The Effect of Stimulus Frequency on Classification Accuracy and Response Time

by

David John Peebles

A thesis submitted to the School of Psychology
of the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology
Faculty of Science
University of Birmingham
Edgbaston
Birmingham
B15 2TT
England

September 1997

Abstract

Recent research in categorisation has focused upon the investigation of response time (RT) and the development of existing formal models to account for data from speeded classification experiments. The research reported here examines the alternative theories of categorisation RT and compares two current models, the *Exemplar Based Random Walk* model (*EBRW*; Nosofsky and Palmeri, 1997) and the *Extended Generalised Context Model* (*EGCM*; Lamberts, 1995, in press, submitted) in a series of experiments. In addition, a new model of categorisation RT is proposed, *ALCOVE(RT)*, based on Kruschke's (1992) connectionist *ALCOVE* model of category learning, and tested. This thesis also attempts to resolve a current debate in the categorisation literature by investigating the effect of stimulus frequency on categorisation performance. The results of the research provide strong evidence for the effect of stimulus frequency on both classification accuracy and RT and show that accurate predictions of these performance measures cannot be based solely on information relating to the abstract category structure.

Dedication

This thesis is dedicated to my wife, Helen and to my daughter, Fiona.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor, Koen Lamberts, for his invaluable guidance and advice throughout the course of this research. I am also indebted to the members of the Categorisation Group: Noellie Brockdorff, Steven Chong and Richard Freeman, for their help and comments. Thanks are also due to Andrew Olson and Mike Harris for their advice and practical assistance and to members of my family for their constant support. Finally, I thank my wife, Helen, for her love, encouragement and understanding, and my daughter, Fiona, for being such a source of happiness and for keeping my feet firmly on the ground.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Categorisation

The ability to divide the various elements of our environment into classes is a fundamental skill upon which most other cognitive processes rely. Categorisation has been called "the gateway between perception and cognition" (Barsalou, 1992) because it is the first cognitive operation performed once stimulus information has been acquired by the perceptual system. The primary functional role of categorisation is to facilitate inference; once an object is recognised as an instance of a particular category, further information about the object which has been previously associated with members of the category can be accessed and utilised. Knowledgeable and experienced gardeners, for example, upon classifying a plant as a member of the category *Common Chickweed*, may be able to recall, amongst other things, that it is also known as *Stellaria Media*, that it is a member of the category *Common Annual Weed*, that it smothers nearby plants and that it should be controlled before or during its seeding stage in Autumn, preferably with Gluphosinate Ammonium.

In the cognitive sciences, categories are regarded as mental representations of classes of objects in the environment grouped by structural, functional or theoretical similarities. It has been suggested that some categories, namely facial expressions, may be specified innately (e.g. Eckman, 1982; Eckman & Oster, 1979) but most, if not all, are generally believed to be constructed from experience of individual category members during an individual's lifetime. The precise structure of category representations is disputed and three principal alternative theories have been advanced. A great deal of progress has been made over the last thirty years, however, in the assessment of various hypotheses and the development of sophisticated formal models of categorisation behaviour.

In this introductory chapter, I provide a broad context for this research by outlining the essential elements of the three main theories, noting their advantages and limitations and illustrating how one, known as *Exemplar Theory*, has been most successful in explaining categorisation behaviour observed in the laboratory. Exemplar theory is the foundation of the research reported here. Therefore, I narrow the context further by describing two current exemplar-based models of categorisation—Nosofsky's (1986) *Generalised Context Model* (GCM) and Kruschke's (1992) *ALCOVE* together with an alternative proposal based upon a multidimensional generalisation of signal detection theory known as *Decision Bound Theory* (Ashby & Gott, 1988; Ashby & Lee, 1991, Ashby & Maddox, 1993).

Following these preliminary discussions, the main subject of the research is introduced. This concerns recent developments in which existing categorisation models have been extended to permit the prediction of classification response times (RTs). Three models are discussed, the *RT-distance* model (Ashby & Maddox, 1991, 1992; Ashby, Boynton and Lee, 1994) based on decision bound theory, the *Exemplar-based Random Walk* (EBRW) model (Nosofsky & Palmeri, 1997) based, in part, on the GCM, and Lamberts' *Extended Generalised Context Model* (EGCM) (Lamberts 1995, in press, submitted; Lamberts & Brockdorff, 1997) which, as its name suggests, is also founded on the GCM. An alternative theory of categorisation RTs is then proposed which employs the representational and processing assumptions of the ALCOVE model and augments them with a mechanism used in a number of connectionist models which allows the gradual accumulation of information over time.

Following these initial discussions, the predictions of the three models are then compared in four experiments which attempt in different ways to determine the precise nature of the information used in categorisation. Each of the models embodies different assumptions concerning the type and amount of information required for categorisation. The first two experiments examine the ability of the models to account for combined classification accuracy and RT data in order to determine the predictive power of the models' alternative assumptions. The third and fourth experiments continue this examination and extend it by investigating the effect of stimulus frequency on categorisation performance. These experiments are designed to investigate the nature of this effect and to discover if the predictions of the alternative models reflect the experimental results, in the expectation that inferences about the information used in categorisation may be made from differences in the models' ability to account for the data.

I will now proceed by setting the context for the research in the remaining sections of this chapter and will then conclude by giving a brief overview of the thesis.

Theories of categorisation

The earliest theory of categorisation—and consequently generally known as the *Classical Theory*—has a long history in philosophy and psychology and is characterised as the position that categories are describable wholly in terms of definitional rules specifying singly necessary and jointly sufficient characteristics. The rules defining a category specify the properties which an object must have in order to be a member of that category. For instance, according to the classical model, a person is a bachelor if and only if that person satisfies the criteria human, adult, unmarried and male.

Although this type of analysis can be successfully applied to certain categories, particularly those relating to kinship, legal, mathematical and other technical domains, the classical model is now widely regarded as being inadequate to account for most natural kind and artefact categories. Wittgenstein (1953) argued that the task of specifying a comprehensive set of defining properties is impossible for the majority of categories because, in most cases, category members can usually be found to which at least one of the properties does not pertain. As an example, Wittgenstein suggested the category game, arguing that this category is not amenable to a classical analysis because there are no common features which apply to all and only games:

> Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "there must be something common, or they would not be called 'games' "—but look and see whether there is anything common to all—For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that...
>
> And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail. (*Wittgenstein, 1953, I, 66*)

A related limitation of the classical model concerns the inherent rigidity of sets of definitional rules and the resultant all-or-nothing nature of the category membership criteria which they create. According to the classical account, there are no degrees of category membership—an object is either a member of a particular category or it is not. This constraint, however, does not correspond with the results of numerous empirical studies which demonstrate that people do judge some category members as being more typical of their category than others, (e.g. Barsalou, 1985; Mervis, 1980; Rosch, 1978; Rosch & Mervis, 1975; Smith & Medin, 1981).

Further shortcomings of the definitional rule theory are revealed when it is applied to some ambiguous real world cases. For example, although priests and gay men satisfy all the conditions for membership of the category bachelor outlined above, it is unlikely that many people would be inclined to classify them as such, suggesting that people are more flexible in their categorisation than the classical model allows. It should be noted, however, that a rule-based model can always be modified to accommodate irregular instances by the addition of further rules, (in the example above, these could include "is not a member of the clergy" or "is not gay"), but this only serves to highlight the potential unfalsifiability and generally unprincipled nature of the approach.

Wittgenstein's alternative proposal, that categories are structured in terms of the family resemblance of their members, has been widely accepted as providing a more powerful and flexible tool for the study of categorisation. The central notion behind this concept of family resemblance, that of similarity, has been adopted and given a more formal basis by the two other theories discussed below. The increased power and flexibility of a similarity-based approach derives from the inherent variability in similarity relationships between objects which underlies the continuously graded structure of categories uncovered by the typicality studies mentioned above. It is the employment of a similarity-based category membership criterion which allows the two alternative theories to account for a much wider range of phenomena than that captured by the classical theory.

The approach to quantifying similarity relationships adopted by the two models is based upon the assumption of a geometric representation of stimuli and categories in a psychological space. The dimensions of this space correspond to the attributes of the stimuli and similarity is taken to be a decreasing function of distance in the space (see Shepard, 1958, 1987; Nosofsky, 1992). A detailed formal analysis of this approach to similarity will be presented in Chapter 2.

The first of the alternative approaches to categorisation is known as *Prototype Theory* because it assumes that categories are represented as single prototypes abstracted from category exemplars. In prototype models, categorisation of an object is a function of its similarity to the alternative prototypes, the object being classified as a member of the category to which it is most similar. In addition, typicality effects are explained in prototype models by relating the typicality of a category instance to the similarity of the instance to the category prototype.

A great deal of experimental evidence for prototype effects on categorisation has been compiled over the years. Early studies (e.g. Posner & Keele, 1968; Rips, Shoben & Smith, 1973; Rosch, 1973, 1978; Rosch, Simpson & Miller, 1976) demonstrated that similarity to prototype is highly correlated with both categorisation accuracy and speed of response. Further evidence for prototype representations was provided by an experiment carried out by Homa, Sterling and Trepel (1981) in which people were trained to categorise pictures of abstract patterns into a number of different categories. When the subjects were subsequently asked to categorise the pictures again together with pictures of the prototype patterns for each category, it was found that they were more accurate in categorising the previously unseen prototypes than the original pictures upon which they had been trained. Homa *et al.* also found that, when the subjects were asked to categorise the full set of pictures after a period of a week, categorisation accuracy for the training stimuli had decreased but accuracy for the prototypical stimuli had actually increased.

Despite these findings, a number of difficulties associated with prototype theory have been identified which are seen by many as casting doubt on the value of the approach. The most important problem concerns the loss of information about individual instances resulting from the abstraction process. In prototype representations consisting solely of the central tendency of the category distribution, information concerning, amongst other things, possible correlations between various properties and the presentation frequency of individual exemplars is lost. However, there is experimental evidence that people do take correlational (Medin, Altom, Edelson & Freko, 1982; Ashby & Gott, 1988) and frequency (Nosofsky, 1988, Nosofsky & Palmeri, 1997) information into consideration when making judgements of category membership.

The shortcomings of prototype theory are further revealed by a number of experiments which have demonstrated that information about individual category exemplars influences classification. The argument proceeds as follows. If, as prototype theory assumes, categorisation is determined solely by similarity to a prototype representation, category responses to previously experienced exemplars should be no more accurate or faster than those to novel instances. If, however, information about individuals is retained during training, responses to old and new items will differ. Evidence supporting the latter hypothesis has come from a categorisation experiment carried out by Jacoby and Brooks (1984) which demonstrated that subjects, after a period of training, in the subsequent testing stage involving the previously seen exemplars, the category prototypes and novel exemplars, were faster at categorising the old exemplars than the others. Further evidence supporting exemplar effects comes from Malt (1989) who used a priming paradigm in a number of experiments to determine the type of information retrieved during the categorisation process. One aim of Malt's study was to discover whether items upon which subjects had been previously trained would be responded to more rapidly if a similar but novel item had been presented immediately before. The rationale behind the paradigm is that this should prove to be the case if exemplar information is present because the information remains accessible for a short time after being retrieved during the categorisation of the similar item. The result of Malt's experiments showed that this is indeed what happens.

The evidence for the storage and retrieval of specific exemplar information revealed by these and other experiments has led to the wide acceptance of an approach which assumes that representations of individual instances are all that is stored in memory. In contrast to prototype models, this approach, commonly known as *Exemplar Theory*, assumes that the typicality of an object is a function of the similarity of the object to each of the stored exemplars belonging to the category. Exemplar theory has been given a precise formal treatment in the Context Model of Medin and Schaffer (1978) and the Generalised Context Model (GCM) of Nosofsky (1984, 1986). As the latter model forms the basis for much of the discussion to follow, a detailed description of the principles and mechanisms it employs will be given in Chapter 2.

The GCM is designed to account for categorisation accuracy after training has taken place, (i.e. when the categories are assumed to have been learned to some specified extent). One exemplar model which embodies many of the assumptions of the GCM, Kruschke's (1992) ALCOVE, has been developed to account for the actual process of category learning. ALCOVE (an acronym of "Attention Learning COVEring map") is a connectionist model that incorporates exemplar-based stimulus representations and an error-driven learning mechanism to determine association strengths between exemplars and categories and attention strengths on stimulus dimensions. As ALCOVE also plays an important role in this research, the underlying principles and mechanisms of the model will be explained in detail in Chapter 2.

One of the successes of exemplar models has been their ability to account for the prototype effects described above. This ability is a consequence of the association of the typicality of a category instance with the summed similarity of that instance to all the exemplars of that category. According to this account, prototypical stimuli and stimuli very similar to the prototype, being close to the central tendency of the category distribution, will have a high typicality value and will therefore be responded to more accurately and rapidly.

Current exemplar theory is extremely successful in accounting for a wide range of categorisation phenomena and the formal models which have been developed have been able to provide precise quantitative fits to classification performance data in a variety of experimental conditions. Over the last few years, extensions of previously existing categorisation models have been proposed in order to explain flexibility in classification (Lamberts, 1994; Lamberts & Chong, submitted) and classification under time pressure (Lamberts, 1995). The most recent models, however, have been developed to investigate the time course of categorisation, (e.g. Ashby, Boynton & Lee, 1994; Lamberts, in press, submitted; Nosofsky & Palmeri, 1997).

The first of these recent models is the RT-distance (RT-D) model proposed by Ashby and his colleagues, (Ashby and Maddox, 1991, 1992; Ashby, Boynton and Lee, 1994). This model is derived from the Decision Bound model of categorisation, (Ashby, 1992; Ashby & Gott, 1988; Ashby & Lee, 1991, 1992) which assumes that each experience of a stimulus is represented as a point in a multidimensional perceptual space. According to decision bound theory, category learning involves the partitioning of the space and the association of a category label with each region. Subsequent categorisation of an object consists in the determination of the region into which the object falls. The line separating two categories is known as the decision bound and the RT-D model states that the categorisation response time for a particular object is an inverse function of the object's distance from the decision bound. The intuition behind this relationship is that exemplars that are similar to members of more than one category, (and therefore, in terms of the decision bound model are close to the decision bound), being ambiguous in relation to their category membership, will be categorised more slowly and less accurately than those similar to members of one category only. Ashby, Boynton and Lee (1994) have carried out a number of experiments which support this model.

An alternative proposal has recently been put forward by Nosofsky and Palmeri (1997). Their approach combines the GCM with a previously existing model of skill development, Logan's (1988) instance-based model of automaticity, and a random walk process to create an explicit mechanism for the classification decision process. This Exemplar-based Random Walk (EBRW) model incorporates a stochastic exemplar retrieval mechanism which allows the gradual accumulation of evidence for and against the alternative categories. This evidence drives the random walk process which implements the function relating classification response time to stimulus similarity.

A third recently proposed model of categorisation RTs, the Extended Generalised Context Model (EGCM; Lamberts, submitted), also incorporates a number of assumptions from the GCM. The EGCM differs from the GCM, however, in that it assumes that perceptual representations of stimuli are constructed by a time-dependent information accumulation process involving the sampling of stimulus features and that the similarity between this representation and stored exemplar representations in memory is constantly computed throughout this process. According to the EGCM, this information accumulation process terminates when a sufficient amount of evidence has been obtained for a category decision to be made and that categorisation RT is a function of the duration between stimulus onset and the completion of information accumulation.

The RT-D, EBRW and EGCM models are all relatively recent developments in the area of categorisation research, as is the study of categorisation response times itself. Consequently, the models have been subjected to a limited number of empirical tests and there have been few comparative analyses carried out. Although it shares a number of features with the EGCM and EBRW, the alternative model introduced in this thesis, *ALCOVE(RT)*, also embodies several distinct assumptions about the mechanisms involved in the categorisation process. One of the main aims of this research, therefore, is to test the ability of the EBRW, EGCM and ALCOVE(RT) models (for various reasons discussed below, the RT-D model is not included in this study) to jointly predict response accuracy and RT data from speeded classification experiments. A second goal is to examine the effects of stimulus frequency on categorisation performance and, in so doing, to discover whether the different underlying assumptions of the models are also sufficient to account for the data from experiments in which stimulus frequency is manipulated.

Overview of the thesis

As a broad background to the forthcoming discussion, the three central theories of categorisation were introduced in this chapter—Classical theory, Prototype theory and Exemplar theory, and it was described how, because of various shortcomings of the classical and prototype theories, most investigators have adopted exemplar-based models of category representation.

In Chapter 2, the concept of similarity is introduced because this notion plays a central role in all theories of categorisation discussed in this thesis. The various attempts to frame a more precise definition of similarity for the purposes of formal modelling are also described in order to facilitate subsequent discussion of formal models of categorisation. Three such categorisation models are then introduced and described in detail—the Generalised Context Model (GCM) of Nosofsky (1986), the Decision Bound theory proposed by Ashby and his colleagues (Ashby, 1992; Ashby & Gott, 1988; Ashby & Lee 1991, 1992), and Kruschke's (1992) neural network based ALCOVE model. The representational assumptions and categorisation mechanisms of the three models are compared and contrasted.

In Chapter 3, after a brief review of the historical progress of the psychological study of response times (RT) and an outline of some of the basic assumptions held by RT investigators, recent developments in categorisation research in which formal models have been applied to the analysis of categorisation response times are introduced. The mechanism for generating RT predictions from the Decision Bound model is discussed and two recent exemplar models of categorisation response times based on the GCM—the Exemplar-Based Random Walk (EBRW) model (Nosofsky & Palmeri, 1997) and the Extended Generalised Context Model (Lamberts, submitted) are described in detail.

In Chapter 4, after an initial discussion of connectionist approaches to the modelling of response times and a brief summary of the central principles of connectionist models relating to time-dependent information processing, a new model of categorisation response times based on the ALCOVE model is proposed. The new model, ALCOVE(RT), augments the ALCOVE model with a cascade activation function which gives the model the ability to make predictions of categorisation response times based on the gradual accumulation of information in the processing units.

In Chapters 5 and 6, two speeded categorisation experiments are described. The EGCM, EBRW and ALCOVE(RT) models are applied to the combined classification accuracy and RT data and the results of these applications analysed. Several conclusions about the capability of the various assumptions embodied in the models to account for the data are drawn.

In Chapter 7, the relationship between stimulus frequency and categorisation performance is discussed. A number of early experimental findings for and against the notion that frequency affects graded category structure are outlined. Next, two more recent investigations carried out by Nosofsky (1988) and Ashby, Boynton and Lee (1994) into the effects of frequency on different measures of categorisation performance are described in detail. As the discussion illustrates, the results from these later experiments provide conflicting evidence about the existence of frequency effects. This chapter serves to provide a theoretical background for the discussion which follows in subsequent chapters.

In Chapter 8, A recent experiment carried out by Nosofsky and Palmeri (1997, Experiment 2) designed to compare classification RT predictions from the RT-D and EBRW models in relation to stimulus frequency is described and details of the application of the EBRW and EGCM to the data are outlined. The ALCOVE(RT) model is then applied to the data.

In Chapters 9 and 10, two speeded classification experiments designed to analyse the effect of stimulus frequency on categorisation performance and to compare the categorisation accuracy and RT predictions of the EGCM, EBRW and ALCOVE(RT) models in respect to the manipulation of stimulus frequency are described. The first experiment looks at the development of categorisation performance throughout the course of training whereas the second, while also investigating the category learning process, includes a speeded testing stage after training has taken place. Kruschke's original ALCOVE model is also employed to analyse the course of category learning in relation to manipulations of stimulus frequency.

Finally, in Chapter 11, a review of the research is conducted and a general discussion of the results is undertaken in which conclusions are drawn concerning the effect of stimulus frequency on categorisation. Finally, implications of the models' varying degrees of success for theories of categorisation response times are discussed.

Chapter 2

Models of Categorisation

Introduction

In this chapter, because formal measures of similarity play a fundamental role in all current theories of categorisation, the concept of similarity is first introduced and various attempts to construct an explicit formulation of the notion for use in cognitive analyses are discussed. Four models of categorisation are then outlined—the *Context Model* (CM; Medin & Schaffer, 1978), the *Generalised Context Model* (GCM; Nosofsky, 1986), the *Decision-bound Model* (Ashby & Townsend, 1986), and the *ALCOVE* model (Kruschke, 1992) and their individual representational assumptions and processing mechanisms are described in detail. This chapter serves to provide the necessary theoretical background for the discussion in later chapters.

Similarity

The concept of similarity lies at the core of all theories of memory and cognition. The idea that similarity underlies mental processing can be traced as far back as Aristotle's notion of *association by resemblance* (Sorabji, 1972) and was also at the heart of the empiricist philosophical doctrine of the *association of ideas* propounded by Locke (1706) and Hume (1739) in the seventeenth and eighteenth centuries. Later, in the nineteenth century, the central role played by similarity in cognition was acknowledged by William James (1890) who asserted that "this sense of Sameness is the very keel and backbone of our thinking".

One of the most elementary of cognitive processes, the retrieval of information from memory, requires an ability to calculate a measure of similarity between a perceived stimulus and representations of previously encoded stimuli. Upon this foundation, basic cognitive skills such as recognition, identification and categorisation can be built. Not only is similarity essential for such basic functions, however. Much recent research in the cognitive sciences has revealed the important role that the perception of similarity has in scientific discovery and creativity (Vosniadou & Ortney, 1989).

Despite its long history in philosophy and psychology, the earliest explicit formulation of the notion of similarity, however, is to be found in the early twentieth century in

Thorndike's (1913) common elements theory of generalisation. In this theory, Thorndike related the similarity between objects to the number of their common components.

An early formal model of learning and choice behaviour based upon Thorndike's theory—Estes' (1950) *Stimulus Sampling Model* (SSM), equated the similarity between two stimuli with the sum of their common components. This measure of similarity between two stimuli, i and j, denoted by η_{ij}, can be stated formally as

$$\eta_{ij} = \sum_{k=1}^{K} s_k \qquad s_k = \begin{cases} 1 & \text{if } i_k = j_k \\ 0 & \text{if } i_k \neq j_k \end{cases} \tag{2.1}$$

where s_k is the similarity value between the two stimuli on dimension k. The stimulus sampling model predicts that, given a situation in which participants are trained to associate a unique response with each of two stimuli, subsequent generalisation of these responses to novel stimuli is determined by the proportion of common elements between the new and previously learned stimuli. Table 2.1 illustrates this situation by showing the design of a simple experiment carried out by Estes and his associates to test this prediction (for a detailed description of the experiment see Atkinson & Estes, 1963; Estes, 1994). The experiment involved four stimuli, S, the components of which are indicated by lower case letters in Table 2.1. In the experiment, participants were trained to associate a particular response with each of two training stimuli response $R1$ with stimulus $S1$ and response $R2$ with stimulus $S2$. Following training, participants underwent a transfer task in which, in addition to stimuli $S1$ and $S2$, they were also required to respond to the transfer stimuli $S3$ and $S4$. These transfer stimuli had either one or two components in common with the training stimuli, producing the SSM-predicted similarity values (Sim) shown in Table 2.1 and a ratio of similarity between the training and transfer stimuli of 2:1. The stimulus sampling model predicts that the probability of a transfer stimulus being responded to with response $R1$ is proportional to this similarity ratio. The probabilities of response $R1$ for the transfer stimuli predicted by the SSM (Pred) and actually observed in the experiment (Obs) are shown in Table 2.1.

Table 2.1: Transfer Via Common Elements in Stimulus Sampling Model

Stimuli (S)	Stimulus						
Training	Elements			Response (R)			
S1	a	b	c	R1			
S2	d	e	f	R2		P(R1)	
Transfer				Sim(S1)	Sim(S2)	Obs	Pred
S3	b	c	d	$b+c=2$	$d=1$	0.667	0.669
S4	c	d	e	$c=1$	$d+e=2$	0.333	0.332

Although the predictions of the SSM are very accurate in this situation, it was soon discovered that the use of a simple additive similarity rule leads to a serious deficiency in the model. This deficiency came to be known as the *overlap problem* because it concerned a prediction by the SSM that people should never be able to learn to distinguish with 100% accuracy between two stimuli which consist of overlapping components, for example two stimuli having the components (a, b, d) and (b, c, d) respectively. This prediction arose because, according to the SSM, any stimulus presented during training would be equally similar to the two sets of representations of previously observed stimuli to which responses $R1$ and $R2$ had been associated. Experimental evidence, however, (e.g. Uhl, 1964; Robbins, 1970) shows that people can learn to discriminate such stimuli perfectly. A further shortcoming of the SSM is that it is only applicable to situations in which stimuli have binary-valued dimensions.

The overlap problem revealed a serious limitation of additive similarity rules and led to the formulation of an alternative proposal—that common stimulus components combine multiplicatively, (Estes, 1972, 1973; Medin, 1975; Medin & Schaffer, 1978).

According to this new formulation, the similarity of two stimuli on a particular component dimension k is represented by a real valued parameter s_k which can take values ranging between 0 and 1. If two stimuli have a common component, the value of s_k corresponding to that component equals 1 whereas if the stimuli differ on dimension k, s_k is set to a value between 0 and 1. This can be summarised formally as

$$\eta_{ij} = \prod_{k=1}^{K} s_k \qquad s_k = \begin{cases} 1 & \text{if } i_k = j_k \\ (0 \leq s_k \leq 1) & \text{if } i_k \neq j_k \end{cases} . \tag{2.2}$$

Typically, the individual values of s_k are estimated when fitting the model to experimental data. The multiplicative similarity rule does not suffer from the overlap problem as the additive rule does. This can be shown in relation to the example described above. If the value of s_k for components a and c is estimated to be 0.3, a stimulus with components (a, b, d) presented during training would have a similarity value of 1.0 to other previously observed stimuli with the same components but a value of only 0.3 to those stimuli with components (b, c, d). One important limitation of the product rule, however, is that, like the additive rule, it can be applied only in circumstances where stimulus dimensions are binary valued.

An alternative approach to the analysis of similarity comes from the field of multidimensional scaling (MDS) theory. MDS theory refers to a class of mathematical techniques for producing spatial representations of objects representing their similarity relationships in a geometric form based on proximity data. The proximity data upon which such analyses are carried out are typically stimulus similarity ratings, same-different errors or identification confusions (Carroll & Wish, 1974; Kruskal & Wish, 1978). The output of a MDS analysis is usually a multi-dimensional diagram in which stimulus positions are plotted according to their similarity relationships. The importance of MDS theory for cognitive psychology, however, is that it provides a notion of a multidimensional psychological space in which stimuli are represented by the subject, the similarity relationships between stimuli in this space not necessarily corresponding to their actual physical correspondences (Shepard, 1958, 1980; Nosofsky, 1992). A further important advantage of the MDS approach is that it can be applied to stimuli with real-valued dimensions.

According to MDS theory, similarity is a monotonically decreasing function of distance in the co-ordinate space. The distance between two stimuli i and j, d_{ij} is given by the

Minkowski-r power metric formula

$$d_{ij} = \left(\sum_{k=1}^{K} |x_{ik} - x_{jk}|^r \right)^{\frac{1}{r}} \tag{2.3}$$

where K is the number of spatial dimensions, x_{ik} is the value of exemplar i on dimension k and r determines the distance metric employed. Two forms of distance metric are commonly used depending on the nature of the stimuli (Garner, 1974). For stimuli with integral dimensions which are perceived as relatively unanalysable wholes such as colours varying in hue, brightness and saturation, or tones varying in pitch and volume, a Euclidian metric is most suitable, ($r = 2$) whereas for analysable stimuli with separable dimensions such as shapes varying in size and orientation, a city-block metric is more appropriate ($r = 1$).

A great deal of empirical evidence has shown that the function relating similarity to distance is exponential in form (Shepard, 1987; Luce, 1963). This relationship is given by

$$\eta_{ij} = \exp\left(-d_{ij}^q\right) \tag{2.4}$$

where q, which determines the shape of the function equals 1, (if q equals 2 the function is Gaussian). A relationship between the multiplicative similarity rule proposed by Medin and Schaffer (1978) and the MDS conception of similarity was revealed by Nosofsky (1984) who showed that the multiplicative rule is a special case of psychological distance in MDS using a city-block metric and of similarity when it is an exponential decay function of distance.

MDS theory has provided an indispensable framework in which to conceptualise psychological space and a number of powerful techniques with which to understand and investigate the structure of mental representations. The importance of MDS is acknowledged by Nosofsky (1992) who claimed that the incorporation of MDS theory had been one of the most significant advances in cognitive psychology in the last 40 years. The MDS framework underlies much of the analysis and research reported below.

The Context Model

The *Context Model* (CM; Medin and Schaffer (1978) was the first of a number of related *exemplar-similarity* models of categorisation (Estes, 1994). The CM combines the multiplicative rule for computing similarity (Equation 2.2) with a rule for stimulus categorisation based on the ratio of the similarities between the stimulus and the alternative categories. The CM assumes that stimuli are represented as individual instances or *exemplars* in memory and that a category is represented by the set of stored exemplars which have been associated with that particular category label. Formally, the context model states that the probability of categorising stimulus i as belonging to category J, $P(J|i)$, is given by

$$P(J|i) = \frac{\sum_{j \in J} \eta_{ij}}{\sum_K \left(\sum_{k \in K} \eta_{ik}\right)} \tag{2.5}$$

where η_{ij} represents the similarity between stimuli i and j and is computed according to Equation 2.2 and $j \in J$ refers to all stored exemplars which have category label J. The rule described by Equation 2.5 defines the probability of categorising stimulus i as a member of category J as the quotient of the division of the summed similarity of i to J by the summed similarity of i to all other categories. This formulation of stimulus categorisation can be viewed as a category-level extension of Luce's (1963) *Similarity Choice Model* (SCM) of stimulus identification, according to which the probability of incorrectly identifying stimulus i as stimulus j is given by

$$P(j|i) = \frac{\beta_j \eta_{ij}}{\sum_k \beta_k \eta_{ik}} \tag{2.6}$$

where η_{ij} is the similarity between stimuli i and j, ($\eta_{ij} \leq 1$, $\eta_{ij} = \eta_{ji}$, $\eta_{ii} = 1$) and β_j, ($0 \leq \beta_j$, $\sum \beta_j = 1$) is the bias for making response j (Nosofsky, 1984, 1986).

