaling Analysis. Behavioural Science, 12, 498

Zaidel D W (1987) Hemispheric asymmetry in long term semantic relationships. Cognitive Neuropsychology, 3 (4), 1-25

Appendix 1: The Dissimilarities Pre-Test

Putting Words in Groups.

Here are some jumbled up words. Write them out again putting them into groups of words that you think go together in some way. Use as many groups as you like and give a name to each group.

Write your groups here.

(a)

spider
table
horse
sheep
cat
chair
bed
dog
wardrobe
goat
snake

(b)

knife
pen
spoon
towel
soap
pencil
fork
water
paper

(c)

orange
apple
potato
pear
carrot
grape
cabbage
lettuce

---

2. Putting things in order.

Here are some jumbled up things.

ice cream

a summers day

a hot cup of tea

ice cube

rain water

HOT ______________________________ COLD

Find the hottest thing and put it right at the HOT end of the line above. Then put the coldest thing right at the COLD end of the line. Then put the other things in between to show how hot (or cold) you think they are. Mark the line with a cross and write the name of the thing next to this.

Do the same for this set of things

a laugh

a shout

ordinary talking

a whisper

a scream

NOISY ______________________________ QUIET

3. Look at this shape: 

Now look at these shapes: 

Which of the bottom shapes is most like the top one?
Put a circle around the one you choose.

Now do these questions in the same way. Which of the bottom things is most like the top thing?

(a) man

boy puppy

(b) apple

cabbage orange

(c) lorry

car bicycle

(d) 

4. Read all these words and think about them:

Cardigan
Coat
Jeans
Jumper
Shirt
Shoes
Socks
Trousers

Now think about the two words:

trousers jeans

How alike are trousers and jeans? Quite alike really.

What about?

shirt socks

These words mean quite different things. See if you can choose a number between 1 and 7 which shows how alike two things are.
Choose 1 if you think the two things are very alike in meaning.
Choose 7 if you think the two things are very different in meaning.
If you think the two things are in between very alike and very different in meaning, choose a number in between 1 and 7. Put a circle around the number you choose.

trousers	jeans	VERY ALIKE	1	2	3	4	5	6	7	VERY DIFFERENT
shirt	socks	VERY ALIKE	1	2	3	4	5	6	7	VERY DIFFERENT
cardigan	coat	VERY ALIKE	1	2	3	4	5	6	7	VERY DIFFERENT
jumper	trousers	VERY ALIKE	1	2	3	4	5	6	7	VERY DIFFERENT
shoes	socks	VERY ALIKE	1	2	3	4	5	6	7	VERY DIFFERENT
jumper	coat	VERY ALIKE	1	2	3	4	5	6	7	VERY DIFFERENT

Appendix 2: Instructions for dissimilarity and unidimensional scale rating

1. Instructions for the Dissimilarities Task
I want you to think about the animals you have just read. Think about ways in which they are alike and ways in which they are different from one another.

To show you what I mean, suppose I asked you to think about things you can spread on bread or biscuits. Tell me some of these things? (Pause for examples)

How alike do you think butter and margarine are?

Quite alike really, don't you agree? But what about jam and butter? Or jam and Marmite?

Suppose you had a ruler of 'alikeness' like this one (show ruler) where 1 equals VERY ALIKE and 7 equals VERY DIFFERENT. Could you give butter and margarine a number to show me how alike you think they are?

Now do the same for jam and Marmite. Can you see that this might be a very different number?

I'd like you to do this for pairs of the animals you've been thinking about. Look at each pair and circle the number on this 1 to 7 ruler (show booklet) that you think is best for that pair. If you think the two animals are very alike you would circle 1 : if you think the two animals are very different you would circle 7. If the two animals are not very alike and they are not very different, find the number inbetween that you think is best. Choose numbers that YOU think are best - there's no right or wrong answer. Don't take too long over each one and try to think about everything you know about the animals. Try these first nine pairs for practice and ask if you need help.

2. Instructions for the Unidimensional Scale Rating Task
This is like the booklet you've just filled in. The animals are the same, but this time I want you to do something different. Each page has an animal and a list of things about that animal. Is it a FAT animal or a THIN animal? Or is it somewhere inbetween? Is it LARGE or SMALL? Or again, is it inbetween? Use the 1 to 7 ruler again to show me what YOU think. For each thing

choose the number that best describes the animal. Sometimes you will want to choose a number near the ends of the ruler, but other times neither end will seem right to you and you will want to choose a number near the middle. Do whatever YOU think is right for the animal. There is no right or wrong answer - only what you think.

Appendix 3: Occurrences of features in children's and experts' descriptions of animal and taxonomic terms

	11 ANIMAL TERMS		5 TAXONOMIC TERMS	
	CHILDREN	EXPERTS	CHILDREN	EXPERTS
1. SIZE & SHAPE				
big/large	19	32	05	03
small	17	26		
long	06	11		01
fat	03	03		
thin	03	05		
heavy	03	01		
fish-shape	01		01	
not tall	01			
powerful		01		
short		01		
tube		02		
solid			01	
2. COLOURING				
white	07	10		
black/blue	05	10	01	
brown	05	13		
grey	04	05		
stripes	03			
ginger	01	04		
colours		01	02	
green			01	
spots			01	
3. SKIN TYPE				
fur	15	08		
feathers	04	03	02	03
rough	02			
slimey	01			04
slippery	01			
hairy		02	01	02
no hair		01		
hard		01		02
scaley		01	02	03