The context model has been extremely successful in accounting for data in numerous categorisation experiments (Busemeyer, Dewey, & Medin, 1984; Medin, Altom, Edelson, & Freko, 1982; Medin, Altom, & Murphy, 1984; Medin, Dewey, & Murphy, 1983; Medin & Schaffer, 1978; Medin & Smith, 1981) and has been the foundation upon which a number of recent exemplar-based categorisation models have been constructed.

The Generalised Context Model

As its name suggests, the *Generalised Context Model* (GCM; Nosofsky, 1984, 1986) is one such model developed from Medin and Schaffer's context model. The GCM extends the CM by incorporating the original multiplicative similarity analysis into the framework of MDS theory described above and augmenting the similarity computation with variables representing overall stimulus discriminability and selective attention to stimulus dimensions. These changes are implemented by modifying the Minkowski-r metric formula (Equation 2.3) so that it becomes

$$d_{ij} = c \left(\sum_{k=1}^{K} w_k \left| x_{ik} - x_{jk} \right|^r \right)^{\frac{1}{r}} \tag{2.7}$$

where w_k is the *attention weight* of stimulus dimension k, ($w_k \leq 1, \sum w_k = 1$) and c is a scaling parameter that determines overall sensitivity to differences between exemplars in psychological space ($0 \leq c < \infty$). A large value of c produces a rapid decrease in similarity with increases in difference whereas when c is small, the reduction of similarity is more gradual. The value of c typically increases with exposure to stimuli to reflect the greater level of discriminability which occurs over the course of training. The dimension weights are model parameters which correspond to the differential similarity values, s_k in the CM and represent the application of selective attention to stimulus dimensions. The result of increasing the selective attention applied to a dimension is to stretch the space in that dimension and shrink the space in other dimensions thereby increasing the significance of the difference on that dimension in the distance computation. This process is illustrated in Figure 2.1 which depicts eight stimuli varying on three binary dimensions, size, colour and shape.

The position of each stimulus corresponds to its value on each of the dimensions and the three dimensions combined form the axes of a rectangle. Figure 2.1A shows the distance relationships between the stimuli when equal attention is applied to the three dimensions.

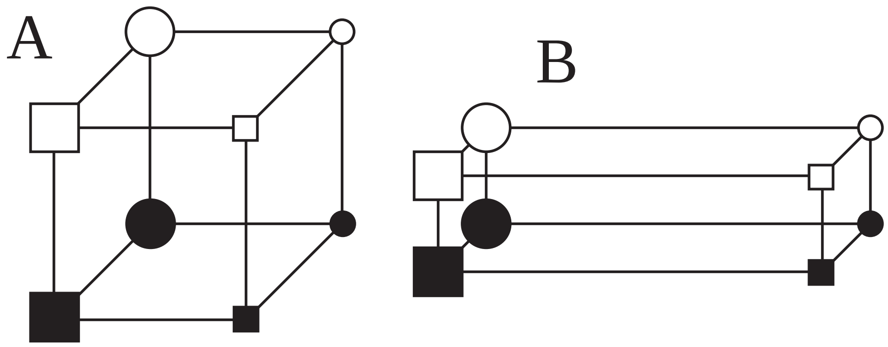

Figure 2.1: Schematic illustration of the role of selective attention in modifying similarities among exemplars. Adapted from Nosofsky (1986)

Figure 2.1B shows the situation when selective attention is applied to the size dimension the space is stretched along that dimension and shrunk on the shape and colour dimensions. This has the effect of accentuating the perceived difference between the large and small stimuli and increasing the perceived similarity between circular and square and black and white stimuli.

In addition to the extension of the Minkowski-r metric formula, the GCM also modifies the context model's response rule with the addition of a bias parameter for each category, β_J which represents the bias for making category response J, $(0 \leq \beta_J, \sum \beta_J = 1)$, to form

$$P(j|i) = \frac{\beta_j \sum_{j \in J} \eta_{ij}}{\sum_k \left(\beta_k \sum_{k \in K} \eta_{ik}\right)} \tag{2.8}$$

This modification also has the effect of making Luce's SCM a special case of the GCM when each stimulus defines its own category, (i.e. in a stimulus identification situation).

Like the CM, the GCM has been extremely successful in providing quantitative accounts of many categorisation experiments (e.g. Nosofsky, 1984, 1985, 1986, 1987, 1988, 1989, 1991) and demonstrating the close relationship between stimulus identification, old-new recognition and categorisation (Nosofsky, 1986, 1988, 1992).

The Decision Bound model

The decision-bound model of categorisation is related to the GCM in that it also assumes a MDS geometric representation of stimuli. However the two models differ regarding their information access and category decision mechanisms. The decision-bound model is derived from *General Recognition Theory* (GRT; Ashby & Townsend, 1986; Ashby & Gott, 1988; Ashby & Perrin, 1988) which is a multidimensional generalisation of *Signal Detection Theory* (Green & Swets, 1966). The central assumption of GRT is that individual experiences of the same stimulus vary in terms of their representation due to perceptual noise and that consequently, each stimulus is represented by the subject, not as a single point, but as a multivariate probability distribution of points in perceptual space. According to this approach, categories are represented as probability mixtures of the individual exemplars distributions which define the category. During category learning, participants are assumed to partition the space and associate a category label with each created region. The line or curve separating two response regions is known as the *decision bound* and it is the position and shape of this bound which are assumed to be optimised to maximise response accuracy when categories are learned.

Tests of the decision-bound model have commonly relied on a number of assumptions about the nature of the stimuli and category structures. Typically stimuli vary continuously on two dimensions and categories take the form of a bivariate normal probability density function. An example of this type of function is shown in Figure 2.2.

In this surface plot, the height of the function at each point represents the probability that a stimulus chosen at random has values on the x and y axes associated with that point. In a two choice categorisation paradigm, two such distributions are defined, one representing each category. A useful method of depicting such situations is in terms of *contours of equal likelihood*. An example is shown in Figure 2.3A which depicts probability distributions representing two alternative categories and Figure 2.3B which shows the corresponding contours of equal likelihood and optimal decision bound. The sets of concentric circles in Figure 2.3B are similar to the contours on a geographical map representing height above sea-level in that they are slices parallel to the (x, y) axis plane at different values on the z axis. As such, they represent regions of the (x, y) plane in each probability distribution which have equal probability density.

The category decision process in GRT is specified in terms of a deterministic discriminant function $f(x)$ where x is a vector of stimulus co-ordinates in perceptual space. This function is defined so that values of x which fall on the decision bound satisfy $f(x) = 0$, whereas those which fall on either side of the bound satisfy $f(x) < 0$ and $f(x) > 0$ respectively. This is the discriminant function constructed by the optimal classifier (Fukunaga, 1990; Bishop, 1995) which can be reformulated as the categorisation rule

- If $f(x) < 0$ then assign x to Category A
- If $f(x) = 0$ then guess
- If $f(x) > 0$ then assign x to Category B

However, there is experimental evidence (e.g. Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992) to suggest that participants do not employ an optimal discriminant function during categorisation.

Ashby and his colleagues have suggested four principal reasons for this suboptimal behaviour: (a) participants do not know the locations of every exemplar in all relevant categories, (b) the presence of noise in perceptual and memory retrieval processes, (c)

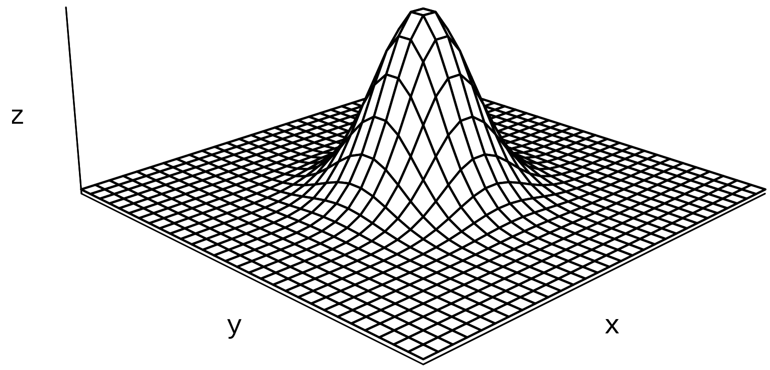

Figure 2.2: Surface plot of a bivariate normal probability density function

participants are not aware of the parameters defining the perceptual noise distributions, (d) participants may have an irrational bias towards one of the alternative categories (Maddox & Ashby, 1993). Consequently, Ashby *et al.* have proposed a suboptimal discriminant function having the form

- If $f(x + e_p) + e_c < \delta$ then assign x to Category A
- If $f(x + e_p) + e_c = \delta$ then guess
- If $f(x + e_p) + e_c > \delta$ then assign x to Category B

where e_p is a multivariate normally distributed random vector with mean 0 and variance σ^2 representing *perceptual noise*, e_c is a normally distributed random variable also with 0 mean and variance σ^2 representing *criterial noise*, (i.e. trial-by-trial variability in the participant's recall of the decision bound location), and δ is a parameter representing category bias. The category bias works by shifting the decision bound towards the centre of one distribution and away from the other.

A positive value of δ represents a bias towards category A by moving the decision bound closer to the mean of the category B distribution. One consequence of the inclusion of perceptual and criterial noise and a category bias parameter is that the discriminant function is no longer deterministic in the sense that presentations of the same stimulus no longer always evoke the same response. The decision rule remains deterministic, however, in the sense that if $f(x + e_p) + e_c < \delta$ the probability of a category A response = 1. The discriminant function defining the decision bound can have several other forms and Ashby and his colleagues have analysed a number of them (Ashby, 1992; Ashby & Gott, 1988). These functions can be distinguished by whether they integrate stimulus dimension information into one decision boundary, (e.g. the general quadratic classifier, the general linear classifier, and the minimum distance classifier), or whether they make separate

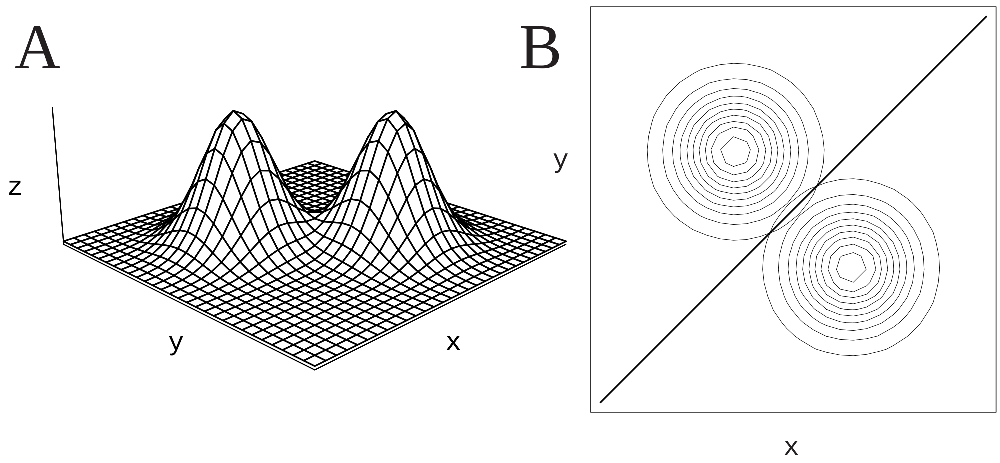

Figure 2.3: Surface plot of two bivariate normal probability distributions (A) with corresponding plot of contours of equal likelihood and optimal decision bound (B)

decisions about individual stimulus dimension values and then combine this information in a subsequent stage to select a response (e.g. the independent decisions classifier).

Over the last ten years, general recognition theory has been one of the leading proposals for an account of multidimensional perceptual categorisation. GRT has been the framework for a theory of similarity (Ashby & Perrin, 1988; see also Nosofsky, 1992) and models based on GRT have been successfully utilised to investigate the relation between identification, similarity judgement and categorisation (Ashby & Lee, 1991) and the effect of category base-rate information on classification performance (Maddox, 1995). The significant application of GRT in relation to the subject of this research, however, is in the modelling of categorisation reaction times, (Ashby, Boynton & Lee, 1994; Ashby & Maddox, 1991, 1994; Maddox & Ashby, 1996). This subject will be discussed at length in Chapter 3.

The ALCOVE model

ALCOVE (Attention Learning COVEring Map; Kruschke, 1992; Nosofsky & Kruschke, 1992; Nosofsky, Kruschke, & McKinley, 1992), is a connectionist model of category learning in which the exemplar-based category representations of the GCM are combined with the error-driven supervised learning mechanism used in Gluck and Bower's (1988a, 1988b) adaptive network models of categorisation.

The ALCOVE model is a feed-forward network consisting of three layers of processing units; a layer of input units representing stimulus dimension values, a hidden layer of radial-basis function units representing the stored exemplars' positions in multidimensional space, and an output layer of units which represent the alternative categories. Although the main architectural features of ALCOVE are founded on the representational assumptions of the GCM, the models differ in two important respects. Firstly, ALCOVE assumes that participants learn associations between exemplars and categories and provides a mechanism for this associative learning. Secondly, ALCOVE includes a mechanism for learning dimensional attention strengths rather than simply estimating them from the data. These two additional features are represented in ALCOVE by the connection weights between the exemplar and category units and the attention weights gating the input units respectively. The basic architecture of the ALCOVE model with three input units, six hidden units and two output units is illustrated in Figure 2.4.

Two versions of ALCOVE have been formulated and tested although several others are possible, (see Kruschke, 1992). In the first, hidden units are randomly distributed throughout the multidimensional space to create a *covering map*, whereas in the second, a hidden node is explicitly fixed at the specific location in the multidimensional space of each of the training exemplars. The main results of testing ALCOVE reported by Kruschke (Kruschke, 1992; Nosofsky & Kruschke, 1992; Nosofsky, Kruschke & McKinley, 1992) concern the latter type, and therefore this is the version discussed here, (see Kruschke, 1990a, 1990b for reports of the covering map version). This version is also more similar to the GCM in that it embodies an exemplar representation of stimuli. Typically, the locations of the stimuli in psychological space are determined by multidimensional scaling as in the case of the GCM or alternatively can be assumed to be identical to the actual physical dimensions of the stimuli.

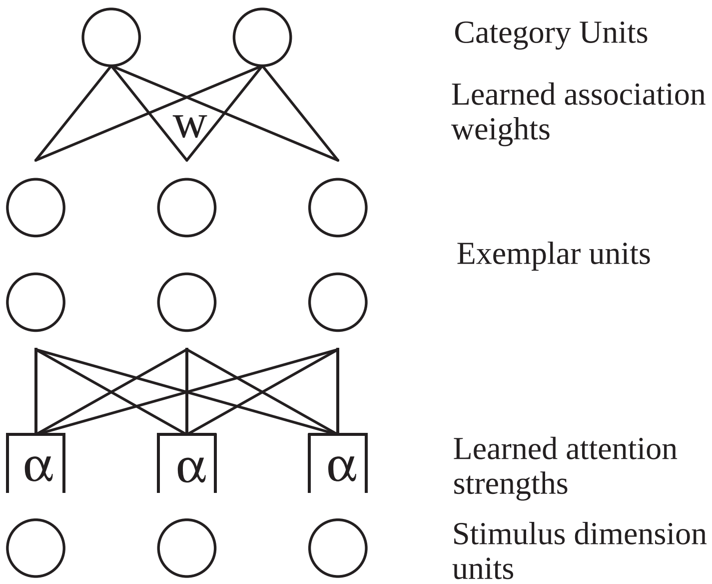

Figure 2.4: The architecture of ALCOVE (adapted from Kruschke, 1992)

During the category learning phase, the association weights between exemplar and category units and the dimensional attention strengths are adjusted by the method of gradient descent on error according to the generalised delta rule (Rumelhart, Hinton, & Williams, 1986). This procedure has the effect of minimising an error value, E defined as a point in a multidimensional error surface corresponding to the vector of association or attention weights. Weight values are adjusted in proportion to the size of their contribution to the error so that the corresponding change in the weight vector and the value of E is via the steepest descent of the error surface. When a training stimulus is presented to the model, the hidden units are activated according to their similarity to the pattern. The function governing the activation of hidden unit i, denoted a_i^{hid}, $(0 \leq a_i^{hid} \leq 1)$, is formally identical to the similarity function used in the GCM (Equations 2.4 and 2.7) and is given by

$$a_i^{hid} = \exp\left[-c\left(\sum_j \alpha_j \left|h_{ij} - \alpha_j^{in}\right|^r\right)^{\frac{q}{r}}\right] \tag{2.9}$$

where c is a freely estimated parameter determining the specificity of the hidden unit, h_{ij} is the value of hidden unit i on dimension j, α_i is the attention strength associated with dimension i, $(0 \leq \alpha_i)$, r is the constant which determines the psychological distance metric as in Equation 2.7, and q is a constant which determines the similarity function. The c parameter corresponds to the generalisation parameter c in the GCM and governs the rate at which a hidden unit is activated (i.e. how steep is the gradient of the units' activation function). A hidden unit identical to a stimulus pattern has a maximum activation value of 1. When a city-block distance metric and an exponential similarity gradient are employed, the activation profile of the hidden units takes a characteristic pyramidal form as shown in Figure 2.5.

Category units are activated according to the weighted sum of the activations of the hidden units, a_j^{hid}, to which they are connected. The activation of category unit i, denoted a_i^{out}, is given by

$$a_i^{out} = \sum_{\substack{hid \\ j}} w_{ij} a_j^{hid} \tag{2.10}$$

where w_{ij} is the association weight between output unit i and hidden unit j. Output error is computed and fed back to the association weights and dimensional attention strengths which are adjusted accordingly. The error term, E, used in the generalised delta rule is defined as

$$E = \frac{1}{2} \sum_{\substack{out \\ i}} \left(t_i - a_i^{out}\right)^2 \tag{2.11}$$

where t_i is the *teacher value* provided to each output unit. The teacher value for category unit i on presentation of stimulus S_j is computed according to the rule

$$t_i | S_j = \begin{cases} \max\left(+1, a_i^{out}\right) & \text{if } S_j \in I \\ \min\left(-1, a_i^{out}\right) & \text{if } S_j \notin I \end{cases} \tag{2.12}$$

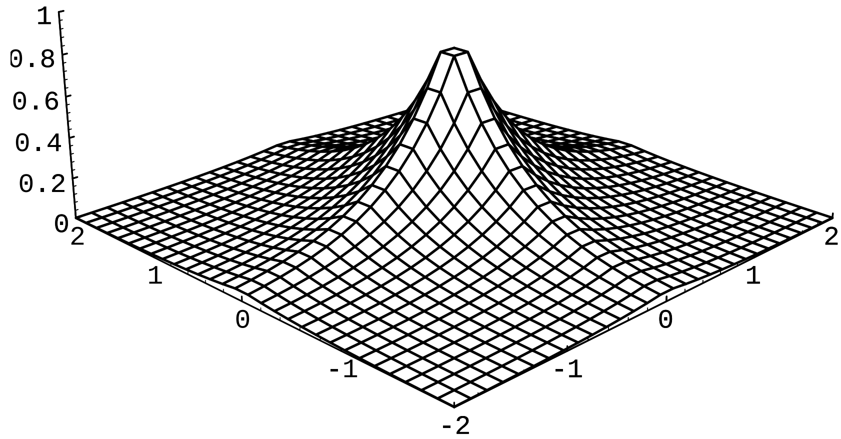

Figure 2.5: Surface plot of exemplar unit activation profile in ALCOVE model (Equation 2.9) at location $(0, 0)$ in 2D space with $r = q = 1.0$, $\alpha_1 = \alpha_2 = 1.0$ and $c = 1.5$. Adapted from Kruschke (1993)

where $S_j \in I$ is to be read as "stimulus j is a member of category I". This definition allows category units to produce activation values greater or less than the teacher values, the justification for this being that in categorisation experiments, participants are generally only given feedback regarding a stimulus' category membership and not to the degree of membership.

In the error back-propagation phase of training, the change in the association weight between category unit i and exemplar unit j, w_{ij}^{out}, is given by

$$\Delta w_{ij}^{out} = \lambda_w \left(t_i - a_i^{out}\right) a_j^{hid} \tag{2.13}$$

where λ_w $(0 < \lambda_w)$ is a single freely estimated *association learning rate* parameter applied to all output weights which determines the rate of change. The corresponding function which determines the change in the dimensional attention strengths is defined as

$$\Delta\alpha_i = -\lambda_\alpha \sum_{\substack{hid\\j}} \left[\sum_{\substack{out\\k}} \left(t_k - a_k^{out}\right) w_{kj} \right] a_j^{hid} c \left| h_{ji} - a_i^{in} \right| \tag{2.14}$$

where λ_α $(0 < \lambda_\alpha)$ is the *attention learning rate* parameter applied to all attention strengths. Output unit activation values are converted to response probabilities according to Luce's (1963) choice rule

$$P\left(I\right) = \frac{\exp\left(\phi a_i^{out}\right)}{\sum_{\substack{out\\i}} \exp\left(\phi a_i^{out}\right)} \tag{2.15}$$

where ϕ is a real valued mapping constant to scale the activation value.

The training phase consists of the presentation of the set of training stimuli for a fixed number of times or until the difference between the desired output and the actual out is sufficiently small. This procedure may also be embedded within a parameter estimation algorithm to optimise the values of the four free parameters, (the exemplar unit specificity, c in Equation 2.9, the response mapping constant, ϕ in Equation 2.15, the association learning rate, λ_w in Equation 2.13 and the attention learning rate, λ_α in Equation 2.14) from experimental data.

The assumptions about category learning embodied in ALCOVE's learning mechanisms are significantly different from those of the GCM. Firstly, according to the GCM, category learning takes place by the incremental addition of exemplars in memory over training trials. In contrast, the learning process in ALCOVE is driven by categorisation performance so that changes in the model only occur when there is error between the actual and desired output. Secondly, category learning in ALCOVE is *interactive* in the sense that, on each trial, the associative connections between all of the exemplar units and the category units are adjusted relative to their activations and the amount of error, rather than just the connection between the most similar unit, (i.e. that corresponding to the stimulus pattern), and the category unit. Kruschke has argued that it is this property of ALCOVE which allows it to account for the *base-rate neglect* phenomenon originally observed by Gluck and Bower (1988a, 1988b) and subsequently examined by several other researchers (Estes, Campbell, Hatsopoulis & Hurwitz, 1989; Shanks, 1990; Nosofsky, Kruschke & McKinley, 1992).

The effect of base-rate neglect has been most frequently studied in a two-choice category learning paradigm in which the categories, represented as two diseases, are learned on the basis of a set of symptoms which are probabilistically associated with both of the alternative categories. On each learning trial, a hypothetical patient with a set of symptoms is shown to a participant who is required to make a decision as to which disease is present. The frequency of occurrence for the two diseases during the training stage is unequal, one (the *common* disease, C) occurring three times more often than the other (*rare* disease, R). One of the symptoms (s_1) has a relatively high conditional probability (.6) of being present when the rare disease is correctly diagnosed and a relatively low conditional probability (.2) of being present when the common disease is correctly diagnosed. According to Bayes' theorem, therefore, although s_1 is highly correlated with R, the relatively low frequency of occurrence of R in relation to C means that neither disease can be predicted on the basis of the presence of s_1 alone. The primary finding of the studies is that subjects' behaviour diverges from the predictions of Bayes' theorem by showing a systematic overestimation of the predictability of s_1 for R—when presented with s_1 alone, subjects consistently neglect the base rate information and predict that disease R is present. These results have remained unaccounted for by existing exemplar-based models and even the ability of ALCOVE to account for this phenomenon is not generally applicable to the full range of experimental circumstances (Lewandowsky, 1995).

ALCOVE has been very successful, however, in accounting for trial-by-trial category learning and transfer data from several experiments (Kruschke, 1992; Nosofsky, Kruschke & McKinley, 1992) and has demonstrated a number of important properties which are uncharacteristic of standard back-propagation networks. For example, ALCOVE is able to learn nonlinearly separable categories faster than linearly separable ones and can learn categories comprised of stimuli with correlated dimensions. Because of its exemplar-based internal representations, ALCOVE does not suffer from the problem of catastrophic forgetting, a characteristic of back-propagation networks where previously learned associations are forgotten when the network is trained on novel associations (Kruschke, 1992, 1993). Finally, ALCOVE is able to exhibit the three-stage learning of rule exceptions observed in children's development of language acquisition (Kruschke, 1992).

Summary

The primary purpose of this chapter is to introduce the fundamental theoretical background to this research and to place the following discussion within the broad context of formal investigations into multidimensional perceptual categorisation. In the first section, the central role of similarity in cognition was discussed and the various attempts to apply a formal analysis to the concept of similarity were outlined. Following this, four of the most influential models of categorisation were introduced and their structures, mechanisms and theoretical assumptions were described in detail.

In the next chapter, this introductory overview is expanded with a discussion of recent attempts to analyse the time course of the categorisation process with three formal models of categorisation reaction times (RT). The relevance of this initial overview will become apparent when it is demonstrated that each of these RT models is an extension of one of the models discussed above.

Chapter 3

Categorisation Response Times

Introduction

Arguably one of the most interesting developments in categorisation research in recent years has been the attempt by several investigators to formulate theories of categorisation response time (RT). Prior to this, researchers have focused primarily on categorisation accuracy as the dependent variable in their experiments. Following an initial summary of psychological research into reaction times to set the context for the discussion, this chapter reviews the three main theories of categorisation RT.

The study of response times in psychology

The idea that response times contain information about underlying cognitive processes has a long history in psychology. The earliest recognition of the temporal properties of human responses came at the end of the eighteenth century when it was discovered that measurements of stellar transits produced by individual astronomers exhibited systematic and significant differences. This discovery led the astronomer Friedrich Bessel in 1816 to carry out one of the first experimental comparisons of individual response times and to formulate a corrective value known as the *personal equation* for each astronomer to compensate for the variation in individuals RTs.

By the middle of the nineteenth century, reaction time measurements were being employed by leading experimentalists—Wilhelm Wundt, Hermann Helmholtz and F. C. Donders—in simple RT experiments to investigate various aspects of behaviour. Wundt studied, amongst other things, the effect of attention on response time and Helmholtz investigated the speed of impulse transmission in human nerve cells by measuring the relative speed of responses to stimuli applied to various parts of the body. Perhaps the most influential work involving response times, however, was that of Donders (1969) who distinguished three classes of RT tasks and proposed a technique for revealing psychological processes using RT measurements known as the *method of subtraction.*

According to Donders, RT tasks could be classified into three categories according to their complexity, *simple* RT tasks, in which participants are required to make a single response as quickly as possible to a single stimulus, (e.g. press a button on the illumination of a light-bulb), *discrimination* RT tasks, in which participants are required to make a single response as quickly as possible to one target stimulus among a number of others, (e.g. press a button on the illumination of one predetermined light-bulb in an array of five bulbs, but not on the illumination of the other four), and *choice* RT tasks, in which participants are required to make a unique response as quickly as possible to each of a number of target stimuli, (e.g. from an array of five buttons, each uniquely associated with one of five light-bulbs, press the associated button on the illumination of a light-bulb).

The basic premise behind the method of subtraction is that separate cognitive components are utilised at different stages in the performance of a task and that these stages are carried out serially, each stage taking a certain amount of time to execute its function. Given this premise, Donders reasoned that inferences about the existence and global structure of these components may be made by comparing the reaction times generated by two experimental conditions, one of which includes a particular component or stage absent in the other. Subtracting the mean RT of the eliminated-component/stage condition from the included-component/stage condition gives the mean RT of the component/stage under investigation. Donders' applied this reasoning to the analysis of the three classes of RT tasks described above, arguing that tasks involving more stages take longer to perform. According to Donders' analysis, therefore, simple RT tasks, (which require stages to perceive the stimulus and execute a response) take less time than discrimination RT tasks, (which require the same stages as simple tasks plus an additional discrimination stage to distinguish which of the alternative stimuli was presented), which, in turn, take less time than choice RT tasks, (which require the same stages as discrimination tasks plus an additional choice stage to decide which of the alternative responses to make). Using the method of subtraction, Donders was able to calculate the time required for individual stages in each of the three task categories, (on the assumption that discrimination stage time is equal to discrimination task time minus simple task time and that choice stage time is equal to choice task time minus discrimination task time).

Although Donders' approach has been widely criticised, both for its methodological and theoretical assumptions (e.g. Posner, 1978; Pylyshyn, 1979, 1984; Sternberg, 1975) and the method of subtraction largely discredited (Woodworth, 1938; Woodworth & Schlosberg, 1954), it is important for introducing the idea of cognition as a series of time-dependent information processing stages. This idea continues to play an important role in current research, primarily as a result of Sternberg's (1969a, 1969b) highly influential extension of Donders' approach called the *additive factors method*. Sternberg's method differs from Donders' in that stages are no longer considered to be simply concatenated in a series with no effect upon each other and which can be simply inserted or removed in the accomplishment of different tasks. In contrast, the additive factors method rests on the weaker premise that, rather than using (potentially incompatible) tasks to test hypotheses about the existence of stages, experimental manipulations may be employed to affect individual or multiple stages of a single cognitive task. Like Donders, Sternberg does assume, however, that only one processing stage can be active at any one time and that the duration of a process is not affected by that of any other. According to the additive factors method, effects of manipulations can be used to generate hypotheses about the existence of and relationships between processing stages. For example, if two factors have additive effects on RT, they can be assumed to be influencing two separate processing stages. Alternatively, if two factors have an interactive effect on performance, it can be assumed that they are influencing a single stage. By this method, factors affecting all of the supposed stages can be accumulated to account for the complete processing course of a task.

Sternberg's additive factors methodology is still commonly employed by experimental psychologists and many of the basic assumptions behind the techniques are widely (although often implicitly) held (Luce, 1986). The underlying assumption that cognitive processing consists of a set of separate subprocesses or stages is not without its critics, however (e.g. McClelland, 1979). The rise of connectionism since the mid 1980's has resulted in the construction of several cognitive models in which the individual components process information simultaneously and continuously over time (e.g. Anderson, 1991; McClelland, 1993; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). This approach will be discussed in greater detail in Chapter 4.

Models of categorisation response times

There are currently three alternative formal theories of categorisation RTs and each is related to one of the categorisation models described in Chapter 2. The first proposal put forward was the *RT-Distance* (RT-D) hypothesis (Ashby, Boynton & Lee, 1994) which is based on the decision bound model of categorisation. The other two models are the *Exemplar-Based Random Walk* (EBRW) model (Nosofsky & Palmeri, 1997) and the *Extended Generalised Context Model* (EGCM; Lamberts, submitted) which are both extensions of the GCM.

The RT-Distance model

The central claim of the RT-Distance (RT-D) model is that, like categorisation accuracy, categorisation reaction time is inversely related to the distance between the perceptual representation of the stimulus and the category decision bound. This relationship can be expressed by the intuition that stimulus representations near the decision bound, being similar to representations from the alternative categories, are ambiguous in terms of their category membership and are therefore categorised less accurately and more slowly. In contrast, the category membership of representations further away from the decision bound is more determinate and consequently categorisation is more rapid and accurate. One prediction which the RT-D model makes in relation to this claim therefore, is that, under certain general experimental conditions, the median RT of incorrect categorisation responses will be greater than that of correct responses. Original tests of the RT-D hypothesis did not assume any functional form for the relationship between distance from decision bound and RT but simply made predictions at the ordinal level. More recently, Maddox and Ashby (1996) have tested two versions of the model based on different assumptions about the form of this function. The first version assumes that the decision time for a stimulus x, t_x, decreases exponentially with increases in the distance of the stimulus percept from decision bound, d_x,

$$t_x = \alpha e^{-\beta d_x} \tag{3.1}$$

where α and β are constants determining the shape of the function. The second version assumes that the decrease in decision time is a power function of distance from decision bound,

$$t_x = \alpha d_x^{-\beta} \tag{3.2}$$

Ashby *et al.* (1994) tested the original RT-D model in three speeded categorisation experiments and discovered that, in general, median RTs of incorrect categorisation responses were greater than those of correct responses and also found a negative correlation between categorisation and distance from decision bound, which is the relationship predicted by the RT-D model. One further result of these experiments which has relevance for later discussion was the discovery that stimulus frequency had no significant effect on categorisation response time. These findings are examined more thoroughly in Chapter 5.

The Exemplar-Based Random Walk model

The Exemplar-Based Random Walk (EBRW) model (Nosofsky & Palmeri, 1997) combines the representational assumptions of GCM with the processing assumptions of a previously existing model of skill development—Logan's (1988) *instance-based* model of automaticity. The purpose of the EBRW is to provide a more detailed account of the memory access and decision making processes involved in categorisation and to supply a time dependent mechanism for predicting categorisation RTs. The assumptions about how exemplars are represented and how distance and similarity are computed remain largely unchanged from the standard GCM. The EBRW assumes that, upon presentation of a probe stimulus, each exemplar is activated according to its similarity to the stimulus and its *strength* in memory. According to the EBRW, the strength of an exemplar can be affected by various factors such as frequency and recency of presentation. The activation of each exemplar i given presentation of stimulus j, a_{ij} is, therefore given by

$$a_{ij} = \eta_{ij}\mu_i \tag{3.3}$$

where η_{ij} is the similarity between stimulus i and exemplar j and μ_i is the strength of i. When a stimulus is presented, stored exemplars race to be retrieved from memory. The time each exemplar takes to be retrieved is an exponentially distributed random variable with a rate proportional to its activation. The probability density of this occurring at time t is described by

$$f(t) = a_{ij}e^{-a_{ij}t}. \tag{3.4}$$

This race is the basis of the random walk process which produces the category response. A random walk process is defined as a trajectory initialised at the origin which, at each time step, t, moves to the left with probability p or to the right with probability $q = 1 - p$. In the EBRW, the stochastic process governing each movement is the exemplar retrieval mechanism with the direction of each step being determined by the category label of the retrieved exemplar. In a two-choice decision task, the random walk counter is driven towards one of two response barriers. If the exemplar is associated with category A, the random walk counter is incremented in the direction of the category A response barrier whereas if the exemplar has a category B label, the counter is decremented towards the direction of the response barrier for category B. A category response is made when the counter reaches one of the response barriers and the associated RT for that response is a function of the number of time steps taken to make the response. This process is illustrated in Figure 3.1.

The predicted RT for a given stimulus is a function of two factors—the total number of steps taken to reach the decision boundary and the time taken to retrieve each exemplar from memory. The first factor is represented in the model by a constant value, α which is added to each winning exemplar retrieval time, t_w to produce a total time for each step, T_{step}

$$T_{step} = \alpha + t_w. \tag{3.5}$$

Nosofsky and Palmeri interpret α as representing the time required to extract the category label from the winning exemplar and then to increment the random walk counter based upon this information. The central prediction of the EBRW is that reaction time is fastest for stimuli which are least ambiguous and conversely, is slower for those stimuli which are closer to category boundaries. This is essentially the same prediction as that made by the RT-D model.

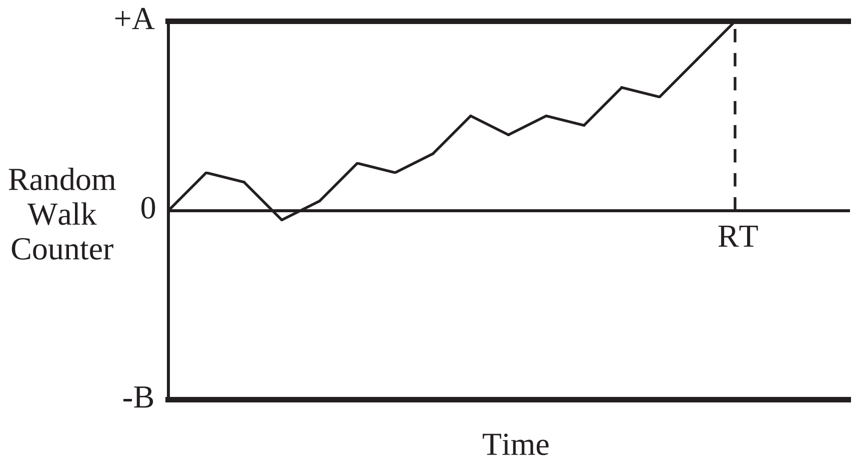

Figure 3.1: Schematic illustration of the random walk process in the EBRW model

The EBRW also predicts that frequency of presentation also has an effect of RT because of a property of the model which relates exemplar retrieval time to the number of exemplars stored in memory. This relationship can be described simply as being one in which an increase in the number of exemplars results in faster winning retrieval times. Because the relationship is a consequence of certain properties of exponential distributions, it will be necessary to discuss some of these properties briefly before continuing with the explanation of the model.

Given an exponentially distributed process p_i with rate λ_i, the expected finishing time of p_i, $E(T_i)$, is $\frac{1}{\lambda_i}$. Furthermore, given a set of n parallel exponentially distributed processes, $p_1, \ldots, p_n$, each with rate λ_i, a separate process may be defined as the minimum finishing time of these processes. This process is also exponentially distributed with a rate equal to $\sum_{i=1}^{n} \lambda_i$. Therefore, the expected minimum finishing time of a set of n exponentially distributed processes, $E(T_n)$, is $\frac{1}{\sum_{i=1}^{n} \lambda_i}$. Consequently, the expected minimum finishing time of a set of processes depends upon the number of processes taken into consideration because the value of $E(T_n)$ decreases as the value of $\frac{1}{\sum_{i=1}^{n} \lambda_i}$ increases. Roughly translated into terms of the EBRW model, the more exemplars there are stored in memory, the higher the probability that one of them will have a very short retrieval time. Finally, the probability that a particular processes, i is the first to finish is $\frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$.

To relate the above analysis to the predictions of the EBRW, the expected retrieval time for the winning exemplar amongst n exemplars given stimulus i, $E(T_w|i)$, is defined as

$$E\left(T_w|i\right) = \frac{1}{\sum_{j=1}^{n} a_{ij}} \tag{3.6}$$

where a_{ij} is the activation of exemplar j given stimulus i. From Equation 3.5, the expected time for taking each step in the random walk for stimulus i, $E(T_{step}|i)$, is

$$E\left(T_{step}|i\right) = \alpha + E\left(T_w|i\right) \tag{3.7}$$

The probability of taking a step towards the criterion $+A$, p_i is computed by summing the retrieval probabilities of all the exemplars in category A and dividing the result by the summed retrieval probabilities of all exemplars in all categories

$$p_i = \frac{\sum_{j \in A} a_{ij}}{\sum_K \left(\sum_{k \in K} a_{ik}\right)} \tag{3.8}$$

Similarly, the probability of taking a step towards criterion $-B$, q_i is defined as $1 - p_i$ or

$$q_i = \frac{\sum_{j \in B} a_{ij}}{\sum_K \left(\sum_{k \in K} a_{ik}\right)}. \tag{3.9}$$

The following equations have been derived in previous analyses of random walk models, (e.g. Feller, 1968), and so are not derived here. Instead, I will follow the example of Nosofsky and Palmeri (1997) and simply present the results relevant to the current discussion.

The expected number of steps in the random walk given stimulus i, $E(N|i)$ is given by

$$E(N|i) = \begin{cases} \frac{B}{q_i - p_i} - \frac{A+B}{q_i - p_i}\left[\frac{1-(q_i/p_i)^B}{1-(q_i/p_i)^{A+B}}\right] & \text{if } p_i \neq q_i, \\ AB & \text{if } p_i = q_i. \end{cases} \tag{3.10}$$

Therefore, the complete expected time of the random walk given stimulus i, $E(T|i)$, can be computed by multiplying the expected number of steps by the expected time for each step given by Equation 3.7.

$$E(T|i) = E(N|i) \cdot E(T_{step}|i). \tag{3.11}$$

The probability of a Category A response given stimulus i is given by

$$P(A|i) = \begin{cases} \frac{1-(q_i/p_i)^B}{1-(q_i/p_i)^{A+B}} & \text{if } p_i \neq q_i, \\ \frac{B}{A+B} & \text{if } p_i = q_i \end{cases} \tag{3.12}$$

whereas the probability of a Category B response given stimulus i is given by

$$P(B|i) = \begin{cases} \frac{(q_i/p_i)^B-(q_i/p_i)^{A+B}}{1-(q_i/p_i)^{A+B}} & \text{if } p_i \neq q_i, \\ \frac{A}{A+B} & \text{if } p_i = q_i. \end{cases} \tag{3.13}$$

The free parameters in the version of the EBRW studied in this article include the c and attention weight, (w) parameters in the similarity and distance equations; the α parameter in the exemplar-retrieval function; the category criteria parameters, $+A$ and $-B$ in the random walk, (note that the $+A$ and $-B$ parameters are integers), and two parameters used in the linear regression procedure to transform the predicted values, (which are generated by the model in arbitrary units) to response times. The first of these two parameters, (denoted here by k), determines the slope of the regression line while the second, μ_R, determines the y-intercept, which in the current context can be interpreted as the mean of the residual perceptual processing and response execution stages.

The EBRW has been tested in a series of experiments (Nosofsky & Palmeri, 1997, in press) and has proved successful in accounting for both categorisation accuracy and response times in speeded categorisation tasks using integral dimension stimuli and the development of automaticity in skilled performance of a visual numerosity judgement task. The EBRW has also provided an account of the effect of stimulus frequency on categorisation performance discovered in one particular experiment (Nosofsky & Palmeri, 1997, Experiment 2). The results of this experiment will be discussed further in Chapter 5.

The Extended Generalised Context Model

Like the EBRW, the EGCM inherits its representational assumptions from the GCM. Despite this underlying similarity, however, the two models differ considerably in their assumptions about the mechanisms affecting the time course of categorisation. Whereas the EBRW postulates that the time dependent functions are related to the memory access and category decision stages after the stimulus has been perceived, the EGCM assumes that response times are affected by the earlier perceptual processing stage in which information about the stimulus is gradually accumulated.

The EGCM was initially formulated to account for perceptual categorisation under time pressure (Lamberts, 1995; Lamberts & Brockdorff, 1997) and was not originally intended as a general model of categorisation RT. Recent extensions of this early model, however, enable it to be applied to a wider range of RT data (see Lamberts, submitted).

Categorisation, according to the EGCM, is essentially a two stage process. The first stage involves the construction of a perceptual representation of the stimulus and the computation of similarity between this representation and stored exemplar representations in memory. The second stage involves the making of a category decision and the initiation of an appropriate response. The time dependent aspect of this process pertains to the initial stage because it is assumed that stimulus information used in the construction of a perceptual representation is accumulated over a period of time. This information accumulation process is assumed to terminate once the required amount of evidence has been acquired for a category decision to be made and it is this criterion upon which the EGCM account of categorisation RT rests. The EGCM also assumes that the information accumulation process differs depending on the whether the stimulus is composed of separable or integral dimensions. For this reason, the two situations are discussed separately below.

Stimuli with separable dimensions

The EGCM postulates that, upon presentation of a stimulus with separable dimensions, the features, (i.e. dimension values) of the stimulus are sampled and that this sampling process is stochastic, parallel, independent and without replacement. At any given time after stimulus onset, a dimension can be in one of two states in regard to whether it has been sampled or not and whether, therefore, it can be included in the similarity computation. According to the EGCM, the probability that a dimension x is included at or before a given time t, is called the *cumulative inclusion probability* and is described by the function

$$\iota_x(t) = 1 - \exp(-q_x t) \tag{3.14}$$

where q_x is a parameter representing the *inclusion rate* of dimension x. The inclusion rate governs the length of time taken, on average, to process a dimension and is chiefly determined by a dimension's salience so that a high value of q_x produces a correspondingly high inclusion probability for small values of t. The addition of a mechanism for processing dimensions over time has the effect of making the similarity function time-dependent. The similarity between the stimulus representation and the stored exemplars is computed each time a stimulus dimension is processed and so similarity is dependent on which and how many dimensions have been processed. The main features of the similarity function employed by the GCM (Equations 2.4 and 2.7) are retained by the EGCM. The similarity between stimulus i and exemplar j, given the set of processed dimensions, Φ, denoted by $s_{ij}(\Phi)$, is given by

$$s_{ij}(\Phi) = \exp\left[-c\left(\sum_{p\in\Phi} u_p \left|x_{ip} - x_{jp}\right|^r\right)^{\frac{q}{r}}\right] \tag{3.15}$$

where c is a generalisation value and x_{ip} is the value of stimulus i on dimension p as in the GCM and u_p is a *utility value* parameter ($0 \leq u \leq 1, \sum u = 1$) which indicates the importance of dimension p in the computation of similarity by representing the *diagnosticity* of the dimension (Lamberts, 1995).

Each time a stimulus dimension is processed and the similarity between the stimulus representation and the stored exemplars is computed, the similarity values are used to produce a measure of confidence, given the set of processed dimensions, that the presented stimulus is stimulus i, denoted by $\kappa_i(\Phi)$, for each stimulus i. In a two-category situation, this confidence measure is given by

$$\kappa_i(\Phi) = \left| \frac{b_A \sum_{j \in C_A} \zeta_j s_{ij}(\Phi) + \gamma}{\left[b_A \sum_{j \in C_A} \zeta_j s_{ij}(\Phi) + \gamma\right] + \left[(1 - b_A) \sum_{k \in C_B} \zeta_k s_{ij}(\Phi) + \gamma\right]} - 0.5 \right| \tag{3.16}$$

where b_A is a category A response bias parameter, $(0 \leq b_A \leq 1)$, $k \in C_A$ refers to all stored exemplars which have category label A, ζ_j is a parameter representing the *strength* of exemplar j, (which is assumed to be related to the frequency of presentation of stimulus j), and γ is a noise parameter $(0 \leq \gamma)$ representing the absolute level of confidence (the greater the value of γ, the lower the absolute level of confidence). The form of this confidence function is very similar to Luce's choice rule (Equation 2.6) in that it is largely determined by the ratio of summed similarities of the stimuli to the exemplars of both categories. If the summed similarity of stimulus i to the exemplars of one of the categories is high relative to its summed similarity to the exemplars of the other category, the confidence that the presented stimulus is stimulus i will be also be high. Conversely, if the similarity ratio is small, the measure of confidence is also small.

The EGCM assumes that dimension processing continues until the measure of confidence is sufficiently high for a category decision to be made and a response initiated. The probability that feature sampling stops given a particular set of included dimensions, $P_{stop}(\Phi)$, is given by the power function

$$P_{stop}(\Phi) = \begin{cases} \left(\frac{\kappa_i(\Phi)}{\Psi}\right)^\theta & \text{if } \Psi > \kappa_i(\Phi) \\ 1 & \text{if } \Psi \leq \kappa_i(\Phi)\text{, or if all dimensions are included} \end{cases} \tag{3.17}$$

where Ψ is a parameter $(0 < \Psi)$ representing the confidence criterion which must be exceeded before a response is initiated and θ is a parameter which determines the shape of the function relating confidence to stopping probability, (increases in the value of θ necessitate larger degrees of confidence to achieve a particular stopping probability).

A category decision is made and a response initiated as soon as dimension processing stops. The probability of a category A response, (in a two-category situation), is given by

$$P_A(\Phi) = \frac{b_A \sum_{j \in C_A} \zeta_j s_{ij}(\Phi) + \gamma}{\left[b_A \sum_{j \in C_A} \zeta_j s_{ij}(\Phi) + \gamma\right] + \left[(1 - b_A) \sum_{k \in C_B} \zeta_k s_{ij}(\Phi) + \gamma\right]}. \tag{3.18}$$

Like the confidence function (Equation 3.16), this function is also based on Luce's choice rule (Equation 2.6) in that response probabilities are a function of the ratio of summed similarities of stimuli to the exemplars of the alternative categories.

Stimuli with integral dimensions

Because stimuli with integral dimensions are perceived as unitary wholes rather than objects composed of individual dimensions (Garner, 1974), the mechanism for stimulus sampling applicable to stimuli with separable dimensions described above is inappropriate. Therefore, instead of a discrete dimension inclusion mechanism, the EGCM proposes that the sampling of features of integral dimension stimuli is a process in which information about features is accumulated continuously over time. This has the same effect in terms of the time dependent availability of stimulus information and the computation of similarity because both processes result in the gradual accumulation of stimulus information over time and a corresponding increase in distinguishability between stimuli and exemplars in memory.

The set of equations governing this continuous process is essentially the same as that proposed for the discrete process summarised above. The main difference between the two is that the relationships described by the continuous dimension equations are all a function of processing time, t, rather than the set of included dimensions, Φ. It is assumed that the accumulation of information about a dimension x, ι_x, is an exponentially increasing function of time

$$\iota_x(t) = 1 - \exp(-q_x t) \tag{3.19}$$

where q_x is an accumulation rate parameter for dimension x corresponding to the q_x parameter in Equation 3.14. Similarity between stimuli and exemplars in memory also becomes a function of time t

$$s_{ij}(t) = \exp\left[-c\left(\sum_{p=1}^{P} \iota_p(t)\, u_p \left|x_{ip} - x_{jp}\right|^r\right)^{\frac{q}{r}}\right] \tag{3.20}$$

as does the equation governing stimulus confidence

$$\kappa_i(t) = \left|\frac{b_A \sum_{j\in C_A} \zeta_j s_{ij}(t) + \gamma}{\left[b_A \sum_{j\in C_A} \zeta_j s_{ij}(t) + \gamma\right] + \left[(1-b_A)\sum_{k\in C_B} \zeta_k s_{ij}(t) + \gamma\right]} - 0.5\right|. \tag{3.21}$$

The stopping probability function is slightly modified to become

$$P_{stop}(t) = \begin{cases} \left(\frac{\kappa_i(t)}{\Psi}\right)^{\theta} & \text{if } \Psi > \kappa_i(t) \\ 1 & \text{if } \Psi \leq \kappa_i(t), \text{ or if } cr < t \end{cases} \tag{3.22}$$

where cr is a parameter representing a self-imposed deadline. This parameter is required to place a maximum bound on the amount of time allowed for perceptual processing and so prevent the possibility of processing continuing indefinitely.

Application of the EGCM to RT data from experiments involves a two stage computation. Firstly, for each stimulus, the probability of every possible permutation of the course of feature sampling is computed. Then, the expected time required for each permutation is calculated and used in the computation of the predicted RT. Given a stimulus with two dimensions, for example, the possible courses of perceptual processing are firstly, that processing stops immediately after either one of the dimensions is sampled and secondly, that both dimensions are sampled one after another before processing terminates. The probability of each of these courses depends on the inclusion rates of the dimensions which are exponentially distributed random variables. As described in the discussion of the properties of exponential distributions above, the probability that an exponentially distributed process p_i with rate λ_i, is the first to finish from a set of n parallel exponentially distributed processes, $p_1, \ldots, p_n$, each with rate λ_i, is given by $\frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$. Therefore, the probability that a dimension x is included first from a set of dimensions, y, can be defined as

$$P(x) = \frac{q_i}{\sum_y q_y} \tag{3.23}$$

The probability that perceptual processing stops after the inclusion of a dimension is given by Equation 32. If processing continues, subsequent dimension inclusion probabilities are computed for the remaining dimensions. The resulting probabilities can then be multiplied to create probabilities for each possible sequence of inclusions.

The expected time to execute each possible sampling sequence is also a function of dimension inclusion rates. Referring again to the earlier discussion of the properties of exponential distributions, the expected minimum finishing time of a set of n parallel exponentially distributed processes, $p_1, \ldots, p_n$, each with rate λ_i, $E(T_n)$, is given by $\frac{1}{\sum_{i=1}^{n} \lambda_i}$. Therefore, the expected time of inclusion, τ, for the first dimension in a set of dimensions (each with inclusion rate q_i), Φ, can be defined as

$$E(\tau|\Phi) = \frac{1}{\sum_{i \in \Phi} q_i} \tag{3.24}$$

The expected inclusion time for each subsequent dimension is computed using Equation 3.24 with the set of remaining dimensions. The total expected time for a sequence of dimension inclusions is equal to the sum of the expected times of the individual dimensions. The expected processing time for a stimulus is calculated by multiplying the probability of each possible sequence of inclusions with the total expected time for that sequence and summing the results. A residual time, t_{res}, which represents the brief period of time immediately after stimulus onset in which dimensions cannot be processed and the time involved in the category decision and response execution stages, is added to this final value to produce the total predicted RT for the stimulus.

Application of the EGCM to category response proportion data involves the same computations as application to RT data—predicted choice proportions are calculated by multiplying the probability of each inclusion sequence by the expected choice for that sequence and summing the results

The EGCM has been able to produce accurate quantitative fits to data from speeded categorisation experiments using stimuli with separable and integral dimensions and compares favourably with the EBRW in the modelling of data from experiments using response deadlines (Lamberts, 1997, submitted). Because the EGCM has a parameter representing the strength of exemplars, it is also able to account for frequency effects on categorisation RT, including those found by Nosofsky & Palmeri (1997, Experiment 2). This aspect of the model will be discussed in detail in Chapter 8.

EBRW and EGCM compared

Although the EBRW and EGCM share a number of assumptions about the representation of information and the computation of similarity between stimuli and stored exemplars, the two models differ widely in their assumptions about the actual processes involved in categorisation and how they relate to response times. The EBRW essentially regards RT as a function of the category structure and the similarity relationships between stored exemplars. The random walk process incorporated in the EBRW is driven by a race between exemplars to be retrieved from memory and the probability of an exemplar winning this race is determined by the similarity between the exemplar and the stimulus. The free parameters in the EBRW related to categorisation allow the model to adapt the distribution of attention across the stimulus dimensions and to modify overall discriminability within the stimulus space. Although the free parameters relating to the prediction of RTs provide an additional degree of flexibility, the essential driving force behind the RT predictions of the model is the category structure.

In contrast, the EGCM assumes that in addition to the category structure, other factors play an important role in determining RT, most notably the perceptual features of the stimulus (i.e. the salience of the stimulus dimensions). Lamberts (1995) showed that a model which distinguished between inclusion rates (related to dimensional salience) and utility values (related to category structure) was able to give a better account of categorisation performance under the pressure of imposed response deadlines than a model which only employed the latter. Given the further assumption that response decisions are initiated when sufficient stimulus information has been accumulated in non-deadline conditions, the additional parameters in the EGCM give the model a degree of flexibility not found in the EBRW. One of the aims of the research reported in the following chapters, therefore, is to determine whether the additional assumptions held by the EGCM allow the model to fit the data more closely and consequently to shed light on the question as to whether these assumptions are justified.

Summary

In this chapter, a general introduction to the study of response times in psychology was given and the three main models of categorisation RT were described in detail. In the next chapter, I discuss more fully approaches to modelling RTs from within a connectionist framework and present a new model of categorisation RTs which extends the ALCOVE model of categorisation with the incorporation of a mechanism for the accumulation of information over time.

Chapter 4

A Connectionist Model of Categorisation Response Times

Introduction

In this chapter, the various connectionist approaches to modelling reaction times are introduced and the basic principles of cascade models are described. Then a new model of categorisation reaction times is proposed which augments Kruschke's (1992) ALCOVE model of categorisation with a cascade activation function. The basic properties of the new model are then described and illustrated by applying the model to several category structures of varying difficulty.

Connectionist approaches to modelling RT

The last few years have seen an increasing number of attempts to model the time course of cognitive processes within the connectionist framework. An important feature of some connectionist models in this regard is that they contain explicit mechanisms for processing information over time. Given the assumption that reaction time is related to the dynamic behaviour of a network, predictions can be made based on the number of cycles taken by a model to reach a stable state given a particular input. Connectionist systems of this sort have been used as the basis for theories of cognitive processing in general and to model the time course of various cognitive functions, for example McClelland's (1979) *Cascade* and (1993) *GRAIN* models, Anderson's *Brain-state-in-a-box* model (Anderson, Silverstein, Ritz & Jones, 1977; Anderson, 1991), and McClelland & Rumelhart's *IAC* model of letter perception, (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), (see also Bullinaria, 1995, and Ratcliff & van Zandt, in press, for recent discussions of connectionist models of reaction times).

In terms of modelling RTs, artificial neural networks can be divided into two classes—those which process information over time due to their inherent structural properties and those which, unless augmented with additional time dependent mechanisms, process information instantaneously. Examples of the former class are the *Hopfield* network (Hopfield,

1982) and *Boltzmann Machine* (Hinton & Sejnowski, 1983a, 1983b), the *Brain-state-in-a-box* model (Anderson, Silverstein, Ritz & Jones, 1977) and the various models based on the principles of *Interactive Activation* (e.g. McClelland, 1981; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; McClelland & Elman, 1986). Information processing in this class of networks (often called "attractor networks") is due to the properties of non-linear dynamic systems and is commonly characterised as resulting from the network's trajectory towards a stable point attractor or its downward movement in an energy landscape. The time course of processing in these networks is typically seen as a function of this trajectory. The most prominent examples of the latter class, (commonly known as "feedforward networks"), are the single-layer (Rosenblatt, 1958) and multilayer perceptron (Rumelhart, Hinton & Williams, 1986), and the *Self Organising Map* (SOM; Kohonen, 1982). In the context of this discussion, the important distinction between these networks and those of the first class is that once feedforward networks have been trained, information processing is generally restricted to the forward propagation of a signal from input to output layers. By adding a time dependent factor to the equations governing the activation of processing units, however, these networks can be transformed into *cascaded* networks in which activation is accumulated over time. As this is the method employed here, an explanation of the underlying principles of the approach is necessary.

Cascaded networks

In the original formulation of the cascade model, McClelland (1979) set out a number of basic principles for an approach to modelling reaction times which challenged the dominant Donderian and Sternbergian assumptions about the discrete nature of information processing stages discussed in Chapter 3. Like the earlier approach, the cascade model assumes that processing consists of several stages or levels. In contrast to the Donderian/Sternbergian approach however, the cascade model assumes that processing is carried out at each level simultaneously and that output from each level is continuously available to the layer immediately following it. In a multilayer cascade model, therefore, upon presentation of a stimulus pattern to the input layer, activation *cascades* across successive layers of the network in a gradual and continuous process until a global stable state is reached. The activity of a processing unit in any layer at a particular time t is a reflection of the weighted input to the unit at time t. The set of output layer unit activations is taken to be the response of the network and the number of cycles taken for the output units to stabilise or reach a threshold value taken to be the time variable.