	11 ANIMAL TERMS		5 TAXONOMIC TERMS	
	CHILDREN	EXPERTS	CHILDREN	EXPERTS
4. MOVEMENT				
runs	07			
slides	07	02		
swings	06	01		
fast	05	08		
climbs	04	01	01	
flies	04	01	07	09
jumps	04			
can't walk	01			
agile		01		
graceful		01		
move			04	
mobile				04
swim			04	05
walk				02
5. HABITAT				
sea	10	16		
water	06	04	08	08
jungle	05	02		
trees	04	03	01	
zoo	03			
houses/domestic	03	10		
Africa/Asia	02	03		
beaches	02			
fences	02			
land	02	04		02
ground	01			
not houses	01			
street	01			
wild places	01			
amphibious		01		01
cliffs		01		01
grass		01		01
nest		01	02	
cage			02	
tropical			02	
Trafalgar Square			01	
Earth				01

	11 ANIMAL TERMS		5 TAXONOMIC TERMS	
	CHILDREN	EXPERTS	CHILDREN	EXPERTS
6. ANATOMY				
tail	25	19	02	03
four legs	11	18	04	02
whiskers	08	04		
mouth	07		03	
beak	06	12	02	04
ears	06	07		
nose	06	03		
trunk	06	08		
eyes	05	01	02	02
jaw	05	06		
teeth/none	05	07		02
fins	04	03	01	03
wings	03	04	02	06
claws	02	01		
face	02		01	
feet	02	03		
tusks	02	03		
spines	01	01		
tongue	01	01		03
two legs		03	01	04
cartilaginous		01		
limbs		07		02
head		02		
opp. thumbs		02		
webbed		01		
no legs			02	
gills			01	05
brain				01
fingers				01
hollow bones				01
lungs/none				04
vertebrate				02
7. PHYSIOLOGY				
breathes	01	01	02	06
cold blooded		03	02	10
warm blooded			02	05
excretes				02
feeds				02
non-photosynthesise				02
grows				01
energy				01
moults				01

	11 ANIMAL TERMS CHILDREN	11 ANIMAL TERMS EXPERTS	5 TAXONOMIC TERMS CHILDREN	5 TAXONOMIC TERMS EXPERTS
8. COMPLEX PHYSIOLOGY OR BEHAVIOR				
spouts	05	01		
chases mice	03			
scared of dogs	02			
infests	02			
out in bad weather	01			
eats thro' trunk	01			
friend to dogs	01			
intelligent	01	07		
scared	01		02	
scavenger	01	02		
sees in dark	01	01		
sheds skin	01			
stamps	01			
alert		01		
is chased		02		
cheeky		01		
hyperactive		01		
independant		01		
flocks		01		
alone			02	
not human			02	
no clothes			01	
can't talk			01	
not fish			01	
conscious				01
stupid				01
hibernate				01
non maternal				01
9. NOISE				
noisey	08		03	
miows	03	03		
squeaks	02	01		
rattles	01			
characteristic cry		05		
chirps		02		
silent		01		
sing/whistle		01	02	

	11 ANIMAL TERMS CHILDREN	11 ANIMAL TERMS EXPERTS	5 TAXONOMIC TERMS CHILDREN	5 TAXONOMIC TERMS EXPERTS
10. DIET				
banannas	04			
eats alot	03			
cheese	02	03		
worms	01	01		01
meat		16		
vegetarian		03		
bones			01	
11. FEROCITY				
dangerous	07	05		02
aggressive/ viscious/ferocious	06	06		
poisonous	06	01		
eats people	05			
hunts	03	03		
bites	03			
frightening	02	01	01	01
harmless	01	03		01
unpleasant	01	01		
growls	01	02		
wild		02	01	01
friendly (most)			01	
hurts			01	
nice (some)			01	
scratches			01	
12. REPRODUCTION				
viviparous		02		03
oviparous		01		03
eggs		01	02	06
live young		01		02
mammaries				03
sexual reproduction				04
13. USES				
funny tricks	04			
pet	01	02	02	
ride	01			
knocks down trees	01			
leather		01		
cats eat			01	
people eat			01	01

	11 ANIMAL TERMS		5 TAXONOMIC TERMS	
	CHILDREN	EXPERTS	CHILDREN	EXPERTS
14. SUPERORDINATES				
mammal	07	27		
bird	06	16		
fish	03	05		
animal	02	16	01	16
not animal	01			
ape family	01			
cat	01	04		
reptile	01	10		
creature		02	01	03
collective term				03

THE BRITISH LIBRARY
BRITISH THESIS SERVICE

TITLE .. The Development of Semantic Memory A Spatial Model of Animal Concepts in Schoolchildren, Novices and Experts

AUTHOR .... Denise Ann Hale ......

DEGREE ....

AWARDING BODY City of London Polytechnic
DATE .... (1991) C.N.A.A

THESIS
NUMBER ....

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

The quality of this reproduction is dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction.

Some pages may have indistinct print, especially if the original papers were poorly produced or if the awarding body sent an inferior copy.

If pages are missing, please contact the awarding body which granted the degree.

Previously copyrighted materials (journal articles, published texts, etc.) are not filmed.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no information derived from it may be published without the author's prior written consent.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

cms 1 2 3 4 5 6

REDUCTION X 20
CAMERA 5
No. of pages

DX
172349