If the network is linear, as in the case of McClelland's cascade model, processing units are simple linear integrators, computing their output as the weighted sum of the activations of units in the preceding layer to which they are connected. The net input to a unit i at level r at time t is given by

$$net_{i(r)}\,(t) = \sum_j a_{j(r-1)}\,(t)\,w_{ij} \tag{4.1}$$

where $a_{j(r-1)}(t)$ is the activation of unit j at level $r-1$ at time t. Note that, as described in Chapter 2, the output units of the ALCOVE model are of this type. Accumulation of unit activation over time is governed by the cascade equation (McClelland, 1979; Cohen, Dunbar & McClelland, 1990; Bullinaria, 1995). In this equation, the activation of unit i in layer r at time t, denoted $a_{i(r)}(t)$, is a function of the current net input to i at time t, the previous net input to i at time $t-1$ and a *cascade rate* parameter, τ_r, $(0 \leq \tau_r \leq 1)$, which determines the speed of activation change for all units in layer r.

$$a_{i(r)}\left(t\right) = \tau_r net_{j(r)}\left(t\right) + \left(1 - \tau_r\right) net_{i(r)}\left(t-1\right). \tag{4.2}$$

If τ_r is large, the activations of units in layer r change rapidly whereas if τ_r is small, change in unit activation is more gradual. Two consequences of Equations 4.1 and 4.2 are firstly, that if the input to a layer of units is fixed, each unit in the layer will reach an asymptotic activation equal to its net input, and secondly, that units with a relatively large absolute net input will reach asymptote more rapidly. This variation in rate is due to the fact that the size of activation change for an individual processing unit at time t increases with the size of the absolute difference between the net input to it at time t and $t-1$ (McClelland, 1979; Cohen, Dunbar & McClelland, 1990; Bullinaria, 1995). This situation is depicted in Figure 4.1 which shows plots of activation accumulation over time for three linear units with net inputs of 0.2, 0.5 and 0.8 respectively. The value of τ for each unit is set to 0.29.

McClelland's (1979) analysis was confined to networks composed of linear units and the computational limitations of this type of network are well known (Rumelhart, Hinton, & McClelland, 1986). Nonlinearities can be incorporated into the cascade equation, however, by simply making unit activation a non-linear (e.g. logistic) function of net input.

There have been a number of recent attempts to model the time course of cognitive functions using cascaded networks. For example, Cohen *et al.* (1990) have successfully constructed a model of the Stroop task which accounts for both the time course of processing and the effects of learning while Bullinaria (1995) has employed a cascaded network to analyse the process of reading aloud (i.e. text to phoneme translation). Bullinaria's model is able to reproduce several important features of phoneme and word naming latencies and some of the effects of word priming on RT.

The ALCOVE(RT) model

The various strands of this discussion serve to highlight a number of important issues relating to the modelling of categorisation RTs. Firstly, in Chapters 2 and 3, it was shown how previously existing categorisation models have been applied successfully to the analysis of response times. In Chapter 3, the issue of whether cognitive processes are to be regarded as composed of discrete stages was also raised and an alternative connectionist conception of processing has been outlined in this chapter. The new model of categorisation RTs proposed here represents an attempt to address these issues by augmenting the ALCOVE model of categorisation with time dependent exemplar and output unit activation functions to produce a cascaded network model of categorisation RTs.

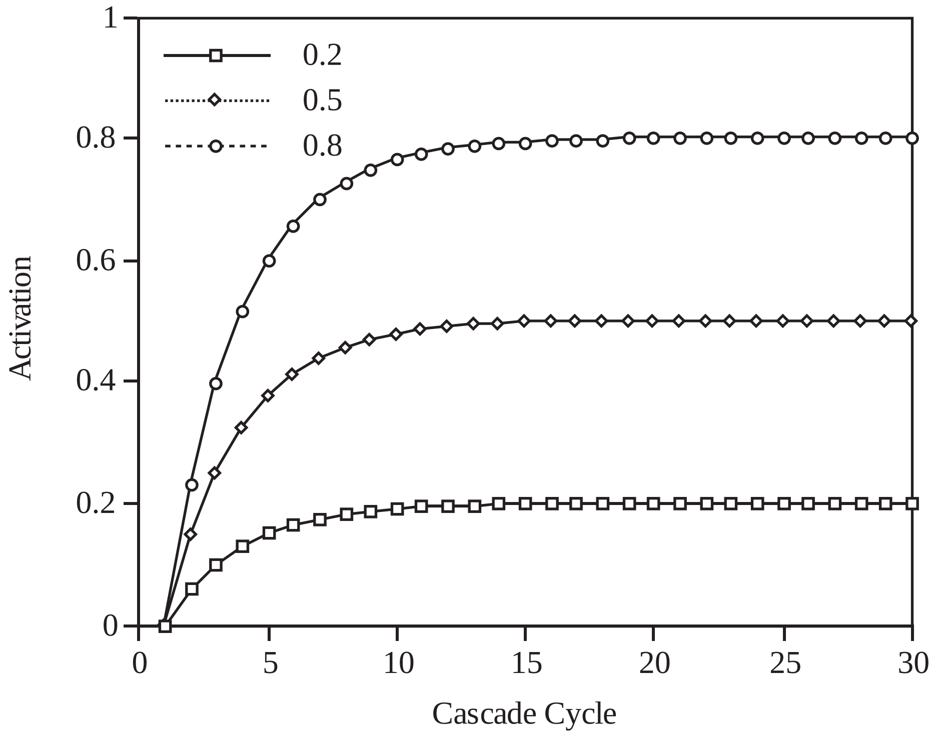

Figure 4.1: Activation accumulation profiles of three linear units with net inputs of 0.2, 0.5 and 0.8. The value of τ for each unit = 0.29.

Adding a time dependent activation function to the exemplar units can be interpreted as creating a model in which, upon presentation of a stimulus, stored exemplars are activated to the extent and at a rate which is dependent on the similarity of the exemplar to the stimulus. The main assumptions of the ALCOVE model are retained by the new RT model, which henceforth will be referred to as *ALCOVE(RT)*. The only modifications made are to the activation functions for the hidden and output units during the testing or *transfer* stage, (unit activations remain instantaneous during training). By inserting the exemplar activation function, (Equation 2.9), into the cascade equation, (Equation 4.2), the activation of exemplar i at time t, $a_i^{hid}(t)$, becomes

$$a_i^{hid}(t) = \tau_{hid} \exp\left[-c\left(\sum_j \alpha_j \left|h_{ij} - \alpha_j^{in}\right|^r\right)^{\frac{q}{r}}\right] + (1 - \tau_{hid})\, a_i^{hid}(t-1) \tag{4.3}$$

where τ_{hid} is the hidden layer cascade rate parameter.

By similarly inserting the output unit activation function (Equation 2.10) into the cascade equation, the activation of output unit i at time t, $a_i^{out}(t)$, becomes

$$a_i^{out}(t) = \tau_{out} \sum_{j}^{hid} w_{ij} a_j^{hid} + (1 - \tau_{out})\, a_i^{out}(t-1) \tag{4.4}$$

where τ_{out} is the corresponding cascade rate parameter for the output layer. The activation of a category output unit at time t can be interpreted as reflecting the amount of evidence for that category at time t.

Two primary factors in the ALCOVE(RT) model have a combinatorial effect on the rate of category unit activation (and hence predicted RT) for a given stimulus. The first is the relative similarity between a stimulus and the exemplars of both categories. The second is the strength of the associative links between the exemplars and the alternative categories. If a stimulus is more similar to exemplars from one category than from another, the activation of these exemplars will accumulate more rapidly with the consequence that the activation of the category unit with which they are associated will also quickly increase. If the exemplars are strongly positively associated with the category unit and strongly negatively associated with the alternative unit, their rapid accumulation of activation will push the category unit activations strongly in opposite directions early in the cascade process. If, however, there are only slight differences in association between the exemplars and the alternative categories, the rate of activation change in the category units will be more modest given the same net exemplar input.

The time course of information processing in the ALCOVE(RT) model can be demonstrated using the example stimulus set used in Chapter 2. The stimuli vary on three binary dimensions, the values of which are represented by 1 and 0. The dimensions are *size* (large = 1, small = 0), *shape* (square = 1, circle = 0) and *colour* (black = 1, white = 0). The dimension values and category structure for this stimulus set are displayed in Table 4.1.

For simplicity, the category structure was constructed along the size dimension so that an optimal decision boundary would classify large stimuli into category A and small stimuli into category B. Because optimal categorisation of the structure shown in Table 4.1 requires differentiation simply in terms of size, training the network should result in the situation depicted in Figure 2.1 in which the amount of attention (represented by α in ALCOVE) paid to the size dimension will be large relative to the colour and shape dimensions.

Table 4.1: Structure of Stimuli in Simulation 1

Category Structure	Stimulus Number	Dimension		
		Size	Shape	Colour
Category A	1	1	1	1
	2	1	1	0
	3	1	0	1
	4	1	0	0
Category B	5	0	1	1
	6	0	1	0
	7	0	0	1
	8	0	0	0

The ALCOVE(RT) model was constructed with three input units (one representing each stimulus dimension), eight exemplar units (one for each training stimulus), and two output units representing the alternative categories. Because the purpose of this demonstration is simply to illustrate cascade processes and not to model actual response proportions or reaction times obtained in experiments, the various model parameters were not optimised but were given values that were deemed appropriate or which had been previously shown by Kruschke (1992) to produce adequate behaviour in similar situations. Specifically, the association learning rate parameter (λ_w) was set to 0.03 and attention learning rate parameter (λ_α) set to 0.0033, the cascade rate parameters for the hidden (τ_{hid}) and output (τ_{out}) layers were set to 0.2, and the exemplar unit specificity parameter (c) and the response mapping parameter (ϕ) set to 2.5 and 2.0 respectively. The dimensional attention strengths were all set to 1/3 to reflect an even distribution of attention before training and the association weights between exemplar and category units were initialised to 0 (as it is assumed that there is no association between exemplars and categories prior to training). The network was trained for fifty *epochs* as described in Chapter 2 using a batch updating procedure. In *batch* updating, changes in connection weights are carried out after each epoch (i.e. when all stimuli in the training set have been presented once to the network), which results in the removal of any effects of stimulus sequence.

After fifty epochs, the network achieved optimal classification performance (i.e. the probability of correct classification = 1.0). As expected, the network had adjusted the attention strengths on the stimulus dimensions to stretch the stimulus space along the size dimension and shrink it on the colour and shape dimensions. The attention strength for the size dimension after training was 0.57 whereas those for the colour and shape dimensions were 0.17. To illustrate the cascaded activation process in the testing stage, the network's response to one stimulus (stimulus 1) will be examined. Table 4.2 shows exemplar and category unit activation values in response to stimulus 1 after training.

Table 4.2: Exemplar and Category Unit Activation Values in Response to Stimulus 1 After 20 Epochs of Training

	Unit	Activation
Exemplar Units	1	1
	2	0.658
	3	0.658
	4	0.433
	5	0.243
	6	0.160
	7	0.160
	8	0.105
Category Units	A	1.0
	B	−1.0

As described in Chapter 2, exemplar unit activations are a measure of the similarity between the exemplar and the stimulus. The effect of the differences in attention strength can be observed in the activations of exemplar units which share the same number of features with stimulus 1. For example, in the case of exemplars 2, 3 and 5, which all have two features in common with stimulus 1, the increase in attention to the size dimension has the effect of increasing the similarity of exemplars 2 and 3 to stimulus 1 because of their shared value on this dimension. The activations of the output units indicate that the amount of evidence for category A is high while that for category is correspondingly low (output units were trained to respond "at least 1" to stimuli in their category and "at least -1" to stimuli not in their category).

In the testing stage, the exemplar and category unit activation values were reinitialised to 0 before each stimulus was presented to the network. Following presentation of each stimulus, the network was allowed to cascade for fifty cycles. The number of cycles for a network to stabilise is subject to manipulation, as the rate of information accumulation to asymptote is determined in large part by the network layer cascade rates. It was found that the network in this simulation had stabilised after about twenty cycles. The accumulation profiles of the exemplar units in response to stimulus 1 are displayed in Figure 4.2.

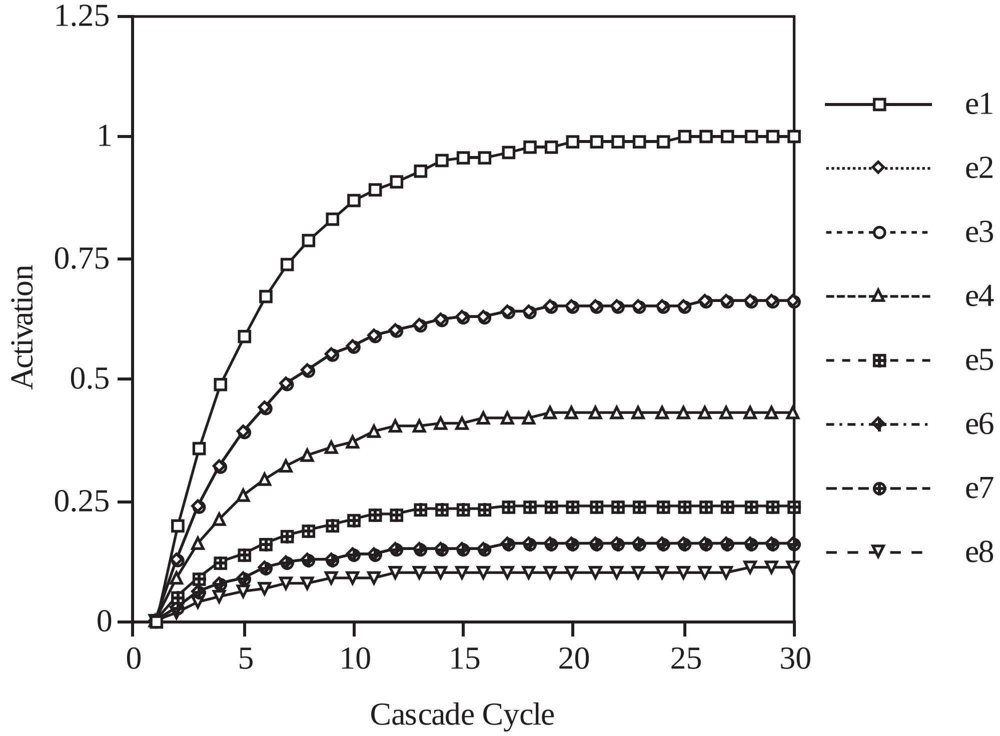

Figure 4.2: Activation accumulation profile of ALCOVE(RT) exemplar units in response to stimulus 1, Simulation 1

The graph clearly shows the relationship between stimulus-exemplar similarity and rate of exemplar unit activation accumulation. Note that, because the activations of exemplar pairs 2 and 3 and 6 and 7 are identical, their accumulation profiles also overlap, resulting in the obscuration of two plots. The activation accumulation of the output units is displayed in Figure 4.3. The highly symmetrical form of the category structure and the use of 1 and -1 for teacher values during training result in a correspondingly symmetric set of network responses to the stimuli. After training, all of the exemplar-category association weights have the final value of 0.481 (negative associations = −0.481) and the set of exemplar and output activations are the same for each stimulus (the only difference being the distribution of activations across units). One of the results of the uniformity of the association weights is that the category unit activations are pushed equally in opposite directions by the combined activations of the exemplar units.

The probability of responding to each of the categories over the course of the 30 cycle cascade is plotted in Figure 4.4. Again, the curves reflect the accumulation of evidence for a category A response over time as the category A exemplar activations rose rapidly to asymptote.

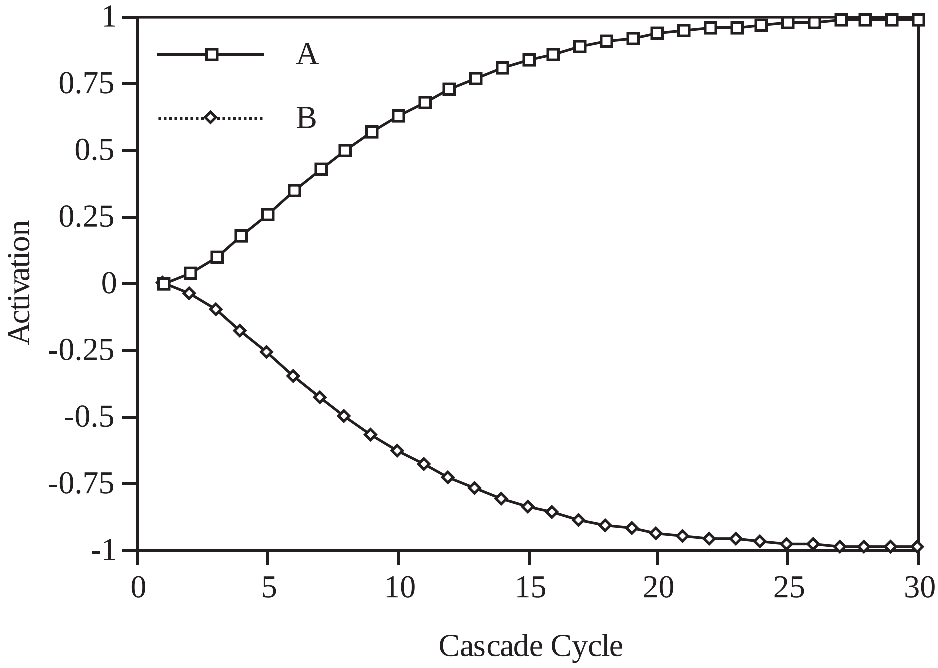

Figure 4.3: Activation accumulation profile of ALCOVE(RT) category units in response to stimulus 1, Simulation 1

The category structure shown in Table 4.1 corresponds to the Type 1 category structure identified by Shepard, Hovland and Jenkins (1961) in their classic study of category learning. Shepard *et al.* found that there are six distinct category structures that can be formed from eight stimuli varying on three binary dimensions and that, because information from only one dimension is required for correct classification, the Type 1 structure above is the easiest to learn of the six. The structures specified by Shepard *et al.* were also used by Kruschke (1992) to explicitly test ALCOVE's selective attention learning mechanism and to show how this mechanism allows ALCOVE to account for the relative difficulty of learning the different category types. The six category structure types can be represented using the scheme introduced in Figure 2.1 in Chapter 2. The resulting structures are presented in Figure 4.5.

For ease of reference, the three dimensions of the cube are numbered according to the trident at the lower right of the figure while the number of each stimulus is indicated at the lower left. The different category structures are represented by the open and filled circles at the corners of each cube. To take the previously encountered Type 1 structure as an example, the diagram representing the structure is to be interpreted as indicating that stimuli 1, 2 ,3 and 4 (the open circles) belong to one category while stimuli 5, 6, 7, and 8 (the filled circles) belong to the other and that only one dimension (dimension 1) distinguishes the categories.

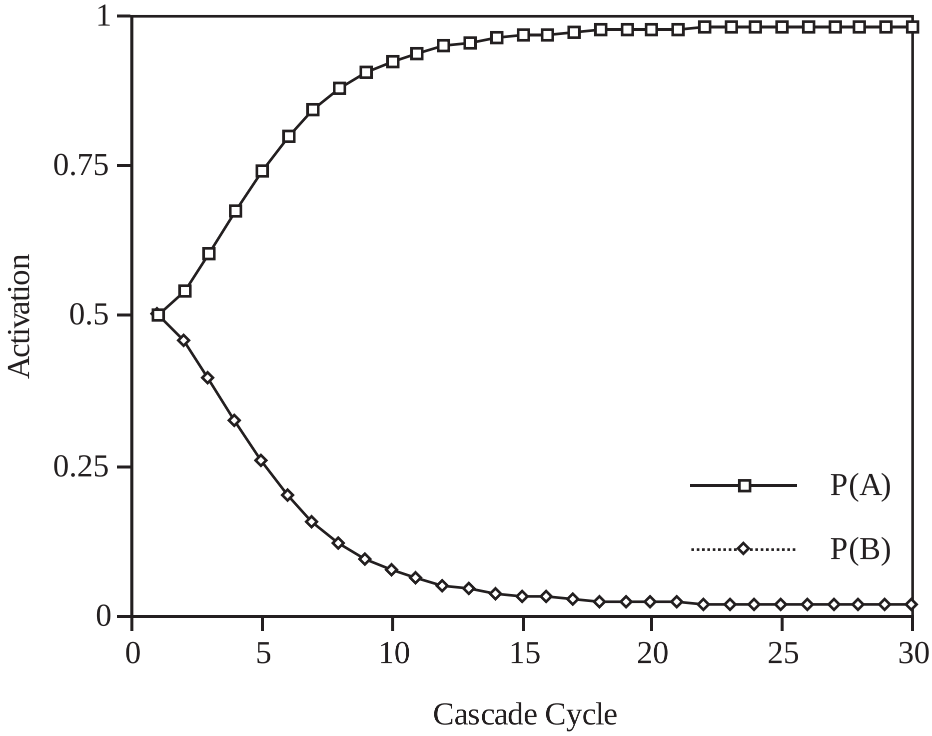

Figure 4.4: Probability of category response as a function of cascade cycle for ALCOVE(RT) model in response to stimulus 1, Simulation 1

Shepard *et al.* rated the category types in terms of the number of dimensions required to classify the stimuli correctly and, on the assumption that additional cognitive effort or capacity is required for each extra dimension to be considered, reasoned that the Type 1 structure should be easiest to learn, followed by Types 2, 3, 4, 5 and 6. To correctly classify the Type 2 structure, two dimensions must be taken into account (dimensions 1 and 2) whereas for Types 3, 4, 5, and 6, correct classification requires the consideration of all three dimensions. Types 3, 4 and 5 differ from Type 6, however, in that, although information about all three dimensions is required for correct classification, all dimensions are not equally informative (i.e. most stimuli in the category type can be correctly classified on consideration of only two dimensions). According to this analysis, Type 6 should be the most difficult to learn because equal attention must be paid to all three dimensions in order to classify the stimuli correctly.

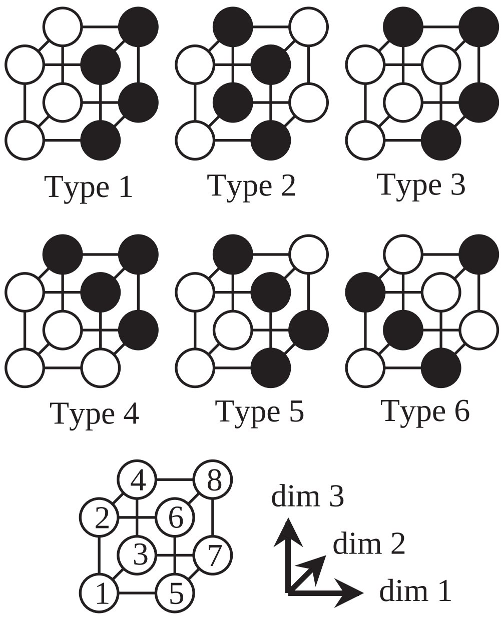

Figure 4.5: The six category types for stimuli varying on three binary dimensions identified by Shepard, Hovland and Jenkins (1961)

Shepard *et al.* (1961) carried out several category learning and memorisation experiments using the six structures and discovered that difficulty of learning (as measured by the number of errors made during learning) did indeed follow the order $1 < 2 < (3, 4, 5) < 6$ as predicted by their hypothesis. Note that Types 3, 4 and 5 are bracketed together because their levels of difficulty were found to be very close. Shepard *et al.* also found that other measures of difficulty, such as speed and accuracy of category recall followed the same ordering.

To explain their findings, Shepard *et al.* proposed a stimulus generalisation hypothesis which stated that ease of category learning is related to the set of similarity relationships between stimuli and in particular, the ratio of within- and between-category similarities. If a category structure is such that stimuli in the same category are highly similar to each other and highly dissimilar to stimuli in alternative categories, it should be relatively easy to learn. In contrast, the learning of category structures in which categories are composed of stimuli which are similar to members of alternative categories and which are relatively dissimilar to each other should be more difficult.

To observe how the cascade process is affected by category difficulty and to determine whether the order of difficulty is reflected in differences in the output activation accumulation rate for the category types, the ALCOVE(RT) model was applied to the remaining five category types. For each type, the network was constructed and initialised as previously described and was then trained on the category structure for fifty epochs as before. After training, the stimuli were again presented to the network which was allowed to cascade for fifty cycles. Table 4.3 shows the probability of a correct response for each of the category types after fifty epochs of training. The value in the table for each category type is the probability of a correct response averaged over the eight training stimuli. The order of results is in line with the ordering found by Shepard *et al.*

Table 4.3: Probability of Correct Category Response by ALCOVE(RT) for Each Category Structure Type After Fifty Epochs of Training. For Each Structure Type, P(Correct) is the Probability of A Correct Response Averaged Over all Stimuli in the Training Set.

Category Type	P(Correct)
1	0.98
2	0.97
3	0.89
4	0.91
5	0.82
6	0.60

The average information accumulation values of the correct category unit for each of the six category types are displayed in Figure 4.6. To form the plot for a category type, the cascade values of the correct category unit for each stimulus were averaged so that each plot can be interpreted as the average accumulation profile of a correct network response to stimuli within a category type. It can be seen in Figure 4.6 that the evidence for a correct category response accumulates on average more rapidly for the easier category structures than for the more difficult ones.

To examine what is happening in the ALCOVE(RT) model to cause these differences in cascade and to show the stimulus generalisation hypothesis is embodied in ALCOVE, one can contrast the network's responses to stimuli of two of the category types, for example, Types 1 and 6. Both structures are symmetrical but differ in the number of dimensions required for correct classification. In the Type 1 structure, because of the increase in the attention paid to dimension 1 during training, the similarity between stimuli within the same category is very high whereas that between stimuli in different categories is correspondingly low, resulting in relatively large association weight values (0.48 and -0.48). When a stimulus is presented to the network, therefore, the high similarity values of stimuli in the same category (average activation value = 0.687) result in rapid accumulation of their activation, a rapid increase in the activation of the category unity to which they are (strongly) positively associated and a correspondingly rapid decrease in activation of the category unit to which they are (equally strongly) negatively associated.

This effect is compounded by the very low activation accumulation of exemplar units from the alternative category (average activation value = 0.167). The activation accumulation of their associated category unit remains modest because of the low positive net input it receives in comparison to the negative net input from exemplars in the alternative category.

In contrast, when the network is trained on the Type 6 category structure, attention is evenly distributed (in the ALCOVE(RT) model, the final value of α for each dimension after training = 0.465) because all three dimensions are equally diagnostic. As a result,

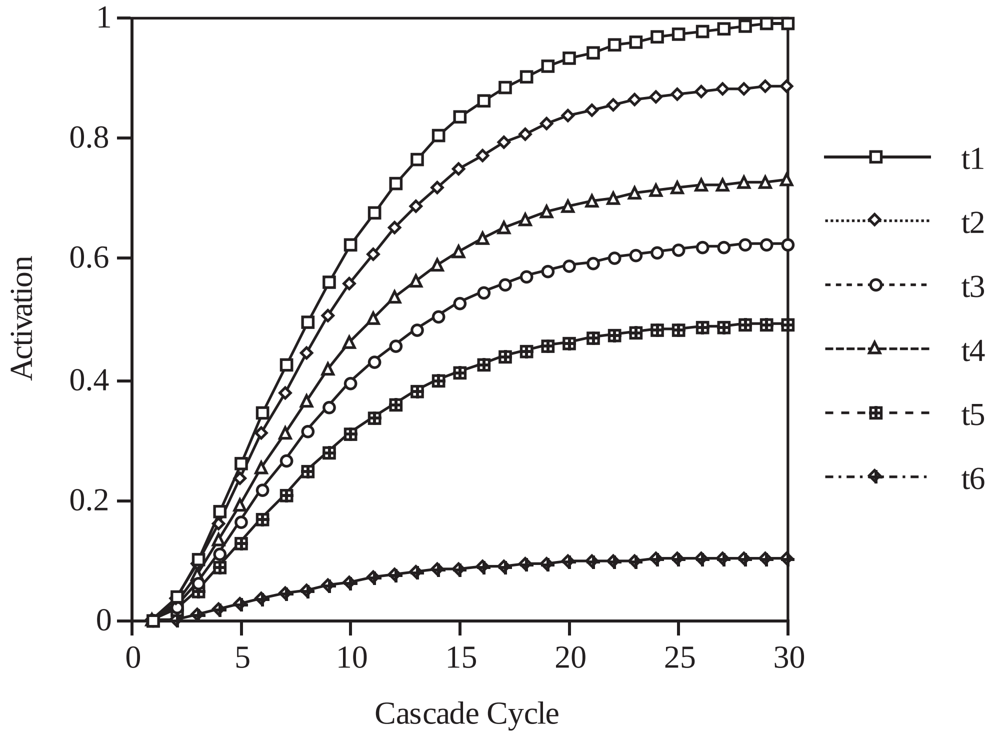

Figure 4.6: Activation accumulation profile of correct category unit averaged over all stimuli for each category type

when a stimulus is presented to the trained network, the difference between similarity between stimuli in the same category (average activation value = 0.434) and that between stimuli in the alternative category (average activation value = 0.138) is relatively small compared with the Type 1 structure. The association weights between exemplar and category units are also not as big (0.32 and −0.32). As a consequence, the accumulation of activation of the correct category unit is less rapid.

ALCOVE(RT), EBRW and EGCM compared

In terms of the assumptions concerning the processes involved in categorisation and their relationship to response times, the ALCOVE(RT) model differs in many respects from both the EGCM and EBRW. Intuitively, the ALCOVE(RT) model does have an underlying similarity to the EBRW model in that both relate response time to the ratio of summed exemplar activations of the alternative categories. Whereas the EBRW sees this ratio as affecting a stochastic process feeding into a random walk, however, the ALCOVE(RT) model assumes that the process is one of simple information accumulation. ALCOVE(RT) has less in common with the EGCM as, like the EBRW, it does not assume that perceptual characteristics of the stimulus (i.e. dimensional salience) have any effect on categorisation or RT. The main difference between the ALCOVE(RT) model and the EGCM and EBRW is the process by which dimensional attention weights are modified and the category structure is represented. In the EGCM and EBRW, attention weights are free parameters for which optimal values are found using a search algorithm during the model fitting process. Similarly, the category structure is represented in the EGCM and EBRW simply as a set of labels attached to the stored exemplars. In contrast, because ALCOVE(RT) is a model of category *learning*, the modification of attention weights is carried out by a process of gradient descent on error using the generalised delta rule. This error-driven learning process is also used to adapt the association weights between exemplar and category units in the network by which the category structure is represented. Therefore the constraints placed upon the categorisation process embodied in the ALCOVE(RT) model are much greater than those upon the EGCM and EBRW because ALCOVE(RT) is required to rely exclusively upon the abstract category structure to which it is exposed during the category learning process.

This characteristic of ALCOVE(RT) is important because it allows alternative assumptions concerning the information required for categorisation and the time course of categorisation to be tested and compared to those held by the EGCM and EBRW. In this respect, therefore, there are three distinct assumptions of increasing complexity embodied in the three models under examination. The first, held by the ALCOVE(RT) model, is that categorisation relies solely on the abstract structure of the categories being learned and that categorisation RT is a function of the gradual accumulation of activation based on the ratio of summed exemplar activations of alternative categories. The EBRW also assumes that categorisation relies on the category structure but, because the dimension attention weights are not constrained by this structure as are those of ALCOVE(RT), the EBRW has an additional degree of flexibility over the learning model. Finally, in addition to the category structure, the EGCM also assumes that RT is related to the creation of a perceptual representation of a stimulus based on the sampling of stimulus features over time, and this represents yet another level of complexity.

One of the primary goals of this research, therefore, is to investigate the ability of the three models to account for experimental data in order to determine whether categorisation RTs can be explained solely in terms of the abstract category structure, or whether (and, if so, which) additional assumptions are required to explain the data.

Summary

In this chapter, the connectionist approach to modelling reaction times was introduced and the principles of cascade models were outlined. Then a new model of categorisation reaction times, the ALCOVE(RT) model was proposed and the basic properties of the model were demonstrated. In Chapter five, the categorisation accuracy and RT predictions of the EBRW, EGCM and ALCOVE(RT) models will be tested on data from a speeded categorisation experiment.

Chapter 5

Experiment 1

Introduction

The primary purpose of Experiment 1 is to test the ability of the EGCM, EBRW and ALCOVE(RT) models to account for joint categorisation accuracy and RT data from a standard speeded category learning experiment. Applying the three models to experimental data permits a comparison of predictions and the possibility to discover limitations in any of the models or to observe any underlying similarities between models' patterns of prediction.

The data with which Nosofsky and Palmeri (1997) test the EBRW were generated by experiments in which participants were required to categorise integral-dimension stimuli varying on two dimensions, (e.g. colours varying in brightness and saturation and tones varying in pitch and loudness). One motivation for this focus on integral-dimension stimuli is Nosofsky and Palmeri's hypothesis that exemplar-based categorisation more often occurs when stimulus dimensions are not perceived or processed separately but are integrated to form unified representations. They claim that, in contrast to this, stimuli with separable dimensions seem to be perceived and processed in terms of their separate dimensions, a factor which may not be as beneficial to the efficient storage of unified exemplar representations. A second reason given is their belief that separable-dimension stimuli may not so easily be encoded in parallel as those composed of integral dimensions. One possible effect of this, it is claimed, is that the serial processing of separate dimensions may add unnecessary error to the response-time data. In contrast, the EGCM has been successfully tested using stimuli varying on more than two binary valued dimensions, (Lamberts, 1996). The question whether the EBRW can be as successfully applied to this type of stimulus remains open, however. A further purpose of Experiment 1, therefore, was to generate categorisation accuracy and reaction time data using such stimuli with which to test the predictions of EBRW model.

Another reason for carrying out this experiment is to determine the nature of the relationship between RT and response accuracy. A broad assumption held by all of the RT models previously discussed is that increased classification accuracy correlates with faster RTs. This experiment is designed to investigate this relationship by having participants learn a relatively simple category structure to a predetermined level of accuracy and then to categorise the same stimuli again in a speeded test stage. Data from the test stage may then be analysed to whether differences in response accuracy are accompanied by corresponding differences in RT.

Because the domain of application of the RT-distance model is restricted to stimuli with continuously varying dimensions and probabilistic category structures, it cannot be included in this test.

Method

Participants

Ten undergraduate and postgraduate psychology students from the University of Birmingham participated in the experiment. The undergraduates who took part were given credit towards the Psychology department's research participation scheme.

Apparatus and stimuli

The experiment was carried out on an Elonex PC-433 computer with a Vale EC 33 cm SVGA colour monitor using a display mode with 640 pixels horizontally and 480 pixels vertically. Participant's responses were registered by two microswitches connected to the computer's parallel port. The stimuli used were drawings of aeroplanes viewed from above which varied on four binary dimensions—shape of *nose* (round or pointed), shape of *wings* (straight or tapered), number of *engines* (two or four), and shape of *tail* (square or rounded). Two example stimuli showing the full range of dimension values are shown in Figure 5.1.

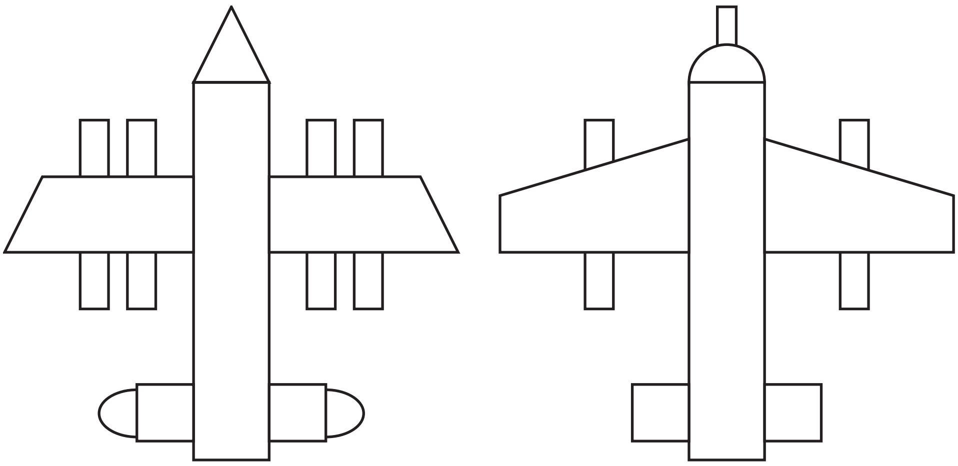

Figure 5.1: Sample stimuli in Experiment 1

Design and procedure

The category structure, which comprised a total of eight stimuli, is shown in Table 5.1. The structure is regular in that each stimulus differs on two dimensions from six of the other stimuli and on all dimensions from one stimulus in the alternative category. This regularity is also apparent in the arrangement of the stimuli as on each dimension, three of the four stimuli in a category have one value while one stimulus has the opposite value. In addition, no dimension is more predictive of a category than any other.

Table 5.1: Structure of Stimuli in Experiment 1

	Stimulus	Dimension			
Structure	Number	Nose	Wings	Engines	Tail
Category A	1	1	1	1	0
	2	1	1	0	1
	3	1	0	1	1
	4	0	1	1	1
Category B	5	0	0	0	1
	6	0	0	1	0
	7	0	1	0	0
	8	1	0	0	0

The experiment was a standard categorisation reaction time experiment which consisted of a training stage, in which participants were required to learn to classify stimuli into two categories, followed by a transfer stage, where the task was to classify the same stimuli again as quickly as possible without sacrificing accuracy. In both training and transfer stages, participant's category responses were recorded and in the transfer stage, the time of each response (in milliseconds) was also recorded. In the training stage, participants were presented with the stimuli in blocks, each block consisting of the complete set of training stimuli presented sequentially in random order. Training continued until two blocks in succession had been categorised correctly. On each training trial, a white fixation cross would appear at the centre of the blank computer screen for a period of 400 ms followed by a period of 100 ms where the screen was again blank. Then one stimulus chosen at random would appear at the centre of the screen. When one of the two response buttons was pressed, an auditory signal indicating the correctness of the response was given for a period of 500 ms and the screen would be cleared. If the category response was correct, a 600 Hz (high) tone was given whereas if it was incorrect, a 100 Hz (low) tone was given. An interval of 1500 ms separated each training trial. To eliminate any effect of response bias due to handedness, category labels were randomly assigned to left and right response buttons.

After a short break, participants underwent a transfer stage in which blocks of stimuli were presented again as during the training stage. Trials in the transfer phase were identical to training trials except that no auditory feedback was given. Participants were instructed to categorise the planes as before but this time to be as fast as they could while trying to remain as accurate as possible. Each participant categorised a total of fifty blocks in the transfer stage, being allowed to rest for a few minutes twice during the session, after the completion of 17 and 34 blocks.

Results

Training

The mean number of blocks required to achieve two consecutive correct blocks was 42.2 ($SD = 23.1$). A simple measure of the difficulty of learning the category structure is the mean error frequency for each stimulus over the course of training across participants. These frequencies are presented in Table 5.2.

Table 5.2: Mean Error Frequencies During Training, Experiment 1

Stimulus	Error Frequency
1 1110 (A)	11.2
2 1101 (A)	21.4
3 1011 (A)	12.3
4 0111 (A)	12.3
5 0001 (B)	8.50
6 0010 (B)	14.8
7 0100 (B)	13.3
8 1000 (B)	12.9

An analysis of variance (ANOVA) on the mean correct responses yielded a significant effect of stimulus $F(7, 63) = 2.49$, $p < .05$, $MSE = .062$. In the training stage, participants' errors were generally evenly distributed across the stimuli, with the exception of stimulus 2 which was on average misclassified more often and stimulus 5 which was in general categorised more accurately. The fact that stimuli 2 and 6 had the highest error rates suggests the possibility that participants were paying greatest attention to the *engines* dimension as these two stimuli differed from the others in their category on this dimension. Conversely, the low error rates of stimuli 1 and 5 suggests that relatively little attention was being paid to the *tail* dimension during the training stage.

According to the MDS analysis of similarity described in Chapter 2, if attention is focused upon a particular dimension, the psychological space in which the stimuli are represented is stretched along that dimension, resulting in a decrease in similarity between stimuli which differ on the dimension. Conversely, if relatively little attention is paid to a dimension, perceived differences between stimuli on that dimension will be reduced, resulting in an increase in similarity between stimuli which differ on the dimension. In terms of the data presented in Table 5.2, a relatively high level of attention to the *engines* dimension will decrease the similarity between stimulus 2 and the category A stimuli and stimulus 6 and the category B stimuli. Therefore, the probability that they are classified correctly as belonging to their respective categories will be reduced. In contrast, according to the same analysis, the relatively high levels of classification accuracy for stimuli 1 and

5 suggests that perceived differences that these stimuli have with other stimuli in their respective categories have been minimised, leading to the conclusion that the dimension upon which they differ is receiving relatively little attention.

Given that the category structure is such that no dimension is more predictive than any other, one plausible explanation for the uneven distribution of participants' attention is that stimulus dimensions may be more or less salient than others, (i.e. the *engines* dimension is highly salient whereas the *tail* dimension is not particularly salient). It is important to note, however, that effects of differences in salience, reflected by mean error rates from the entire course of training, can be expected to reduce as training progresses as participants are required to attend to all dimensions equally in order to achieve the criterion level of accuracy.

Transfer

The proportions of category A responses and mean RTs for each stimulus are shown in Table 5.3. An ANOVA on the mean RTs produced a significant effect of stimulus $F(7, 63) = 6.81, p < .001, MSE = 73,617$. An ANOVA on the proportions of correct responses found no significant effect of stimulus type. This latter result is possibly due to the fact that participants had been trained to a relatively high criterion of performance in the training stage, as is evidenced by the high levels of accuracy for all stimuli in the transfer stage (mean correct response proportion over all stimuli = .91, $SD = 0.022$). To test the hypothesis that more errors during category learning are accompanied by slower RTs during transfer, the mean error frequencies from the training phase were correlated with the mean RTs. The resulting correlation coefficient was .68. This moderately high positive correlation can be confirmed by studying the values for individual stimuli. Stimulus 5, for example, has the lowest error frequency in the training stage and also has the shortest RT and one of the highest accuracy rates in the transfer stage. The opposite pattern is found with Stimulus 2, which has the highest error frequency in the training stage and the longest RT and one of the lowest accuracy rates in the transfer stage.

Table 5.3: Proportions of Category A Responses (RP) and Mean Response Times (RT in ms) for Each Stimulus, Experiment 1

Stimulus	RP	RT
1 1110 (A)	0.904	1061
2 1101 (A)	0.878	1125
3 1011 (A)	0.916	1104
4 0111 (A)	0.932	1041
5 0001 (B)	0.070	854
6 0010 (B)	0.100	1077
7 0100 (B)	0.074	981
8 1000 (B)	0.124	1058

The error rates in the transfer stage were also correlated with mean RTs, producing a correlation coefficient of .61. Again, this correlation is supported by the values of individual stimuli. For example, the three stimuli with the highest accuracy rates in the transfer stage—stimuli 4, 5 and 7, are the stimuli with the shortest mean RTs and the stimulus with the lowest level of accuracy—stimulus 2, has the longest mean RT. There are several differences between accuracy levels in the transfer stage and error rates in the training stage (the correlation between these sets of values = $-.64$). For example, stimulus 4 is classified most accurately in the transfer stage although the error frequency for that stimulus in the training stage is not particularly low. In addition, stimulus 7 is also accurately classified in the transfer stage despite having a relatively high error frequency in the training stage. It should be remembered, however, that the differences in the accuracy levels between the stimuli in the transfer stage are very small.

Model-based analysis

Before applying the models to the data, one may attempt to anticipate some of the parameter values each will estimate based upon knowledge of their processing assumptions and the observed data. For example, the low error frequency in the training stage and high level of accuracy and short RT in the transfer stage for stimulus 5 may suggest that the low salience of the *tail* dimension continued to affect categorisation performance throughout the entire experiment. Similarly, the relatively low accuracy levels and long RTs for stimuli 2 and 6 may suggest that the *engines* dimension was particularly salient and that this also affected performance in the transfer stage, despite the high training criterion. One may expect all three models to allocate low and high attention weights to these dimensions respectively or, in the case of the EGCM, to give them correspondingly low and high values for the inclusion rate parameters. The high level of accuracy for stimulus 4 can also be expected to result in high attention and inclusion rate values for the *nose* dimension.

EGCM

The EGCM was applied jointly to the category A response proportions and RT data from Experiment 1 using the method described in Chapter 3. The predicted category A response proportions and RTs produced by the EGCM are displayed in Table 5.4.

Best fitting parameter values were found by using a search algorithm[1] which maximised the summed coefficient of variation (R^2) for category A response proportions and RTs. This method was used to estimate parameter values for all models throughout this research. category A response proportions were used for model optimisation rather than proportions of correct responses in order to maximise the variability in the response proportions and so reduce their effect in the estimation of total goodness-of-fit. The primary reason for doing this is to increase the effect of RT data on model optimisation because the ability of the models to predict RTs is the main focus of this investigation.

[1] In all cases of model fitting carried out for this research, the EGCM and EBRW models were optimised using the *Newton's Method* search routine built into Microsoft Excel's *Solver* optimisation tool while the ALCOVE and ALCOVE(RT) models were optimised using the *Golden Section Search* method in a computer program written by the author (see Press *et al.* (1992) for detailed discussions of these methods).

Table 5.4: Observed (Obs) and EGCM Predicted (Pred) Category A Response Proportions (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 1

	RP		RT	
Stimulus	Obs	Pred	Obs	Pred
1 1110 (A)	0.904	0.854	1061	1059
2 1101 (A)	0.878	0.865	1125	1095
3 1011 (A)	0.916	0.921	1104	1094
4 0111 (A)	0.932	0.908	1041	1091
5 0001 (B)	0.070	0.047	854	856
6 0010 (B)	0.100	0.043	1077	1072
7 0100 (B)	0.074	0.048	981	978
8 1000 (B)	0.124	0.034	1058	1056

The EGCM had twelve free parameters, four dimension processing rates, q, a generalisation parameter, c, three dimension utility values, u, (the utility value of the fourth dimension is constrained by the values of the other three as all utility values are required to sum to 1), a category response bias parameter, β, the parameters Ψ and θ in the power function which determines the expected duration of dimension processing (Equation 3.17), and a residual time parameter, t_{res}. The best fitting parameter values estimated for the model are shown in Table 5.5.

Table 5.5: Best-Fitting Parameter Values for EGCM, Experiment 1. Note. The value of the utility parameter for the *tail* dimension (in brackets) is constrained by the utility values of the other three.

Parameter	Value
$q(nose)$	10.0
$q(wings)$	0.430
$q(engines)$	2.280
$q(tail)$	0.004
Ψ	0.473
t_{res}(ms)	855
$u(nose)$	0.193
$u(wings)$	0.191
$u(engines)$	0.290
$[u(tail)]$	0.326
c	8.050
θ	9.414
β	0.319

The model provided a good fit to both the RT data ($R^2 = .931, RSS = 3530$) and the choice proportion data ($R^2 = .988, RSS = 0.016$). In particular, the model was able to predict the short RTs for stimuli 5 and 7 and also predicted that stimuli 2 and 3 had the longest RTs. However, the EGCM did not match the observed data by predicting that stimulus 4 was the most accurately classified stimulus but predicted that stimulus 8 was classified more accurately than the rest, despite the fact that this was actually the least accurately classified stimulus of the eight.

As expected, the dimension inclusion rate parameter values estimated by the model indicate that the *nose* and *engines* dimensions were most salient and that the *tail* dimension was the least salient of all. The utility parameter for the *tail* dimension was given the highest value, possibly indicating that, in order to generate a set of optimal predictions for the stimuli, (in particular the relatively low level of accuracy for stimulus 1 which also differs from the other stimuli in its category on the *tail* dimension), the model was required to compensate for the low inclusion rate parameter given to the *tail* dimension. The category bias parameter was relatively low, indicating that, according to the EGCM, participants had a slight tendency to favour a category B response (note that a value of $\beta = 0.5$ indicates no category bias, $\beta > 0.5$ indicates a category A bias and $\beta < 0.5$ indicates a category B bias). Because of the relationship between response accuracy and RT embodied by the model, this is probably due to the fact that the average observed RT for category B stimuli is 360 ms shorter than that for the category A stimuli. This is also reflected in the higher response proportions for the category B stimuli predicted by the model.

EBRW

The EBRW model was also applied to the response proportions and RTs from Experiment 1 using the same goodness-of-fit measure as was used with the EGCM. The predicted response proportions and RTs produced by the EBRW are shown in Table 5.6.

The model had nine free parameters, the generalisation value c, three dimensional weight values, w, (as with the utility values in the EGCM, the value of the fourth weight value was constrained by the values of the other three), the time constant parameter, α, two category response criterion parameters, A and B, and the slope, k, and y-intercept, μ_R parameters used in the linear regression to transform the values produced by the EBRW into predicted RTs. The best-fitting parameter values estimated by the model are displayed in Table 5.7. The EBRW accounted for 88.1% of the variance in the response proportion data ($RSS = 0.158$) and 66.7% of the variance in the RT data ($RSS = 16646$).

The EBRW did predict that stimuli 5 and 7 have the shortest RTs and also predicted that stimulus 2 had one of the longest mean RTs. However, the overall closeness of the predicted RTs to the observed values was less than that of the EGCM. The accuracy predictions for stimuli 2 and 4 were very short of the observed values and, in line with the EGCM, the EBRW also predicted generally higher levels of accuracy for the category B stimuli than for those from category A. Also like the EGCM, this latter result is most likely due to the relationship between response accuracy and RT implicit in the EBRW and the 360 ms difference between average RTs for category A and category B stimuli.

Note that in Table 5.7, the category B response criterion is less than that for category A. This is due to the average RT difference between the categories. Because of the model's assumption that higher accuracy is correlated with shorter RT, the model predicts higher

Table 5.6: Observed (Obs) and EBRW Predicted (Pred) Category A Response Proportions (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 1

	RP		RT	
Stimulus	Obs	Pred	Obs	Pred
1 1110 (A)	0.904	0.954	1061	1086
2 1101 (A)	0.878	0.611	1125	1104
3 1011 (A)	0.916	0.866	1104	1096
4 0111 (A)	0.932	0.662	1041	1104
5 0001 (B)	0.070	0.002	854	927
6 0010 (B)	0.100	0.119	1077	1022
7 0100 (B)	0.074	0.016	981	954
8 1000 (B)	0.124	0.092	1058	1009

response proportions for the category B stimuli than for stimuli in category A. The best-fitting attention weight parameters found by the model were as expected, with the *nose* and *engines* dimensions being given a high weight and the *tail* dimension given no weight at all.

Given the explanation above, it is likely that the poor fit to the response proportion data is due to the effect of the model attempting to fit the RTs. Therefore, to test the ability of the model to predict response proportions alone, the model was applied to the response proportion data with six free parameters (the k and μ_R parameters were omitted because they are involved in RT predictions and only one parameter was used for the $+A$ and $-B$ category response criteria). This model was able to fit the response proportion data much more closely than the combined model, ($R^2 = .998, RSS = 0.002$). The model was then reapplied to the RT data by fixing the new parameter values and estimating the k and μ_R parameters again. However, the EBRW was unable to provide a close fit to the RT data, ($R^2 = .186, RSS = 41891$). As the combined R^2 of this model was considerably less than the first model, however, a third model with seven free parameters was applied to the response proportion data (the k and μ_R parameters were omitted as before but two parameters were used for the $+A$ and $-B$ category response criteria). This model provided a similar fit to the data ($R^2 = .978, RSS = 0.029$) but when the parameters were again fixed and the model applied to the RT data, (estimating the k and μ_R parameters), a slightly better fit resulted, ($R^2 = .332, RSS = 34400$). The difference between the fits of the two models is due to the values of the category response criteria parameters (in the second model, $+A = -B = 2$ whereas in the third model $+A = 2, -B = 1$).

One may conclude, therefore, that although the EBRW model is able to provide a close fit to the response proportion data, it is unable to give as accurate an account of the combined response proportion and RT data than the EGCM. This is due to the difference in average RT between the categories, a difference which also had an effect on the predictions of the EGCM. These differences in performance between the models are in line with recent findings of Lamberts (submitted, Experiment 1) who also found that the EBRW fitted the choice accuracy and RT data worse than the EGCM. Although the

Table 5.7: Best-Fitting Parameter Values for EBRW, Experiment 1. Note. The value of the weight parameter for the *tail* dimension (in brackets) is constrained by the values of the other three.

Parameter	Value
$w(nose)$	0.37
$w(wings)$	0.18
$w(engines)$	0.45
$[w(tail)]$	0.00
c	5.97
α	27.04
A	2
B	1
k	6.03
μ_R(ms)	745

EBRW model performed considerably worse than the EGCM in both Lamberts' experiment and the experiment reported here, it is possible that this is due to the fact that it had 3 parameters less than the EGCM. The most plausible reason for the difference in performance is that the combination of inclusion rate parameters and utility values give the EGCM the additional flexibility to overcome the additional constraints involved in accounting for combined response proportion and RT data.

ALCOVE(RT)

To apply the ALCOVE(RT) model to RT data, a suitable criterion for stopping the network to create a dependent time variable has to be decided upon. Two criteria are commonly used. The first is to take the number of cycles required for the activation of one or more output units to reach a threshold value. The second is to take the number of cycles required for the activation of one or more output units to settle to a stable state (i.e. when changes in output unit activation fall below a specified value). As there are no a priori reasons for choosing one criterion over the other, the former was used. In applications of the model to the data, therefore, a threshold parameter, ϑ, is estimated, the value of which must be reached by one of the output units for the network cascade process to terminate. The number of cascade cycles required by the network to reach ϑ is taken as the time variable and transformed by a simple linear regression procedure to produce a predicted RT. The linear regression function requires two further parameters, υ to determine the slope of the regression line and ω to determine the y-intercept.

When applied to combined response proportion and RT data, the ALCOVE(RT) has a total of nine parameters, the threshold and linear regression parameters mentioned above, together with the c parameter representing the specificity of the hidden units, the association and attention learning rate parameters, λ_w and λ_α, the category unit response mapping parameter, ϕ and the hidden and output layer cascade rate parameters, τ_{hid} and τ_{out}. As with the EGCM and EBRW models, best-fitting parameter values were obtained by maximising summed R^2 for response proportions and RTs.

To apply the model, the network was first trained on the full set of eight stimuli for thirty epochs using the batch updating method described in Chapter 4. The actual number of epochs required to train the network is arbitrary when combined with a parameter estimation procedure because differences in the number of epochs are compensated for by the values of the association and attention learning rate parameters. After training, the network was again presented with each stimulus and allowed to cascade until one of the category units reached the threshold value. The cascade cycle at which the threshold value was reached, (having first been transformed into a predicted RT by the linear regression function), and the category A response probability at this cycle are then recorded and employed in the parameter optimisation routine.

The final predicted response proportions and RTs produced by the ALCOVE(RT) model are displayed in Table 5.8 while the best-fitting parameter values are shown in Table 5.9. Although the model was able to account for 99.7% of the variance in the response proportion data ($RSS = 0.003$), it did so by predicting two values representing the average response proportion for each category and was therefore unable to predict choice proportions for individual stimuli. Note that the choice proportion predicted for category B stimuli equals $1 -$ the predicted proportion for stimuli in category A and vice versa. Despite the fact that ALCOVE(RT) produces just two average response proportions for the eight stimuli, the predicted values account for a greater percentage of the variance than those produced by the EGCM and EBRW models. This is because (as the ANOVA indicated) there is very little difference between the observed response proportions for the stimuli in each category. Therefore there will be very little difference between individual observed values in a category and the mean value for that category, resulting in a very small value for the residual sum of squares used in the computation of the R^2 value.

The network predictions are primarily the result of the regularity of the category structure which created a uniform pattern of connection weights in the network during training (after 30 epochs, the output-exemplar association weights all had a value of 1 (in the case of positive associations) or -1 (for negative associations) while all input-exemplar attention strengths had a final value of 0.808).

Table 5.8: Observed (Obs) and ALCOVE(RT) Predicted (Pred) Category A Response Proportions (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 1

	RP		RT	
Stimulus	Obs	Pred	Obs	Pred
1 1110 (A)	0.904	0.908	1061	1038
2 1101 (A)	0.878	0.908	1125	1038
3 1011 (A)	0.916	0.908	1104	1038
4 0111 (A)	0.932	0.908	1041	1038
5 0001 (B)	0.070	0.092	854	1038
6 0010 (B)	0.100	0.092	1077	1038
7 0100 (B)	0.074	0.092	981	1038
8 1000 (B)	0.124	0.092	1058	1038

Given such a category structure, this situation will occur for any set of parameter values because the model's predictions are constrained by the error-driven learning mechanism which is determined to a large extent by the category structure and teacher values. In this respect, the ALCOVE model has a relatively limited ability to account for categorisation data compared to the EGCM and EBRW models.

The RTs predicted by the network are not close to the observed values, ($R^2 = 0.00, RSS = 51489$). The model produces one value which represents the average RT over all stimuli. This is also due to the category structure because all stimuli produce the same output from the category units, (in line, of course, with the category label associated with the stimulus). The rate of activation accumulation of the category units in the cascade process, therefore, is also the same for each stimulus, resulting in one predicted response time.

Table 5.9: Best-Fitting Parameter Values for ALCOVE(RT), Experiment 1

Parameter	Value
c	3.199
λ_w	0.096
λ_α	0.096
ϕ	4.216
τ_{hid}	0.586
τ_{out}	0.222
ϑ	0.191
υ	0.684
ω	1036.258

Although the correlation between classification accuracy and mean RT is moderately positive at .61, which is what ALCOVE(RT) predicts, the inflexibility of the model, brought about by the method of category learning embodied in the network, means that it is not able to produce fine-grained predictions of response probabilities for individual stimuli in the same way as the EGCM and EBRW.

Discussion

There were three main reasons for conducting Experiment 1. The first was to test the EGCM, EBRW and ALCOVE(RT) models on data from a standard speeded category learning experiment and to analyse and compare their predictions. The second was to generate accuracy and RT data using stimuli with binary valued dimensions so that the ability of the EBRW to model such data could be analysed for the first time. The third was to investigate the relationship between categorisation accuracy and RT.

In the context of these goals, a number of issues have emerged from the results of the experiment. Firstly, the three models had varying success in accounting for the data. The EGCM provided the closest fit to the combined data followed by the EBRW and then the ALCOVE(RT) model. It was found in both the EGCM and EBRW (in particular the latter) that differences between observed and predicted values were the result of the relationship between accuracy and RT embodied in the models. In attempting to account for the shorter average RT for the category B stimuli, both predicted higher levels of accuracy for those stimuli than was actually observed. The greater flexibility of the EGCM in terms of its ability to accommodate dimensional salience and category structure-related attention distribution separately is likely to be the main reason for the model's closer fit to the combined data. An examination of the inclusion rate parameter values yielded by the EGCM model indicated large differences in dimension salience. In particular, the model suggested that the *nose* and *engines* dimensions were most salient with the *tail* dimension being least salient of all. The EBRW model also reflected these differences in its final dimension attention weight values. The *nose* and *engines* dimensions are weighted most strongly whereas the *tail* dimension is not weighted at all. These differences in dimension processing suggested by the model come entirely from the differences in RTs. When the EBRW was applied to the category A response proportions separately, it accounted for 99% of the variance in the data and produced much more evenly distributed dimension weight parameter values ($w_{nose} = 0.258$, $w_{wings} = 0.225$, $w_{engines} = 0.284$, and $w_{tail} = 0.233$). This lends further weight to the suggestion that the success of the EGCM against the EBRW is due to its ability to reflect differences in dimension salience while still maintaining the importance of all dimensions in categorisation as determined by the category structure (note that the dimension utility parameter values of the EGCM were relatively evenly distributed—$w_{nose} = 0.193$, $w_{wings} = 0.191$, $w_{engines} = 0.290$, and $w_{tail} = 0.326$, compared to those of the original EBRW model). In comparison, the limited resources of the EBRW (i.e. a single attention weight for each dimension) cannot adequately capture the opposing influences of dimension salience and category structure.

If the reason for the relative inability of the EBRW to provide a reasonable account of the combined data compared to the EGCM is because of its inflexibility in dealing with dimensional salience and category related attention distribution, there is little evidence that the nature of the stimuli (i.e. the binary valued dimensions) provide any further source of difficulty for the model. However, any firm conclusions about this cannot be made on the basis of results from this experiment alone.

The ALCOVE(RT) model was the least successful of the three in accounting for the data. There are a number of plausible explanations for this lack of success. The first point that has to be remembered, however, is that the ALCOVE model is very dissimilar to the GCM, EGCM and EBRW in that it is a model of category learning and therefore has a number of constraints not shared by the other models. For example, fitting ALCOVE to data is not just a matter of optimising the free parameters because the learning mechanism involves a nominal training feedback value which results in category responses being determined largely by the structure of the category. Unlike the GCM, EGCM and EBRW models, adjustments to parameter values in the ALCOVE model do not result in a fine tuning of the networks responses to more closely fit the observed data but in rates of attention and exemplar-category association learning. Given these differences, it is perhaps unrealistic to expect the ALCOVE(RT) model to fit combined response accuracy and RT data as closely as the EGCM and EBRW models. The limitations of the ALCOVE(RT) model have been highlighted by two factors—a category structure which is relatively unusual because of its high level of regularity and the rather unexpected level of variation in RTs given the lack of variation in the accuracy data.

Concerning the relationship between categorisation accuracy and RT—given the regular nature of the category structure, the relatively high performance criterion in the training stage and the high levels of categorisation accuracy for all of the stimuli in the transfer stage, the fact that there were significant differences in RTs shows that the relationship between classification accuracy and RT is not simple. In this respect, Experiment 1 has been successful in showing that, although category structures can be learned to a high level of overall accuracy (to the extent that there are no significant differences between the accuracy for individual stimuli in the transfer stage), significant differences in response times may remain which may give important clues about information being used in categorisation (i.e. differences in dimensional salience), which would otherwise not be available in an examination of response proportions alone.

In Chapter 6, a second experiment is described which attempts to build upon the results of Experiment 1 by further testing the EGCM, EBRW and ALCOVE(RT) models while including an additional level of complexity in the form of a more irregular category structure.

Chapter 6

Experiment 2

Introduction

The results of Experiment 1 suggest a number of logical developments for a second experiment. Experiment 2 is intended to address essentially the same issues as the first—to test of the ability of the EGCM, EBRW and ALCOVE(RT) to account for combined categorisation accuracy and RT data, to determine the ability of the EBRW to model categorisation data with binary valued stimuli, and to investigate the relationship between categorisation accuracy and RT. The primary distinction between Experiments 1 and 2 is the category structure employed. Although the EGCM and EBRW provided relatively reasonable fits to the data from Experiment 1, the regularity of the category structure posed a particular difficulty for the ALCOVE(RT) model. The category structure used in Experiment 2, therefore, is designed to increase the variation in stimulus typicality by being less regular. In Experiment 1, the summed similarity of each stimulus to the other stimuli in its category and to the other stimuli in the alternative category (given equal attention to stimulus dimensions) is equal. The category structure in Experiment 2 is such that this is not the case because one stimulus in each category is more similar to three members of the opposite category than it is to members of its own and the exact opposite of the fourth alternative category member. In Experiment 1, the regularity of the category structure also had the consequence that no stimulus dimension was more predictive of a category than any other. This is not the case in Experiment 2 because on of the dimensions is more predictive than the others, result in an expected increase in attention to that dimension. Finally, the distribution of *features* (i.e dimension values) across categories in Experiment 1 was uneven so that, for example, the probability that a category A stimulus had a value of 1 on each of the dimensions was .75 whereas that for a category B stimulus was .25. In Experiment 2, these probabilities are more evenly distributed so that, apart from the more predictive dimension where the probabilities are the same as in Experiment 1, the probability that a stimulus from either category has a value of 1 on a dimension is .5. This may have the effect of making the category structure harder to learn because the distribution of dimension values over categories is more even. Secondly, the different stimulus in each category should be hardest to learn and can therefore, being closer to the category decision boundary than other members of its category, be expected to have the longest mean RT.

Apart from the category structure used, all aspects of Experiment 2 are the same as Experiment 1. This will allow some general qualitative comparisons to be made between the data from the experiments. The variation in category structure between experiments, however, precludes a more detailed statistical analysis.

Method

Participants

Ten undergraduate and postgraduate psychology students from the University of Birmingham participated in the experiment. The undergraduates who took part were given credit towards the Psychology department's research participation scheme.

Apparatus and stimuli

The apparatus and stimuli used in Experiment 2 were the same as those employed in Experiment 1.

Design and procedure

The category structure used in Experiment 2 is shown in Table 6.1. The structure is identical to that used in Experiment 1 except that stimuli 4 and 8 are given opposite values on the *wings*, *engines* and *tail* dimensions. Note that a consequence of the structural irregularities referred to earlier is that one stimulus dimension (dimension 1) is now more predictive of category membership than the other three. This should result in an increase in attention to the *nose* dimension. The experimental procedure was the same as that used in Experiment 1.

Results

Training

The mean number of blocks required to achieve two consecutive correct blocks was 46.9 ($SD = 18.0$), four blocks more than in Experiment 1. The mean error frequency for each stimulus over the course of training across all participants is presented in Table 6.2.

Although there were substantial differences in error scores between stimuli, an analysis of variance (ANOVA) on the mean correct responses did not yield a significant effect of stimulus $F(7, 63) = 2.13, p = .053, MSE = .037$.

In the training stage, participants made more errors classifying category A stimuli (mean error frequency = 19.1) than when classifying those in category B (mean error frequency = 15.5).

Table 6.1: Structure of Stimuli in Experiment 2

Structure	Stimulus Number	Dimension Nose	Wings	Engines	Tail
Category A	1	1	1	1	0
	2	1	1	0	1
	3	1	0	1	1
	4	0	0	0	0
Category B	5	0	0	0	1
	6	0	0	1	0
	7	0	1	0	0
	8	1	1	1	1

Table 6.2: Mean Error Frequencies During Training, Experiment 2

Stimulus	Error Frequency
1 1110 (A)	16.8
2 1101 (A)	19.5
3 1011 (A)	18.6
4 0000 (A)	21.6
5 0001 (B)	15.4
6 0010 (B)	14.5
7 0100 (B)	13.0
8 1111 (B)	19.0

One notable feature of the data is that stimuli 4 and 8, mentioned earlier as being particularly untypical of their respective categories, were the stimuli with the highest error rates in their categories. This indicates that participants, at least in the early stages of training, considered these stimuli to be more similar to stimuli from the alternative category than other stimuli in their respective categories. Together with stimuli 4 and 8, the stimuli which was categorised least accurately were stimuli 2 and 3. Stimulus 2 actually had a higher error rate than stimulus 8 while that for stimulus 3 was also very close. This may suggest that, after the *nose* dimension, participants were paying the greatest deal of attention to the *wings* and then the *engines* because they share the same values on those dimensions with two stimuli from category B and stimulus 4 which is being misclassified most often. However, this explanation is not supported by the error frequencies for stimuli 6 and 7, which also share the same values on the *wings* and *engines* dimensions with two stimuli from category A and stimulus 8, because these stimuli are categorised most accurately during training. It is difficult, therefore, to draw many clear conclusions from the error data concerning the distribution of attention over stimulus dimensions because no simple or distinct pattern is discernible.

Transfer

The proportions of category A responses and mean RTs for each stimulus are shown in Table 6.3. An ANOVA on the mean RTs produced a significant effect of stimulus $F(7, 63) = 2.7, p < .05, MSE = 193,078$. Category responses in Experiment 2 were on average 309 ms slower and generally less accurate than those in Experiment 1.

Table 6.3: Proportions of Category A Responses (RP) and Mean Response Times (RT in ms) for Each Stimulus, Experiment 2

Stimulus	RP	RT
1 1110 (A)	0.832	1370
2 1101 (A)	0.868	1391
3 1011 (A)	0.738	1575
4 0000 (A)	0.774	1157
5 0001 (B)	0.158	1279
6 0010 (B)	0.130	1253
7 0100 (B)	0.120	1252
8 1111 (B)	0.132	1497

The mean error frequencies from the training phase were again correlated with the mean RTs, resulting in a correlation coefficient of 0.22, while a similar correlation between the error rates in the transfer stage and RTs produced a coefficient of 0.24. Both of these correlations are considerably smaller than those from Experiment 1. The correlation between accuracy levels in the transfer stage and error rates in the training stage was $-.54$. An ANOVA on the mean proportions correct found no significant effect of stimulus type. So, as with Experiment 1, significant differences in RT during transfer were not mirrored by participants' levels of classification accuracy and, like Experiment 1, this is likely to be a result of the relatively high performance criterion in the training stage. It should be noted that the level of accuracy in the transfer stage of Experiment 2 was, in general, lower than that from Experiment 1.

As with Experiment 1, the average RT for category B stimuli was lower than that for stimuli from category A, although only by 53 ms, rather than 360 ms. The significant differences in average RTs in Experiment 2 are less surprising given the category structure. However, the pattern of RTs does not follow the form suggested earlier. Although stimulus 8 does have one of the slowest mean RTs of the eight stimuli, the level of accuracy for the stimulus is not particularly low. Stimulus 4, however, is the stimulus most rapidly responded to on average, although the level of accuracy for stimulus 4 is relatively low. This suggests that some form of speed/accuracy trade-off may be occurring in participants' responses to these stimuli. A speed/accuracy trade-off does not seem to be occurring with all stimuli, however, because the stimulus with the lowest level of accuracy (stimulus 3) is also the stimulus with the slowest average RT while the stimulus most accurately responded to (stimulus 7) also has one of the shortest RTs. The large differences in accuracy and RT for these two stimuli are surprising given that both stimuli differ from the other stimuli in their category on the same dimension. Taking one of the stimuli in isolation, an explanation of a high accuracy/short RT or low accuracy/long RT combination may be given in terms of a low or high level of attention being paid to the *wings* dimension. The combination of results for stimuli 3 and 7 rules out this analysis, however. As with the error data from the training stage, it is difficult to draw many distinct conclusions from the transfer data about the distribution of attention over stimulus dimensions because no simple pattern can be identified.

Model-based analysis

As with Experiment 1, one may try to forecast some of the estimated parameter values for the three models before applying the models to the data. Firstly, because of the category structure, one may expect all three models to allocate a high level of attention to the *nose* dimension, or, in the case of the EGCM, a high utility value. Secondly, the relatively small difference between the average RT for category A and category B stimuli compared with that of Experiment 1 can be expected to result in a smaller value of the β parameter in the EGCM and a similar value for the A and B response boundary parameters in the EBRW.

EGCM

The EGCM was applied jointly to the category A response proportions and RT data from Experiment 2 using the same procedure employed for Experiment 1. The predicted category A response proportions and RTs produced by the model are shown in Table 6.4 while the parameter values estimated by the model are shown in Table 6.5.

Table 6.4: Observed (Obs) and EGCM Predicted (Pred) Category A Response Proportions (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 2

	RP		RT	
Stimulus	Obs	Pred	Obs	Pred
1 1110 (A)	0.832	0.810	1370	1356
2 1101 (A)	0.868	0.936	1391	1426
3 1011 (A)	0.738	0.789	1575	1497
4 0000 (A)	0.774	0.533	1157	1266
5 0001 (B)	0.158	0.073	1279	1293
6 0010 (B)	0.130	0.130	1253	1238
7 0100 (B)	0.120	0.099	1252	1184
8 1111 (B)	0.132	0.109	1497	1518

The EGCM accounted for 91.8% of the variance in the category A response proportion data (RSS = 0.074) and for 81.6% of the variance in the RTs (RSS = 24882). Neither fit was as close as that to the data from Experiment 1. The greatest difference between the observed and predicted values can be seen for stimulus 4 which is predicted by the model to have a much lower level of accuracy than actually occurred in the experiment.

The greatest difference between observed and predicted RTs (109 ms) is also for stimulus 4. The EGCM does predict that stimulus 8 has the longest average RT and that the mean RT for stimulus 4 is relatively short.

The EGCM does not predict that the *nose* dimension has a higher utility value that the other dimensions as was expected, although, as in Experiment 1, dimension 1 is given the highest inclusion rate value of the four. Also consistent with Experiment 1 is the value given to the Category response bias parameter, β, which indicates a slight bias in favour of a category B response. When comparing the values of the inclusion rate and utility parameters across the experiments, it is apparent that in three of the four dimensions, a lower value for one of the parameters in an experiment is accompanied by a correspondingly higher value in the other. This may indicate that some form of interaction is occurring and that these parameters tend to compensate for each other.

Table 6.5: Best-Fitting Parameter Values for EGCM, Experiment 2. Note. The value of the utility parameter for the *tail* dimension (in brackets) is constrained by the utility values of the other three.

Parameter	Value
$q(nose)$	5.0
$q(wings)$	0.002
$q(engines)$	0.003
$q(tail)$	1.360
Ψ	0.731
t_{res}(ms)	900
$u(nose)$	0.281
$u(wings)$	0.147
$u(engines)$	0.381
$[u(tail)]$	0.191
c	14.94
θ	1.560
β	0.342

EBRW

The EBRW was applied to the data using the same procedure as in Experiment 1. It was found that using a model parameter for each response boundary gave no better fit to the data and so only one parameter representing both boundaries was used. The predicted response proportions and RTs produced by the EBRW are displayed in Table 6.6 while the parameter values estimated by the model are shown in Table 6.7.

The EBRW gave reasonable fits to the RT and response proportion data ($R^2 = 0.666$, $RSS = 45247$ and $R^2 = 0.794$, $RSS = 0.186$ respectively) although, as with the EGCM, these fits were not as close as those for Experiment 1. The EBRW predicts that responses to stimulus 4 are less accurate than was actually found and also that stimulus 3 is the most accurately classified category A stimulus rather than the least. However, the model

Table 6.6: Observed (Obs) and EBRW Predicted (Pred) Category *A* Response Proportions (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 2

	RP		RT	
Stimulus	Obs	Pred	Obs	Pred
1 1110 (A)	0.832	0.775	1370	1425
2 1101 (A)	0.868	0.785	1391	1422
3 1011 (A)	0.738	0.878	1575	1507
4 0000 (A)	0.774	0.500	1157	1170
5 0001 (B)	0.158	0.336	1279	1208
6 0010 (B)	0.130	0.218	1253	1365
7 0100 (B)	0.120	0.235	1252	1307
8 1111 (B)	0.132	0.301	1497	1370

also accurately predicts the relatively large RT for stimulus 3 and the relatively small RT for stimulus 4.

Table 6.7: Best-Fitting Parameter Values for EBRW, Experiment 2. Note. The value of the weight parameter for the *tail* dimension (in brackets) is constrained by the values of the other three.

Parameter	Value
$w(nose)$	0.037
$w(wings)$	0.136
$w(engines)$	0.089
$[w(tail)]$	0.738
c	14.141
α	0.644
A	1
k	949.106
μ_R(ms)	95.955

One interesting pattern in the parameter values across the experiments is the almost doubling of the value of the discrimination parameter, c, given by both models for Experiment 2. This suggests that participants in Experiment 2 were required to discriminate the stimuli to a larger extent than those in Experiment 1 which is what would be expected given the more complex category structure used in Experiment 2.

ALCOVE(RT)

The ALCOVE(RT) model was applied to the data from Experiment 2 using the same procedure as was used in Experiment 1. The resulting values predicted by the model are shown in Table 6.8 and the best-fitting parameter values found by the model are shown in Table 6.9.

Table 6.8: Observed (Obs) and ALCOVE(RT) Predicted (Pred) Category *A* Response Proportions (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 2

	RP		RT	
Stimulus	Obs	Pred	Obs	Pred
1 1110 (A)	0.832	0.838	1370	1346
2 1101 (A)	0.868	0.838	1391	1346
3 1011 (A)	0.738	0.838	1575	1346
4 0000 (A)	0.774	0.821	1157	1346
5 0001 (B)	0.158	0.162	1279	1346
6 0010 (B)	0.130	0.162	1253	1346
7 0100 (B)	0.120	0.162	1252	1346
8 1111 (B)	0.132	0.179	1497	1346

As with Experiment 1, the ALCOVE(RT) was unable to account well for the combined data. The model gave a close fit to the response proportion data ($R^2 = 0.980$, $RSS = 0.018$) but was not able to account for accuracy levels of individual stimuli. The model did predict less accurate responses for stimuli 4 and 8, however. At the end of training, the values of association weights between exemplar units representing stimuli 1, 2, 3, 5, 6 and 7 and the category units were either 0.633 or -0.633 depending on the association. The same values for exemplar units representing stimuli 4 and 8, however, were 0.565 and -0.565, indicating that the network had learned the relative ambiguity of category membership for these stimuli.

The attention strengths between stimulus dimensions and exemplar units after training were as expected. The value for the *wings*, *engines* and *tail* dimensions was 0.809 whereas that for the *nose* dimension was 0.841. The predicted values for the RTs were, as for Experiment 1, simply the mean of the observed values ($R^2 = 0.000$, $RSS = 135538$). This is because stimuli 1, 2, 3, 5, 6 and 7 produce the same overall category unit output. Optimising the parameters of the regression equation which transforms cascade terminations into predicted RTs on such a homogenous output yields a relatively low scale value, υ, and a large value for the intercept, creating a situation in which small differences in cascade termination values do not result in corresponding differences in predicted RTs.

Discussion

The results of Experiment 2 follow roughly the same pattern as Experiment 1. After learning the category structure to a relatively high level of accuracy, participants' classification accuracy remained relatively consistent across stimuli in the transfer stage, producing no significant differences. However, participants' average response times to individual stimuli in the transfer stage did differ significantly. If the cause of these differences between accuracy and response time performance is the training to criterion procedure in the training stage, the use of an alternative training procedure—training for a fixed, and possibly small, number of trials—may have the effect of increasing the variance in classification accuracy.

Table 6.9: Best-Fitting Parameter Values for ALCOVE(RT), Experiment 2

Parameter	Value
c	3.0
λ_w	0.052
λ_α	0.038
ϕ	3.822
τ_{hid}	0.473
τ_{out}	0.114
ϑ	0.191
υ	0.213
ω	1344.512

In fact, this is the training procedure adopted for the remaining experiments, although for different reasons. The high level of overall classification accuracy in Experiments 1 and 2 has the further complicating effect of reducing the amount of variance in the data. One consequence of this is that models can yield a very high value of R^2 for category A response proportions while still producing predicted values quite dissimilar to the observed values.

The ability of the three models to account for the data produced by Experiment 2 also followed the same pattern as for Experiment 1. The EGCM provided a close fit to the combined data, followed by the EBRW and the ALCOVE(RT) models. The less regular category structure in Experiment 2 did produce differences in response proportion predictions for the unusual stimuli but, for reasons discussed above, these were not sufficiently different compared to the values of the other stimuli to be reflected in predicted response times.

Having carried out two preliminary speeded classification experiments to test the three models, in Chapter 7 attention is turned to the central theme of this research—the effect of stimulus frequency or *familiarity* on classification reaction time. In the following chapter, the relationship between stimulus frequency and categorisation performance is introduced and previous research into frequency effects on various aspects of categorisation is outlined. This chapter is a precursor to three subsequent chapters describing experiments designed to reveal frequency effects on classification RT and to further test the EGCM, EBRW and ALCOVE(RT) models. Because stimulus frequency is the factor of interest in classification experiments studying familiarity effects, the training procedure used in these experiments typically involves a fixed number of stimulus presentations. As discussed above, it may be expected, therefore, that variation in classification accuracy will increase. The manipulation of stimulus frequency can also be expected to increase variability in both response accuracy and RTs. It is hoped, therefore, that these changes may result in better performance from the ALCOVE(RT) model in accounting for the data.

Chapter 7

Stimulus Frequency and Categorisation Performance

Introduction

In this chapter, the relationship between stimulus frequency and categorisation performance is discussed. Firstly, a number of experimental findings for and against the notion that presentation frequency affects graded category structure are outlined. Then, an investigation into frequency effects on categorisation and a frequency sensitive version of the GCM which accounts for observed frequency effects on categorisation accuracy and typicality judgements undertaken by Nosofsky (1988) are described in detail. Finally, research carried out by Ashby, Boynton and Lee (1994) which suggested that there was no significant effect of stimulus familiarity on categorisation RT is described. The purpose of this chapter, therefore, is to introduce the various results of recent research into frequency effects to provide a theoretical background for a series of experiments in the following chapters.

Stimulus frequency and graded category structure

Although the effect of stimulus frequency on various measures of categorisation performance has been repeatedly investigated during the last thirty years, there still remains a certain amount of disagreement concerning the magnitude (or sometimes even the existence) of familiarity effects. Intuitively, it seems reasonable to assume that the more often an object is experienced as a member of a category, the more likely it is that the object will be regarded as a good example of the category. However, an early investigation by Rosch, Simpson and Miller (1976) into the factors affecting graded category structure found that frequency effects on stimulus typicality judgements were relatively insignificant compared to effects of stimulus similarity. Rosch *et al.* conducted an experiment in which participants were required to learn a category structure in which the frequency of stimulus presentation was negatively correlated with interstimulus similarity. They discovered that, in a subsequent typicality rating task, participants rated stimuli which were very similar to others in the same category but had been presented less frequently as more typical of their category than stimuli which had been presented more frequently but were less similar to other stimuli in the same category. Although they acknowledged that stimulus familiarity must have some effect on category learning, Rosch *et al.* argued that their results demonstrated that interstimulus similarity was the primary determinant of graded category structure. Research carried out by Mervis *et al.* (1976) led them to draw similar conclusions.

Other researchers, however, have uncovered firm evidence for stimulus frequency effects on typicality judgements (e.g. Ashcraft, 1978; Barsalou, 1985; Hampton & Gardiner, 1983; Malt & Smith, 1982; Schwanenflugel & Rey, 1986) and category learning (Barsalou, 1981; Knapp & Anderson, 1984). For example, Barsalou (1985, Experiment 1) found that, for a number of goal-directed and common taxonomic categories, an object's *frequency of instantiation* (defined as the number of times the object is experienced as a member of a category) was positively correlated with other measures of stimulus typicality such as judgements of goodness of example and frequency of production on request of a category example.

The joint effects of similarity and frequency on perceptual categorisation accuracy and typicality judgements have also been investigated in two category learning experiments by Nosofsky (1988). In the first of these experiments, participants were required to categorise twelve colour stimuli varying on two continuous dimensions, brightness and saturation. In two of the three conditions, a (different) stimulus from one of the categories was presented five times more often than the other stimuli in the category learning stage. In the third condition, all stimuli were presented with equal frequency during training. The two frequency manipulated stimuli were relatively good examples of their category (i.e. were relatively distant from the category boundary). In the subsequent testing stage, participants in all conditions were required to categorise all twelve stimuli again and to rate on a scale of 1 (lowest) to 10 (highest) how confident they were about the correctness of their category judgement. Participants were also asked to rate on a similar scale how typical or how good an example each stimulus was of its category and to judge from all paired combinations of stimuli from the frequency manipulated category, the better example of the category. Nosofsky found that the number of errors during the category learning stage for the two more frequently presented stimuli in the manipulation conditions was significantly less than in the other conditions. Nosofsky also found that the frequency manipulation had a significant effect on classification confidence, typicality ratings and typicality pair comparisons for the same stimuli in the testing stage.

In the second experiment, Nosofsky manipulated the presentation of a relatively poor category example (i.e. one relatively close to the decision bound). There were three conditions, the manipulation between the conditions being the presentation frequency of one stimulus during the category learning stage. In the first condition all stimuli were presented with equal frequency during training. In the second and third conditions, the ratio of presentation frequencies for the manipulated stimulus was 3:1 and 5:1 respectively. All other aspects of the second experiment were the same as the first. Nosofsky found that the frequency manipulation in the 5:1 condition had a significant effect on the number of classification errors during the training stage compared to the equal frequency condition and also significantly affected typicality ratings for the manipulated stimulus relative to the equal frequency condition in the testing stage. A further important finding from Nosofsky's investigation was the observation that classification accuracy and typicality ratings for stimuli which were very similar to the high frequency stimuli and in the same category increased whereas the same measures for stimuli which were very similar to the high frequency stimuli but were in the opposite category decreased.

To quantitatively model the classification learning data in the two experiments, Nosofsky modified the GCM to take presentation frequency into account. This frequency sensitive GCM differs from the frequency insensitive version in that, rather than assuming that multiple presentations of the same stimulus are represented by a single memory trace, it is assumed that each stimulus presentation is stored as a separate representation in memory. This assumption is also found in Hintzman's (1986, 1988; Hintzman & Block, 1971) multiple trace model of recognition memory and classification, *MINERVA 2*, and in Estes' (1986a, 1986b) array model of category learning. Frequency information is formally incorporated into the GCM by the addition of a variable, μ_j indicating the frequency of exemplar j in the category choice rule (Equation 2.8)

$$P(J|i) = \frac{\beta_J \sum_{j \in J} \mu_j \eta_{ij}}{\sum_K \left(\beta_K \sum_{k \in K} \mu_k \eta_{ik}\right)}. \tag{7.1}$$

The frequency sensitive GCM was able to account for the category learning and typicality rating data from both experiments better than the frequency insensitive GCM and frequency sensitive and insensitive prototype models.

Stimulus frequency and categorisation RT

The effect of stimulus frequency on categorisation response time has been relatively little studied and the investigations which have taken place have had conflicting conclusions. Ashby, Boynton and Lee (1994) carried out three speeded category learning experiments using stimuli varying on two continuous dimensions to test the predictions of the RT-Distance (RT-D) hypothesis generated by the decision bound model. The categories used in the experiments took the form of bivariate normal distributions and the values of stimulus dimensions on each trial are randomly sampled from one of the distributions (see Chapter 2 for a detailed explanation of this approach). The shape of each category distribution and the amount of overlap between the categories were manipulated to create three conditions. To illustrate this manipulation, the contours of equal likelihood from the three conditions of Experiment 1 are displayed in Figure 7.1. In the diagram, the probability density of the two categories, A and B are indicated by $f(A)$ and $f(B)$ respectively while the decision boundary for each condition is described by the diagonal line.

The stimuli used in the experiments were either horizontal and vertical lines of varying length joined to form a corner, rectangles varying in height and width, or semicircles varying in size and orientation of an internal radial line. Participants were required to learn to categorise one of the stimulus types according to one of the category structures depicted in Figure 7.1. Each participant was presented with a total of 300 training trials

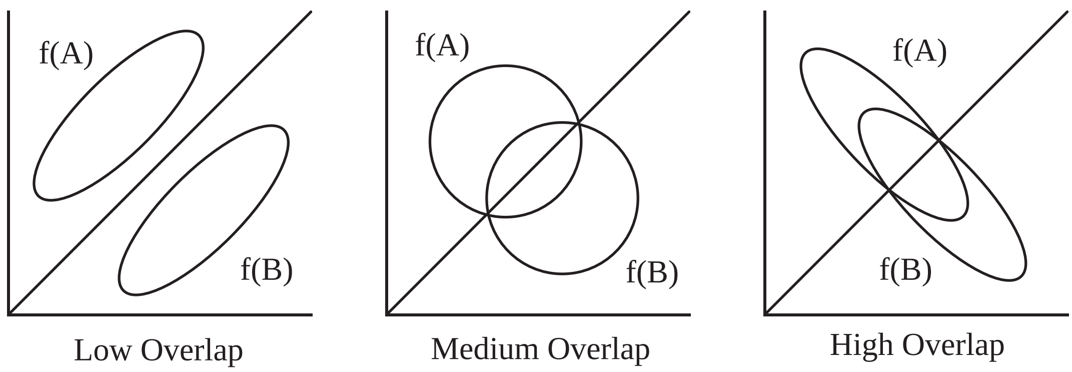

Figure 7.1: Contours of equal likelihood from the three conditions in Ashby, Boynton & Lee (1994), Experiment 1

and was asked to respond as quickly as possible. Note that, because of the overlap between the category distributions, perfect classification performance was unachievable.

In two of the experiments (Experiments 1 and 2), as well as testing the RT-D hypothesis, Ashby *et al.* also tested the predictions of, amongst others, an *RT-familiarity* (RT-F) hypothesis which states that categorisation RT is inversely proportional to stimulus *familiarity*. In their formulation, the familiarity of stimulus i, F_i, is defined as the summed similarity of i to all exemplars, j, in all categories, J, under consideration

$$F_i = \sum_J \left(\sum_{j \in J} \mu_j \eta_{ij} \right). \tag{7.2}$$

Note that this conception of familiarity also rests on the assumption that each individual stimulus presentation is represented as a trace in memory. The measure of similarity used in the computation of familiarity is the same as that used in the GCM (Equations 2.4 and 2.7).

In their subsequent analysis of the data from Experiments 1 and 2, Ashby *et al.* found no significant correlation between stimulus familiarity and response time (the RT-F hypothesis predicts a negative correlation) but that there was a modest negative correlation between RT and distance from decision bound, as predicted by the RT-D model. This led them to conclude that their experiments gave general support for the RT-D hypothesis but provided no evidence to support the claim that stimulus familiarity played a role in determining categorisation RT.

More recent research undertaken by Nosofsky and Palmeri (1997, Experiment 2), however, has provided support for a relationship between stimulus frequency and categorisation RT. In the following chapter, I will describe Nosofsky and Palmeri's experiment in detail and examine the ability of the EGCM and EBRW models to account for their findings. I will then describe the results of applying the ALCOVE(RT) model to the data from this experiment.

Summary

As the above discussion illustrated, the nature and extent of stimulus frequency effects on the various measures of categorisation performance is still a matter of disagreement and active investigation. In studies where frequency effects have been observed, the effects are relatively small compared to those produced by the similarity structure of the stimuli. The aim of this chapter was to outline recent research into frequency effects on stimulus typicality judgements (i.e. on graded category structure), classification accuracy and categorisation RT. The purpose of the following chapters is to investigate more fully the effects of stimulus frequency on categorisation RT and to compare and contrast the ability of the EGCM, EBRW and ALCOVE(RT) to account for data from speeded categorisation experiments in which individual stimulus frequency is manipulated.

Chapter 8

Nosofsky & Palmeri (1997) Experiment 2

Introduction

The conclusions of Ashby *et al.*'s (1994) study concerning the lack of familiarity effects on categorisation RT contradict the predictions of Nosofsky and Palmeri's EBRW model. As discussed in Chapter 3, the EBRW predicts that categorisation RT is affected by stimulus frequency because each stimulus presentation results in a separate memory trace and the probability of a fast exemplar retrieval time is a function of the number of stored exemplars. Nosofsky and Palmeri (1997, Experiment 2), therefore, set out to test the predictions of the EBRW and RT-D models by conducting an experiment in which the frequency of individual exemplars was manipulated while keeping the distance of exemplars from the supposed categorisation decision bound fixed. The result of the experiment is crucial in that it provides positive support for one of the models and negative support to the other, (if RT is affected by stimulus familiarity, the predictions of the EBRW are confirmed while the RT-D model's assumption that distance from decision bound is the single determinant of RT is challenged and vice versa).

Because the EGCM also predicts that stimulus frequency affects categorisation RT, Lamberts (submitted) has also used the data from Nosofsky and Palmeri's experiment to test the RT predictions of the EGCM in relation to stimulus frequency. In this chapter, I will outline the experiment and briefly summarise the outcome of the application of the EBRW and EGCM models to the data. I will then describe the application of the ALCOVE(RT) model to the same data. Apart form the manipulation of stimulus frequency, there is a further interesting difference between Nosofsky and Palmeri's Experiment 2 and Experiments 1 and 2 described above. Stimulus dimensions vary continuously in Nosofsky and Palmeri's experiment rather than being binary valued and have values determined by multidimensional scaling techniques applied to pairwise similarity judgement data. A consequence of this is that the category structure is less regular in respect to the position of stimuli and the decision boundary in psychological space. It is expected that these differences may result in a more flexible response from the ALCOVE(RT) model and in predictions which more accurately reflect relevant information in the data.

Details of the experiment

Nosofsky and Palmeri's investigation took the form of a standard classification response time experiment, consisting of a category learning stage followed by a speeded transfer stage. The category structure used in the experiment is illustrated in Figure 8.1.

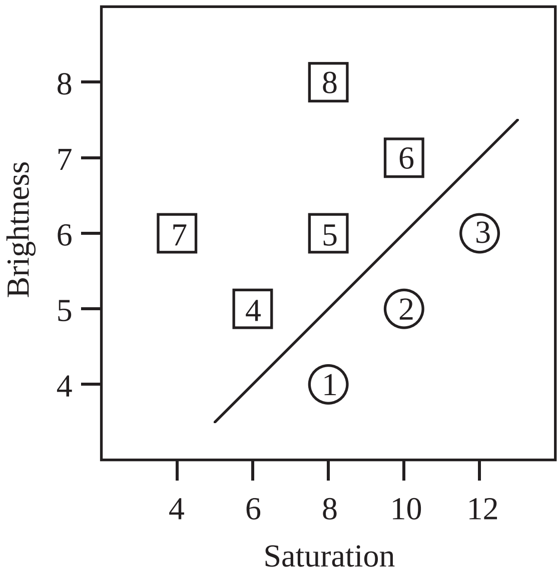

Figure 8.1: Schematic illustration of category structure used in Nosofsky & Palmeri (1997), Experiment 2. Stimuli represented by circles belong to category A while those represented by squares belong to category B. The diagonal line represents a linear decision bound.

The stimuli used in the experiment were eight colours varying in two dimensions, *saturation* and *brightness* (all stimuli had the same red hue). Stimuli 1–3 (represented by circles in Figure 8.1) belong to Category A while stimuli 4–8 (represented by squares in Figure 8.1) belong to category B. The frequency manipulations in the two conditions involved stimuli 7 and 8. In condition U7 (*unfamiliar-7*), stimulus 7 was not presented during the training stage whereas in condition U8, (*unfamiliar-8*), stimulus 8 was not presented during training. Stimuli 1–6 were presented with equal frequency in both conditions.

The category structure was designed on the assumption that an optimal decision boundary predicted by the RT-D model would fall in the region between stimuli 1, 2 and 3 and stimuli 4, 5 and 6 (as indicated in Figure 8.1) and that this would be the same whether stimuli 7 and 8 were presented during training or not. The assumption underlying the design of the structure is that stimuli 1 6 should play a large part in determining the position of an optimal decision bound because of their closeness to the presumed category boundary. Conversely, the influence of stimuli 7 and 8 on the position of the decision bound should be negligible because of their relative distance from the supposed boundary. One consequence of this design, according to Nosofsky and Palmeri, is that the RT-D hypothesis would predict no difference in categorisation RTs between stimuli 7 and 8 across the conditions. In contrast, the EBRW should predict that the frequency manipulation will result in a slower mean RT for stimulus 7 than for stimulus 8 in condition U7 and a slower mean RT for stimulus 8 than for stimulus 7 in condition U8.

The observed proportions of category A responses and mean RTs for each stimulus in both conditions are shown in Table 8.1. The data provide evidence for the claim that distance from decision bound affects categorisation RT because, in both conditions, stimuli 7 and 8—the furthest away from the boundary—are categorised more rapidly than the other stimuli (apart from stimulus 3). However, the results of the experiment support the predictions of the EBRW over those of the RT-D model in that stimulus 8 was, on average, categorised more rapidly than stimulus 7 in condition U7 but more slowly than stimulus 7 in condition U8. These results were confirmed by statistical analysis of the RT data for the two stimuli which showed a significant interaction between stimulus and condition.

Table 8.1: Proportions of Category A Responses (RP) and Mean Response Times (RT in ms) for Each Stimulus in Conditions U7 and U8 in Nosofsky and Palmeri (1997), Experiment 2

	Condition U7		Condition U8	
Stimulus	RP	RT	RP	RT
1	0.964	750	0.948	795
2	0.927	794	0.948	834
3	0.992	648	0.980	677
4	0.109	859	0.105	897
5	0.012	740	0.028	819
6	0.068	846	0.117	896
7	0.056	703	0.016	672
8	0.004	648	0.077	752

The rapid classification of stimulus 3 was explained by a similarity scaling study carried out on similarity ratings for pairs of stimuli provided by experiment participants after the category learning task. The MDS solution for the similarity ratings showed that stimulus 3 was generally considered to be further away from (i.e. less similar to) the other stimuli than expected. However, the optimal decision boundaries computed for each condition were virtually identical so that the RT-D model using the MDS co-ordinates would still not predict any effects of familiarity. Note that the stimulus dimension values yielded by the MDS analysis were used by Nosofsky and Palmeri in their subsequent modelling of the data and were also utilised by Lamberts (submitted) in his analysis. These values will also be used, therefore, in the test of the ALCOVE(RT) model reported below.

Model-based analysis

EBRW

Nosofsky and Palmeri applied the EBRW to the combined category A response proportion and RT data from both conditions of their experiment. The category A response proportions and RTs predicted by the model are displayed in Table 8.2 while the best-fitting parameter values for the model are shown in Table 8.3.

Table 8.2: Observed (Obs) and EBRW Predicted (Pred) Proportion of Category A Responses (RP) and Mean Response Times (RT) for Each Stimulus in Conditions U7 and U8, Nosofsky and Palmeri (1997), Experiment 2

	Condition U7				Condition U8			
	RP		RT		RP		RT	
Stimulus	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	0.964	0.963	750	776	0.948	0.931	795	832
2	0.927	0.947	794	807	0.948	0.914	834	855
3	0.992	0.996	648	655	0.980	0.991	677	683
4	0.109	0.065	859	856	0.105	0.075	897	872
5	0.012	0.028	740	778	0.028	0.041	819	808
6	0.068	0.042	846	812	0.117	0.069	896	862
7	0.056	0.006	703	705	0.016	0.001	672	649
8	0.004	0.001	648	639	0.077	0.014	752	741

In order to provide adequate fits to the data, Nosofsky and Palmeri were required to add a further parameter, δ, to the model to increase the sensitivity of stimulus 3. This was achieved by modifying the similarity function (Equation 2.4) to form $\eta_{ij} = \exp(-\delta \cdot d_{ij})$. The addition of this parameter has the effect of multiplying the distance between stimulus 3 and the other stimuli.

The EBRW provided a close fit to the RT data, accounting for 92.6% of the variance ($RSS = 7986$). A similar fit resulted from the application of the model to the RT data separately. Most importantly, the model predicts the crossover interaction between RTs for stimuli 7 and 8 across the conditions. The model was also able to provide as close a fit to the choice data, accounting for 99% of the variance ($RSS = 0.015$).

In fitting the model to the data, Nosofsky and Palmeri used the same parameter values for both conditions, except for the c parameter, which had separate values for each condition because mean RTs were shorter overall for condition U7 than for condition U8. Nosofsky and Palmeri's experiment builds upon the previous findings of Nosofsky (1988; see Chapter 7) that, contrary to the claims of Ashby *et al.* stimulus familiarity does affect categorisation performance. One potential uncertainty about the experiment recognised by Nosofsky and Palmeri and also noted by Lamberts, concerns whether the frequency effect found is partly a result of the element of *surprise* involved when unfamiliar stimuli are encountered. Because the unfamiliar stimulus in each condition was not presented at all during the training stage, the relative novelty of the stimulus in the transfer stage may

Table 8.3: Best-Fitting Parameter Values for EBRW, Nosofsky and Palmeri (1997), Experiment 2. Note. The value of the weight parameter for the *brightness* dimension (in brackets) is constrained by the value of the *saturation* dimension.

Parameter	Value
c	$c_{U7} = 1.396, c_{U8} = 1.582$
$w(saturation)$	0.508
$[w(brightness)]$	0.492
α	2.097
A	3
B	4
k	27.71
μ_R(ms)	354.56
δ	1.3

result in a longer time to process stimulus information. This question will be addressed further in the following chapter.

EGCM

In fitting the EGCM to the category A response proportion and RT data from Nosofsky and Palmeri's experiment, Lamberts also used single parameter values for both conditions except for the c parameter for which a value was estimated for each condition. The EGCM yielded a close fit to the category A response proportion data ($R^2 = .99, RSS = 0.024$) but provided a less accurate fit to the RT data ($R^2 = .85, RSS = 17152$) than the EBRW. The proportion of category A responses and mean RTs predicted by the EGCM for each stimulus in both conditions are displayed in Table 8.4. The best-fitting parameter values estimated for the model are shown in Table 8.5. The EGCM did predict that responses to stimuli 7 and 8 were on average faster than those to the other stimuli and also correctly predicted to crossover effect for mean RTs for stimuli 7 and 8 of the familiarity manipulation. As discussed in Chapter 3, the ability of the EGCM to account for familiarity effects is due to the ζ parameter (Equation 3.16) representing exemplar strength which is a function of presentation frequency.

Therefore, if a transfer stimulus is very similar to an exemplar which has a high strength value, the level of confidence for that transfer stimulus will also be high, resulting in a relatively short RT. Note, however, that the predicted accuracy levels and RTs for stimulus 3 do not closely match those found in the experiment.

ALCOVE(RT)

What behaviour can be expected of the ALCOVE(RT) model with respect to the frequency manipulation in Nosofsky and Palmeri's experiment? In terms of the crossover effect found for stimuli 7 and 8, *ceteris paribus*, the total exemplar unit activation for stimulus 7 in condition U7 can be expected to be smaller than that for stimulus 8 because the activation

Table 8.4: Observed (Obs) and EGCM Predicted (Pred) Proportion of Category *A* Responses (RP) and Mean Response Times (RT) for Each Stimulus in Conditions U7 and U8, Nosofsky and Palmeri (1997), Experiment 2

	Condition U7				Condition U8			
	RP		RT		RP		RT	
Stimulus	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	0.964	0.920	750	752	0.948	0.920	795	791
2	0.927	0.920	794	748	0.948	0.920	834	785
3	0.992	0.930	648	713	0.980	0.920	677	740
4	0.109	0.090	859	843	0.105	0.100	897	907
5	0.012	0.080	740	785	0.028	0.090	819	835
6	0.068	0.080	846	829	0.117	0.090	896	891
7	0.056	0.060	703	685	0.016	0.050	672	653
8	0.004	0.050	648	647	0.077	0.070	752	724

Table 8.5: Best-Fitting Parameter Values for EGCM, Nosofsky and Palmeri (1997), Experiment 2. Note. The value of the utility parameter for the *brightness* dimension (in brackets) is constrained by the value of the *saturation* dimension.

Parameter	Value
$q(saturation)$	9.44×10^{-4}
$q(brightness)$	2.55×10^{-4}
$u(saturation)$	0.357
$[u(brightness)]$	0.643
c	$c_{U7} = 11.032, c_{U8} = 11.851$
β	0.592
Ψ	0.500
θ	11.851
t_{res}	579.5

of the exemplar unit representing stimulus 8, on presentation of stimulus 7 will be will be less than when stimulus 8 in presented. This will result in lower activation value for the category *A* unit for stimulus 7 and a greater number of cascade cycles to reach a criterion value, (and therefore a longer predicted RT), for stimulus 7 than for stimulus 8 in condition U7. A similar situation will apply for stimulus 8 in condition U8.

For each condition, the ALCOVE(RT) model was trained on the seven training stimuli for thirty epochs and then presented with the complete eight transfer stimuli and allowed to cascade. Model parameters were estimated by maximising summed R^2 for category *A* response proportions and mean RTs. The choice proportions and RTs predicted by the model for each condition are shown in Table 8.6.

The fits of ALCOVE(RT) to the RT data are not close. Although the model for the U7 condition yielded an R^2 for the response proportions of .99 ($RSS = 0.005$), that for the RTs was only .022 ($RSS = 31552$). The model for the U8 condition also produced an R^2 for the response proportions of .99 ($RSS = 0.005$) whereas that for the RTs was equally low as the U7 model at .052 ($RSS = 50648$). The high R^2 values for the predicted response proportions were due to the small amount of variance in the observed data. The reasons for the poor fits to the RT data will be discussed in detail in the chapter summary below.

Table 8.6: Observed (Obs) and ALCOVE(RT) Predicted (Pred) Proportion of Category A Responses (RP) and Mean Response Times (RT) for Each Stimulus in Conditions U7 and U8, Nosofsky and Palmeri (1997), Experiment 2

	Condition U7				Condition U8			
	RP		RT		RP		RT	
Stimulus	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	0.964	0.954	750	761	0.948	0.948	795	812
2	0.927	0.937	794	742	0.948	0.942	834	802
3	0.992	0.967	648	751	0.980	0.944	677	802
4	0.109	0.004	859	742	0.105	0.059	897	802
5	0.012	0.051	740	742	0.028	0.058	819	792
6	0.068	0.039	846	771	0.117	0.102	896	812
7	0.056	0.034	703	761	0.016	0.031	672	792
8	0.004	0.948	648	742	0.077	0.055	752	802

It should be noted, however, that the crossover effect of the response proportions and RTs for stimuli 7 and 8 is generated by the model as expected. The model also correctly predicts a relatively high level of classification accuracy for stimuli 7 and 8 in both conditions and a high level of accuracy for stimulus 3 in condition U7. The best-fitting parameter values found by the model for each condition are displayed in Table 8.7.

Table 8.7: Best-Fitting Parameter Values for ALCOVE(RT), Nosofsky and Palmeri (1997), Experiment 2

	Condition	
Parameter	U7	U8
c	0.662	0.671
λ_w	0.048	0.042
λ_α	0.033	0.009
ϕ	13.247	13.034
τ_{hid}	0.138	0.139
τ_{out}	0.121	0.206
ϑ	0.100	0.083
υ	9.562	9.610
ω	703.645	763.651

It can be seen in Table 8.7 that the parameter values for the two conditions are very similar, suggesting that a similar fit to the data from both conditions may have been possible for a single model with one set of parameter values. However, as the original model fits were already poor, it was considered unnecessary to apply a such a version of the model to the data.

Discussion

Nosofsky and Palmeri's experiment provides clear evidence for the effect of stimulus familiarity on categorisation response times and lends support to the general claim that frequency information can be an important determinant of categorisation performance. The success of the EBRW and EGCM models in accounting for the response accuracy and RT data from this experiment, and in particular, to model the effect of the frequency manipulation for stimuli 7 and 8 also provides strong evidence for the requirement that formal models allow frequency information to be taken into account.

The inability of the RT-D model to account for the frequency effect in Nosofsky and Palmeri's experiment is a further (negative) indication of this requirement. Although overall fits to the data were much worse than the EBRW and EGCM, the ALCOVE(RT) model was also able to produce the frequency effect on the response times for stimuli 7 and 8. This behaviour was generated by the model because the rate of information accumulation for the category units in response to a previously unseen stimulus will be lower than for a familiar stimulus because the summed exemplar unit activation for an unfamiliar stimulus will typically be smaller than that for a familiar one due to the absence of a specific exemplar unit representing the unseen stimulus.

It was suggested in Chapter 5 that the differences between, on the one hand, the ALCOVE model and on the other, the EBRW and EGCM models, prevent strict comparisons between their ability to predict classification data from experiments such as the three reported above. It must be stressed that ALCOVE was designed primarily to account for the course of category learning and is not simply a model of post-training asymptotic categorisation performance, as is the case of the GCM, EGCM and EBRW models. Previous tests of ALCOVE have, in the main, focused on the category learning process and in particular the mechanisms for selective attention learning (e.g. Kruschke, 1992, 1993; Lewandowsky, 1995). However, one experiment carried out by Nosofsky, Kruschke and McKinley (1992) to test ALCOVE's ability to predict transfer data showed that the model was able to give a good account of classification accuracy after training. This success has not been borne out by Experiments 1 and 2 reported here because of, amongst other factors, the relative simplicity of the category structures used and the procedure of training to a predefined level of accuracy. In their experiment, Nosofsky *et al.* did not train participants to criterion but presented stimuli for a fixed number of times. Nosofsky *et al.* also employed a more complex category structure than those used in Experiments 1 and 2, having 9 training stimuli and a total of 16 transfer stimuli. The ability of the ALCOVE(RT) model to give accurate and detailed accounts for combined response accuracy and RT data is limited in comparison to the EGCM and EBRW models because of the relatively indirect effect that parameter values in ALCOVE(RT) have on model performance. Further constraints are placed upon dimension attention strengths and exemplar-category association weights by the connectionist learning rules which modify them during the course of training.

One potential uncertainty about the experiment recognised by Nosofsky and Palmeri and also noted by Lamberts, concerns whether the observed frequency effect is confounded by an element of surprise which may be involved when an unfamiliar stimulus is encountered. Because the unfamiliar stimulus in each condition was not presented at all during the training stage, the relative novelty of the stimulus in the transfer stage may result in participants taking a longer time to process stimulus information. The experiment reported in the following chapter attempts to address a number of the issues raised above.

Chapter 9

Experiment 3

Introduction

Experiment 3 is designed to allow further investigation into the effect of stimulus frequency on categorisation performance and, in so doing, address the issue concerning the possible confounding effect of stimulus novelty on performance which was raised in the analysis of Nosofsky and Palmeri's experiment. The experiment is also designed to provide data on the effects of stimulus frequency on classification accuracy and RT throughout the course of training by permitting the testing of participants at regular intervals in the training stage. In Experiments 1 and 2, the training scheme required participants to learn the category structure to a predetermined (and relatively high) level of accuracy. One consequence of this scheme was that there was very little variation in the accuracy of responses for different stimuli during the subsequent testing stage. The design of Experiment 3 allows possible differences in accuracy and RT which may exist in the initial stages of training to be observed and facilitates the observation of shifts in attention to different dimensions which may occur as the categories as learned.

One further motivation for the design of Experiment 3 is to test the ability of Kruschke's ALCOVE model to account for the learning rates for individual stimuli in respect to their presentation frequency. As ALCOVE was originally intended to account for the process of category learning, it may be used to analyse the effect of frequency manipulations on category response proportions for individual stimuli during the training stage.

Method

Participants

Twenty undergraduate and postgraduate students from the University of Birmingham participated in the experiment. The undergraduate psychology students who took part were given credit under the Psychology department's research participation scheme.

Apparatus and stimuli

The experiment was carried out on a Vale 486 IBM PC compatible computer with a Vale EC 33 cm SVGA colour monitor using a display mode with 640 pixels horizontally and 480 pixels vertically. Participant's responses were registered using two microswitches connected to the computer's parallel port. The stimuli used were drawings of faces which varied on four binary dimensions—*nose* (large or small), *eyes* (close together or far apart), *mouth* (wide or narrow), and *hair* (with hair or without hair). Two example stimuli showing the full range of dimension values are presented in Figure 9.1.

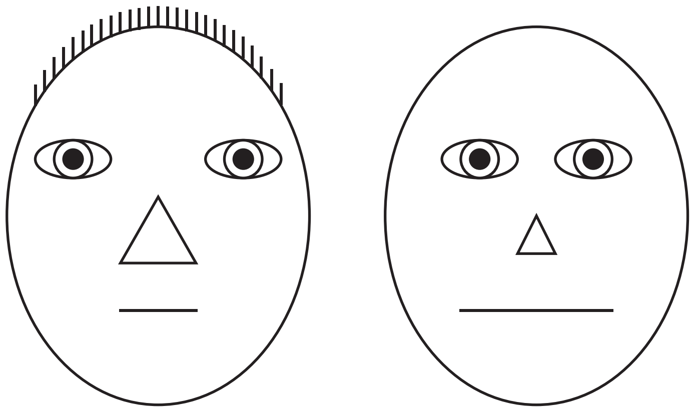

Figure 9.1: Sample stimuli in Experiment 3

Similar stimuli have previously been used in categorisation experiments (e.g. Reed, 1972; Lamberts, 1995). The structure of the eight stimuli used in Experiment 3 (shown in Table 9.1) was the same as that used in Experiment 1. This particular category structure was employed again in an to attempt to keep the similarity relationships between stimuli as equal as possible in order to minimise any effect on RTs due to variation in similarity. The same category structure in Experiment 1 demonstrated a certain degree of inflexibility in the predictions of the ALCOVE(RT) model. A further motive for using the same structure in Experiment 3, therefore, was to discover whether this inflexibility would be reduced by the frequency manipulation, and, if so, whether the model's predictions would be appropriate.

Design and Procedure

The experiment consisted of the presentation of all eight stimuli to the participants in alternating sets of training and testing blocks. In training blocks, participants were required to learn to categorise the stimuli according to the structure in Table 9.1. In testing blocks, the task was to categorise the stimuli as quickly as possible without sacrificing accuracy. Participant's category responses were recorded throughout the experiment and the time of their response (in milliseconds) was also recorded during testing blocks.

There were two conditions in the experiment with ten participants assigned at random to each. In the *equal frequency* (EF) condition, each training block consisted of the complete set of eight stimuli presented sequentially in random order. In the *unequal frequency* (UF) condition, the training blocks were the same as those in the EF condition except that one stimulus, (stimulus 2), was presented four times so that each training block contained a total of 11 trials presented sequentially in random order. Testing blocks were the same in both conditions.

Table 9.1: Structure of Stimuli in Experiment 3

Structure	Stimulus Number	Dimension: Nose	Eyes	Mouth	Hair
Category A	1	1	1	1	0
	2	1	1	0	1
	3	1	0	1	1
	4	0	1	1	1
Category B	5	0	0	0	1
	6	0	0	1	0
	7	0	1	0	0
	8	1	0	0	0

The design of Experiment 3 is intended to minimise a potential effect of stimulus repetition on RT in the UF condition. One possible drawback of an experiment in which frequency effects on RTs are monitored throughout the training stage is the possibility that the sequential repetition of the increased frequency stimulus affects the mean RT for that stimulus. In the current experiment, for example, a training block in the UF condition contains four copies of stimulus 2 randomly distributed amongst the other seven stimuli. The probability that a presentation of stimulus 2 is immediately followed by another presentation of the same stimulus, therefore, is relatively high and it is possible that this repetition will reduce the RT for the second stimulus presentation. If RT data from the training blocks were to be used, therefore, there would be the possibility that the average RT for stimulus 2 in the UF condition would be reduced significantly by this repetition. Therefore the data used in the analysis of this experiment is taken solely from the testing blocks in which all stimuli in both conditions are presented with equal frequency.

At the beginning of the experiment, participants were presented with an initial group of nine training blocks in order to ensure that they had learned the categories sufficiently before testing began. After this initial set of training blocks, participants were presented with alternating sets of four testing and training blocks until a total of 77 blocks, (41 training, 36 testing), had been presented. The total number of stimulus 2 training presentations in the EF condition, therefore, was 41 whereas in the UF condition it was 164.

On each training trial, a white fixation cross was placed at the centre of the blank computer screen for a period of 400 ms when it was removed from the screen. After 100 ms of blank screen, one of the stimuli would appear at the centre of the screen. When a participant responded, an auditory signal indicating whether the category judgement was correct or not was given for a period of 500 ms and the stimulus was removed from the screen. If the category response was correct, a 600 Hz (high) tone was given whereas if it was incorrect, a 100 Hz (low) tone was given. A 1500 ms interval separated each training trial. Test trials in both conditions were exactly the same as training trials except that no auditory feedback was given.

Results

To observe the changes in performance across the training session, participants' category and RT responses for each stimulus were averaged across each of the nine sets of four testing blocks to create nine *test points*. By the ninth point, participants in both conditions had learned the categories well. In the EF condition, the mean correct category response across all stimuli rose from 0.68 ($SD = 0.063$) in point 1 to 0.88 ($SD = 0.102$) in point 9. A similar increase was observed in the UF condition, from 0.70 ($SD = 0.083$) to 0.94 ($SD = 0.067$). The fact that the average scores for point 1 in both conditions were already well above 0.5 indicates that participants had learned the category structure to a certain extent after only nine presentations of each stimulus. Mean RTs also improved during the course of training. Average RTs across all stimuli in the EF condition fell from 1025 ms ($SD = 713$) in point 1 to 818 ms ($SD = 258$) in point 9. In the UF condition this decrease was from 1112 ms ($SD = 405$) to 930 ms ($SD = 298$). An analysis of variance on the mean RTs produced a significant effect of stimulus, $F(7, 126) = 7.65, p < .001, MSE = 150,229$ and a significant interaction between stimulus and condition, $F(7, 126) = 3.47, p < 0.005, MSE = 68,262$. An ANOVA on the mean correct category responses only produced a significant effect of stimulus type, $F(7, 126) = 5.85, p < .001, MSE = 0.143$.

A simplified, though coarser-grained picture of the development of categorisation performance with training was produced by further averaging the nine test points into three new *average test points*. Although the additional averaging process results in a certain loss of detail, it is not significant for this study because the frequency effect is still observable in the averaged data. The three new points were generated by grouping the original test points into sets of three so that the first new point was the average of test points 1, 2 and 3, the second was the average of test points 4, 5 and 6, etc. The resulting Category A response proportions and mean RTs for each stimulus in both conditions are shown in Tables 9.2 and 9.3 respectively.

It is reasonable to assume that attention to stimulus dimensions is liable to change during the course of category learning as participants become more attuned to the diagnosticity of the dimensions in relation to the category structure being learned and become less influenced by differences in dimensional salience. One may observe the overall distribution and trace the development of attention from the results displayed in Table 9.2. Across all three average test points in both conditions, stimuli 1, 5, 4 and 8 are in general misclassified more often than stimuli 2, 6, 3 and 7. This suggests that participants attended to the *hair* and *nose* dimensions more than to the *eyes* and *mouth* dimensions and that this remained the case throughout the experiment. However, as classification improves with training, one can infer that attention becomes more evenly distributed between the dimensions and that overall discrimination increases. A further indication of this redistribution of attention is observed across all the stimuli of the EF condition and in stimulus 2 in the UF condition. In average test point 3, the classification accuracy of stimuli 1, 5, 4 and 8, which had initially been most misclassified, continued to improve whereas that for the initially well classified stimuli dropped slightly.

The effect of the frequency manipulation in the experiment is easily observed in the response proportions for stimulus 2 in both conditions. Whereas in the EF condition, the response proportions rose from 0.833 in average test point 1 to 0.942 in average test point 2 to end slightly lower at 0.908 in average test point 3, those for the UF condition were already close to perfect accuracy with 0.983 in average test point 1, remained at that level at average test point 2 and then also dipped slightly to 0.975 in average test point 3.

Table 9.2: Observed Proportions of Category A Responses for Each Stimulus as a Function of Average Test Point (APnt) in Experiment 3

	EF Condition			UF Condition		
Stimulus	APnt 1	APnt 2	APnt 3	APnt 1	APnt 2	APnt 3
1 1110 (A)	0.467	0.683	0.742	0.683	0.725	0.900
2 1101 (A)	0.833	0.942	0.908	0.983	0.983	0.975
3 1011 (A)	0.858	0.942	0.933	0.783	0.825	0.942
4 0111 (A)	0.742	0.833	0.875	0.642	0.592	0.875
5 0001 (B)	0.475	0.183	0.267	0.392	0.225	0.092
6 0010 (B)	0.150	0.142	0.158	0.200	0.117	0.108
7 0100 (B)	0.183	0.068	0.067	0.342	0.117	0.017
8 1000 (B)	0.267	0.150	0.175	0.308	0.300	0.192

One can also see the effect of frequency on stimulus 2 response proportions relative to the other stimuli in the same condition. In the EF condition, classification accuracy for stimulus 2 is not significantly better than that for stimuli 3, 6 and 7 in all three average test points. In the UF condition, however, classification accuracy for stimulus 2 is 20% higher than the next best classified stimulus in average test point 1 and 10% higher in average test point 2. By average test point 3, however, overall accuracy had increased to a point where there was no significant difference between stimulus 2 and the others.

The effect of the frequency manipulation is also observable in the RT data presented in Table 9.3. In the EF condition, although there is a decrease in mean RT of approximately 100 ms for stimulus 2 at each successive average test point, the values in points 1 and 2 are not at the extreme end of the range. In average test point 3, the mean RT for stimulus 2 does become the shortest, however, but the difference between this value and the next fastest is only 71 ms. Mean RTs for stimulus 2 at all three average test points in the UF condition are short relative to those in the EF condition. At point 1, the mean RT is already 49 ms shorter than the average RT for stimulus 2 at average test point 2 in the EF condition. This falls by 117 ms at average test point 2 and then rises slightly by 20 ms in average test point 3. In contrast to mean RT values of stimulus 2 in the EF condition, those for stimulus 2 in the UF condition are on average 248 ms faster than the next fastest value in each average test point.

Table 9.3: Observed Mean Response Times (in ms) for Each Stimulus as a Function of Average Test Point (APnt) in Experiment 3

	EF Condition			UF Condition		
Stimulus	APnt 1	APnt 2	APnt 3	APnt 1	APnt 2	APnt 3
1 1110 (A)	991	915	828	1232	1121	1044
2 1101 (A)	916	819	716	770	653	673
3 1011 (A)	972	895	810	1002	932	980
4 0111 (A)	865	875	841	1171	1092	1041
5 0001 (B)	943	784	827	1096	985	907
6 0010 (B)	974	1029	918	1112	1058	1048
7 0100 (B)	921	912	824	1277	1062	980
8 1000 (B)	1022	891	787	1159	1081	960

In summary, analysis of the data from Experiment 3 shows that manipulating the frequency of presentation during training can have a significant effect on subsequent categorisation response time. This was demonstrated by the results which showed that participants were significantly faster when categorising stimulus 2 in the UF condition than in the EF condition.

Model-based analysis

EGCM

The EGCM was applied jointly to the response proportions and mean RTs from the three average test points in both conditions as in Experiments 1 and 2. Table 9.4 and Table 9.5 show the observed and predicted category A response proportions and mean RTs for the EF and UF conditions respectively.

Table 9.4: Observed (Obs) and EGCM Predicted (Pred) Proportions of Category A Responses (RP) and Response Times (RT) for Each Stimulus as a Function of Average Test Point (APnt) in EF Condition, Experiment 3

	APnt 1				APnt 2				APnt 3			
	RP		RT		RP		RT		RP		RT	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.467	.750	991	944	.683	.832	915	854	.742	.865	828	808
2	.833	.747	916	949	.942	.827	819	867	.908	.860	716	817
3	.858	.749	972	949	.942	.831	895	864	.933	.864	810	819
4	.742	.751	865	944	.833	.834	875	851	.875	.867	841	807
5	.475	.274	943	965	.183	.192	784	892	.267	.154	827	838
6	.150	.277	974	968	.142	.199	1029	904	.158	.163	918	855
7	.183	.274	921	968	.068	.194	912	901	.067	.156	824	849
8	.267	.273	1022	964	.150	.190	891	890	.175	.151	787	833

The model was able to provide a reasonable fit to the data from the EF condition, accounting for 89.6% of the variance in the response proportions (RSS = 0.277) and 52.2% of the variance in the RTs (RSS = 69715). Note that the homogeneity of the category structure and the lack of any other complicating factor (such as variation in presentation frequency) results in a rather rigid pattern of accuracy and RT predictions produced by the EGCM.

The fit to the data from the UF condition was even closer than for EF condition, the model accounting for 92.8% (RSS = 0.195) and 73.4% (RSS = 143083) of the variance in the response proportions and RTs respectively. The EGCM captures the differences in classification accuracy and mean RT for stimulus 2 across the conditions, although the model's predictions for the UF condition are consistently less accurate and slower over the three average test points than the observed values. The model also fails to predict the rapid RT for stimulus 5 in average test points 2 and 3 and the high level of accuracy for the same stimulus in average test point 3.

The best-fitting parameter values estimated for the model are presented in Table 9.6. The parameter values are very similar between conditions. In particular, the EGCM predicts that stimulus dimensions are equally salient.

Table 9.5: Observed (Obs) and EGCM Predicted (Pred) Proportions of Category *A* Responses (RP) and Response Times (RT) for Each Stimulus as a Function of Average Test Point (APnt) in UF Condition, Experiment 3

	APnt 1				APnt 2				APnt 3			
	RP		RT		RP		RT		RP		RT	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.683	.772	1232	1133	.725	.843	1121	1071	.900	.873	1044	1039
2	.983	.884	770	849	.983	.915	653	746	.975	.925	673	736
3	.783	.789	1002	1004	.825	.858	932	909	.942	.881	980	884
4	.642	.762	1171	1100	.592	.831	1092	1033	.875	.859	1041	1006
5	.392	.310	1096	1168	.225	.257	985	1106	.092	.229	907	1060
6	.200	.253	1112	1145	.117	.162	1058	1005	.108	.118	1048	910
7	.342	.296	1277	1164	.117	.225	1062	1070	.017	.190	980	1004
8	.308	.314	1159	1174	.300	.257	1081	1112	.192	.227	960	1067

One interesting difference between the parameter estimates between the two conditions, however, is the utility value, u, predicted for the *mouth* dimension. In the EF condition, the utility values for the dimensions are very close with $u(mouth)$ being given a slightly higher value than the other three. In the UF condition, however, the model predicts that the *mouth* dimension is insignificant in the computation of stimulus-exemplar similarity. It is likely that this difference is due to the frequency manipulation of stimulus 2 because it is on this dimension that stimulus 2 differs from the other stimuli in category *A*. To consistently categorise stimuli 2 and 6 accurately in the EF condition, it is necessary for participants to attend to the *mouth* dimension to a certain extent.

The increased exposure to stimulus 2 in the UF condition, however, makes this dimension less important because participants in this condition can learn to make correct category decisions based on a configuration of the other three dimensions. This distribution of attention affects the accuracy of responses to stimulus 6 because this stimulus also has a value on the *mouth* dimension different from the other stimuli in its category. Therefore, stimulus 6 should be more accurately classified than the other stimuli. This is actually what happens in average test point 1 where stimulus 6 is the most accurately classified category B stimulus. In average test points 2 and 3, however, this difference is reduced as the accuracy of the responses to the other stimuli improve, probably as a result of the development of a more even distribution of attention across the dimensions.

Table 9.6: Best-Fitting Parameter Values for EGCM, Experiment 3. Note. The value of the utility parameter for the *hair* dimension (in brackets) is constrained by the values of the other three.

	Condition	
Parameter	EF	UF
$q(nose)$	0.005	0.005
$q(eyes)$	0.005	0.003
$q(mouth)$	0.005	0.002
$q(hair)$	0.005	0.007
γ	3.708	3.722
Ψ	0.388	0.447
t_{res}(ms)	590	594
$u(nose)$	0.215	0.388
$u(eyes)$	0.235	0.381
$u(mouth)$	0.315	0.000
$[u(hair)]$	0.235	0.232
c	12.739	14.522
θ	5.84	4.99
β	0.537	0.516

EBRW

The EBRW was also applied to the response proportions and mean RTs from the three average test points in the two conditions using the same method as was employed in Experiments 1 and 2. The predicted category A response proportions and mean RTs for the EF condition are shown in Table 9.7 while those for the UF condition are displayed in Table 9.8. The model was also able to give a reasonable fit to the data from the EF condition, accounting for 90% of the variance in the response proportions ($RSS = 0.269$) and 46.6% of the variance in the RTs ($RSS = 77847$). The fit of the EBRW to the data from the UF condition was less close than that of the EGCM. The model was able to account for 76% ($RSS = 0.649$) and 58.1% ($RSS = 225409$) of the variance in the response proportions and RTs respectively. A second version of the model with a separate c parameter for each average test point was also applied to the data. The extra parameters in this model did not result in a significant increase in fit, however, the model accounting for 92.2% ($RSS = 0.209$) of the variance in the response proportions and 49.6% ($RSS = 73472$) of the variance in the RTs for the EF condition and for 78.8%

($RSS = 0.573$) of the variance in the response proportions and 64.7% ($RSS = 189847$) of the variance in the RTs for the UF condition.

Table 9.7: Observed (Obs) and EBRW Predicted (Pred) Proportions of Category *A* Responses (RP) and Response Times (RT) for Each Stimulus as a Function of Average Test Point (APnt) in EF Condition, Experiment 3

	APnt 1				APnt 2				APnt 3			
	RP		RT		RP		RT		RP		RT	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.467	.758	991	970	.683	.758	915	875	.742	.758	828	841
2	.833	.924	916	927	.942	.924	819	852	.908	.924	716	826
3	.858	.930	972	924	.942	.930	895	851	.933	.930	810	825
4	.742	.909	865	933	.833	.909	875	855	.875	.909	841	828
5	.475	.426	943	990	.183	.426	784	985	.267	.426	827	848
6	.150	.192	974	968	.142	.192	1029	873	.158	.192	918	840
7	.183	.181	921	966	.068	.181	912	872	.067	.181	824	839
8	.267	.216	1022	972	.150	.216	891	875	.175	.216	787	842

One reason for the relatively low R^2 values produced by the EGCM and EBRW models for the EF condition is the relatively small amount of variance in the data from that condition. In fact, if one compares the sum of the squared deviations from the mean for the observed response proportions and RTs from the two conditions, it becomes apparent that, although the response proportions from the two conditions have approximately similar variance about the mean, the RTs from the EF condition are generally far closer to the mean that are those from the UF condition. In the computation of the R^2 function, this results in a relatively small value for the summed squared deviations and a correspondingly small value of R^2. The EBRW does capture the essential effect of the frequency manipulation, predicting shorter RTs for stimulus 2 than for all of the other stimuli in all three average test points. Note, however, that the model also predicts high levels of accuracy for stimuli 3, 6 and 7.

Table 9.8: Observed (Obs) and EBRW Predicted (Pred) Proportions of Category *A* Responses (RP) and Response Times (RT) for Each Stimulus as a Function of Average Test Point (APnt) in UF Condition, Experiment 3

	APnt 1				APnt 2				APnt 3			
	RP		RT		RP		RT		RP		RT	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.683	.884	1232	1141	.725	.884	1121	998	.900	.884	1044	948
2	.983	1.00	770	901	.983	1.00	653	873	.975	1.00	673	863
3	.783	1.00	1002	937	.825	1.00	932	892	.942	1.00	980	876
4	.642	.884	1171	1141	.592	.884	1092	998	.875	.884	1041	948
5	.392	.385	1096	1232	.225	.385	985	1045	.092	.385	907	879
6	.200	.000	1112	1204	.117	.000	1058	1031	.108	0.00	1048	970
7	.342	.000	1277	1204	.117	.000	1062	1031	.017	0.00	980	970
8	.308	.384	1159	1232	.300	.384	1081	1045	.192	.385	960	979

The reason for this can be seen in Table 9.9 which shows that the model predicts that both the *mouth* and *eyes* dimensions are given very little attention compared to the *nose* and *hair* dimensions.

Table 9.9: Best-Fitting Parameter Values for EBRW, Experiment 3. Note. The value of the weight parameter for the *hair* dimension (in brackets) is constrained by the values of the other three.

	Condition	
Parameter	EF	UF
$w(nose)$	0.217	0.409
$w(eyes)$	0.175	0.000
$w(mouth)$	0.188	0.038
$[w(hair)]$	0.419	0.553
c	0.419	21.760
α	0.419	0.000
A	2	1
B	3	2
k	1368.985	3373.533
μ_R(ms)	771	843

Also note that, in both conditions, the EBRW predicts generally faster RTs for the category A stimuli than for the category B stimuli. This is reflected in the values for the parameters representing the A and B decision criteria.

ALCOVE(RT)

Learning curves

To examine the ability of ALCOVE to account for the effect of the frequency manipulation on the learning curves for stimulus 2, the model was applied to the category A response proportion data from the two conditions. The increased presentation frequency of stimulus 2 in the UF condition is expected to result in a more rapid increase in the accuracy of the models' classification for that stimulus during training because the association weights between the exemplar unit representing stimulus 2 and the category units will be adjusted four times as often for stimulus 2 than for the other stimuli.

Four model parameters were estimated in fitting the model to the data—the hidden unit specificity parameter, c, the association and attention learning rate parameters, λ_w and λ_α, and the category unit response mapping parameter, ϕ. As before, R^2 was used as a measure of goodness-of-fit. In applying the model to the data from each condition, the network was trained for thirty epochs on all eight stimuli using batch updating. In the UF condition, an extra three copies of stimulus 2 were presented in each epoch. The predicted response proportions for each stimulus after ten, twenty and thirty epochs were taken as being the network output corresponding to the observed proportions at average test points 1, 2 and 3 respectively.

Although the model accounted for 88.7% (RSS = 0.301) and 93% (0.189) of the variance in the response proportion data for the EF and UF conditions respectively, it was unable, in the main, to provide accurate fits to the values for individual stimuli. The predicted category A response proportions produced by the model for each stimulus in the two conditions are shown in Table 9.10 while the best-fitting parameter values produced by the model are displayed in Table 9.11.

As with Experiments 1 and 2, the predictions of ALCOVE(RT) were generally restricted to a single average (or close to average) response proportion for four stimuli in a category. However, the frequency manipulation for stimulus 2 did have an effect on the predicted response proportion for that stimulus and also for stimulus 6 as both stimuli are predicted to be more accurately classified than the others for each average test point.

Table 9.10: Observed (Obs) and ALCOVE Predicted (Pred) Proportions of Category A Responses for Each Stimulus as a Function of Average Test Point (APnt) for Equal Frequency (EF) and Unequal Frequency (UF) Conditions, Experiment 3

	EF Condition						UF Condition					
	APnt 1		APnt 2		APnt 3		APnt 1		APnt 2		APnt 3	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.467	.703	.683	.846	.742	.918	.683	.749	.725	.830	.900	.854
2	.833	.703	.942	.846	.908	.918	.983	.937	.983	.985	.975	.991
3	.858	.703	.942	.846	.933	.918	.783	.749	.825	.830	.942	.854
4	.742	.703	.833	.846	.875	.918	.642	.749	.592	.830	.875	.854
5	.475	.297	.183	.154	.267	.082	.392	.415	.225	.280	.092	.181
6	.150	.297	.142	.154	.158	.082	.200	.224	.117	.067	.108	.027
7	.183	.297	.068	.154	.067	.082	.342	.415	.117	.280	.017	.181
8	.267	.297	.150	.154	.175	.082	.308	.415	.300	.280	.192	.181

The final distribution of attention produced by the model for each condition also follows the expected pattern. In the EF condition, all dimensions are given the same attention weight of 0.605. In the UF condition, however, the *nose*, *eyes* and *hair* dimensions are given the value of 0.608 while the *mouth* dimension has a value of 0.0.

Table 9.11: Best-Fitting Parameter Values for ALCOVE, Experiment 3

	Condition	
Parameter	EF	UF
c	2.960	2.637
λ_α	0.017	0.019
λ_w	0.472	0.973
ϕ	3.106	2.361

To get a more accurate picture of the observed and predicted learning curves for stimulus 2 in the two conditions, the category A response probability predicted for that stimulus by ALCOVE after each training epoch were recorded and plotted against the observed values for each of the nine original test points. These plots are displayed in Figure 9.2. In order to map the thirty predicted values onto the nine observed values and to take the relatively high level of observed response accuracy (due to the initial nine training blocks) into account, the first twelve predicted values were omitted from the plot and the remaining eighteen values were averaged to form nine plot points. In Figure 9.2 therefore, the predicted value for the first point on the x axis corresponds to the average of the model's predicted values after epochs thirteen and fourteen.

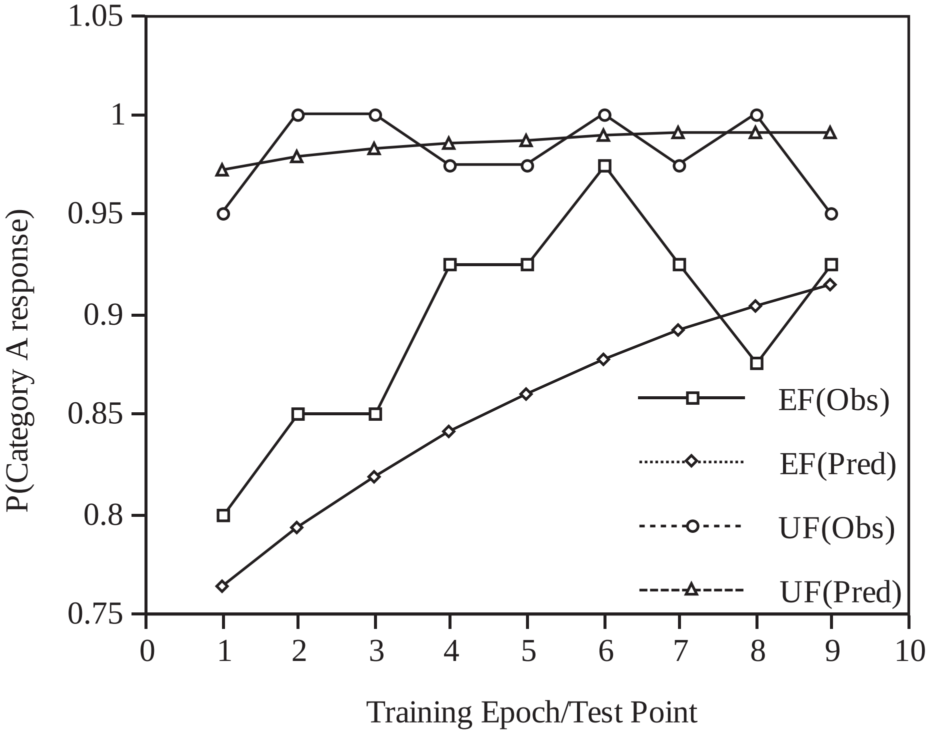

Figure 9.2: Plots of observed (Obs) and ALCOVE predicted (Pred) proportions of category A responses for stimulus 2 in equal frequency (EF) and unequal frequency (UF) conditions during training, Experiment 3

As can be seen in Figure 9.2, the learning curves predicted by ALCOVE do compare favourably with the observed curves. Note that the minimum value on the y axis is 0.5. Both curves from the EF condition start at values close to 0.8 and gradually increase to approximately 0.9 while those from the UF condition both start around a value of 0.95 and remain relatively high throughout the course of training. The relative irregularity of the observed curves is most likely to be a result of a certain degree of noise in the data.

Transfer data

Given the demonstrated inability of the ALCOVE model to provide anything but the most general predictions for the complete set of response proportions using a single set of parameter values, the ability of the ALCOVE(RT) model to give a good account of the combined response proportions and RTs using a similar number of parameters, as was achieved with the EGCM and EBRW models, must be doubted. From past performance in

Experiments 1 and 2, it is to be expected that, given the nature of the category structure and lack of complicating factors in the EF condition, the ALCOVE(RT) will simply predict an average response probability for each category and an average RT for all stimuli in every average test point. In the UF condition, however, the higher presentation frequency of stimulus 2 is likely to affect the predicted RT for (at least) that stimulus. The more limited goal of applying the ALCOVE(RT) model to the combined response accuracy and RT data, therefore, is to determine the nature and extent of the frequency effect on the model predictions in the UF condition.

In an initial application of the model to the data from both conditions, the categorisation specific parameters were set and fixed to the best-fitting values found during the earlier application of the ALCOVE model, while the RT-specific parameters were allowed to vary in the optimisation procedure. While this was sufficient to provide an optimal (for ALCOVE(RT)) fit to the EF data, the model was unable to account for the RT data in the UF condition. Therefore, a version of the model was applied in which all parameters were allowed to vary for each of the average test points in both conditions in the expectation that the resulting predicted values and best-fitting parameter values for each model version will provide some insight into the reasons why a single model version was unable to fit the data.

The predicted category A response proportions and RTs for each stimulus in the EF condition are shown in Table 9.12 while those for the UF condition are shown in Table 9.13. The best-fitting parameter values for the models are displayed in Table 9.14.

Table 9.12: Observed (Obs) and ALCOVE(RT) Predicted (Pred) Proportions of Category A Responses (RP) and Response Times (RT) for Each Stimulus as a Function of Average Test Point (APnt) in EF Condition, Experiment 3

	APnt 1				APnt 2				APnt 3			
	RP		RT		RP		RT		RP		RT	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.467	.728	991	951	.683	.858	915	890	.742	.849	828	819
2	.833	.728	916	951	.942	.858	819	890	.908	.849	716	819
3	.858	.728	972	951	.942	.858	895	890	.933	.849	810	819
4	.742	.728	865	951	.833	.858	875	890	.875	.849	841	819
5	.475	.272	943	951	.183	.142	784	890	.267	.151	827	819
6	.150	.272	974	951	.142	.142	1029	890	.158	.151	918	819
7	.183	.272	921	951	.068	.142	912	890	.067	.151	824	819
8	.267	.272	1022	951	.150	.142	891	890	.175	.151	787	819

Although the model for the EF condition yielded high R^2 values for the response proportions ($R^2APnt1 = .722$, $RSS = 0.160$; $R^2APnt2 = .950$, $RSS = 0.053$; $R^2APnt3 = .957$, $RSS = 0.044$) it did, as expected, predict an average response proportion for each category and an average RT for all eight stimuli in each average test point ($R^2APnt1 = 0.0$, $RSS = 17196$; $R^2APnt2 = 0.0$, $RSS = 36958$; $R^2APnt3 = 0.0$, $RSS = 22169$).

If one compares the categorisation-specific parameter values for the EF condition model in Table 9.14 with those produced by the previous ALCOVE model, shown in Table 9.9, it is clear that all of the values are both very similar to the original values and similar to each other across average test points. Greater differences are apparent in the RT-specific parameter values across average test points, however.

Table 9.13: Observed (Obs) and ALCOVE(RT) Predicted (Pred) Proportions of Category *A* Responses (RP) and Response Times (RT) for Each Stimulus as a Function of Average Test Point (APnt) in UF Condition, Experiment 3

	APnt 1				APnt 2				APnt 3			
	RP		RT		RP		RT		RP		RT	
Stim	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	.683	.728	1232	1060	.725	.787	1121	952	.900	.912	1044	988
2	.983	.818	770	943	.983	.884	653	657	.975	.975	673	755
3	.783	.728	1002	1060	.825	.787	932	952	.942	.912	980	988
4	.642	.728	1171	1060	.592	.787	1092	952	.875	.912	1041	988
5	.392	.328	1096	1178	.225	.233	985	1099	.092	.107	907	988
6	.200	.285	1112	1060	.117	.217	1058	952	.108	.104	1048	988
7	.342	.328	1277	1178	.117	.233	1062	1099	.017	.107	980	988
8	.308	.328	1159	1178	.300	.233	1081	1099	.192	.107	960	988

The model for the UF condition also accounted for a large proportion of the variance in the response accuracy data ($R^2 APnt1 = .900, RSS = 0.052$; $R^2 APnt2 = .899, RSS = 0.081$; $R^2 APnt3 = .987, RSS = 0.018$). The pattern of predicted response proportions is similar throughout the training stage. In each average test point, the model predicts a lower average level of accuracy for category *B* stimuli than for the category *A* stimuli and a higher level of accuracy for stimuli 2 and 6 than that for the other stimuli in their respective categories. The higher predicted accuracy for the category *A* stimuli is due to the increased strength of the positive association between the stimulus 2 exemplar unit and the category *A* unit and the corresponding strength of the negative association between the exemplar 2 unit and the category *B* unit. When a category A stimulus is presented to the network, the category *A* exemplars will be highly activated relative to the category *B* exemplars. The strength of the association between the exemplar 2 unit and the category unit will increase the net input to the category *A* unit, increasing its activation (and the associated probability of a category *A* response) and decrease the net input to the category *B* unit, reducing its activation.

The frequency manipulation also affects ALCOVE(RT)'s predicted response times and a certain amount of variance in the RT data is accounted for by the model ($R^2 APnt1 = 0.463, RSS = 94788$; $R^2 APnt2 = 0.540, RSS = 74502$; $R^2 APnt3 = 0.778, RSS = 23742$). In general the pattern of predictions follows that of the response proportions.

In average test points 1 and 2, the model predicts that stimuli 2 and 6 have a faster mean RT than the other stimuli in their respective categories. In average test point 3, however, the model predicts that the RT for stimulus 6 is the same as that for the other stimuli with the same presentation frequency. This follows the pattern of the observed values in that the response proportions and RTs become closer together as training progresses. In the first two average test points, the model also predicts a faster average RT for the category *A* stimuli than for the category *B* stimuli, for the same reasons given above. It is interesting to note that the values of the input unit-exemplar unit attention strengths

Table 9.14: Best-Fitting Parameter Values for ALCOVE(RT), Experiment 3

	Condition					
	EF			UF		
Parameter	APnt 1	APnt 2	APnt 3	APnt 1	APnt 2	APnt 3
c	3.006	3.000	2.705	5.086	6.129	9.130
λ_w	0.017	0.030	0.012	0.008	0.008	0.010
λ_α	0.483	0.485	0.392	1.002	0.997	0.983
ϕ	3.150	3.148	2.855	0.659	0.595	0.717
τ_{hid}	1.273	1.060	0.643	2.039	1.941	1.960
τ_{out}	0.821	0.595	0.607	1.909	1.847	1.877
ϑ	0.112	0.203	0.302	0.500	1.000	1.303
υ	544.926	427.966	115.993	58.865	73.582	116.390
ω	405.574	462.034	122.917	883.684	583.916	638.932

in the EF condition are all given the same value whereas those in the UF condition are given the same value apart from that for the *mouth* dimension which is given a value of 0.0 in each of the average test points.

In the EF condition, the best-fitting value for the c parameter is relatively constant between the average test points. In the UF condition, however, the estimated value of c increases during the course of training, suggesting that, as may be expected, the overall level of discrimination increases with category learning. In terms of the ALCOVE(RT) model, this development of overall discrimination would be represented by a gradual increase in the specificity of the exemplar units during the course of category learning. Although Kruschke (1992) does discuss the possibility of introducing local attentional effects by adapting exemplar unit specificities using gradient descent on error in the context of accounting for asymmetric similarity data (e.g. Rosch, 1975; Tversky, 1977), a specific mechanism for the global development of discrimination during training has yet to be proposed.

Discussion

The most important result of Experiment 3 was the clear demonstration that differences in stimulus frequency during category learning can have a marked effect on both categorisation accuracy and response times. Significant differences in the speed of responses to stimuli were observed between the two conditions. Differences between the response accuracy and mean RT for stimulus 2 as a result of the frequency manipulation between the conditions were discovered and the general form of the differences in response accuracy over the training period were captured by the ALCOVE model. It was also found that the differences in accuracy levels and RTs between the increased frequency stimulus and the other stimuli in the UF condition decreased over the course of training.

All of the models predicted that in the UF condition, participants' attention is withdrawn from the dimension upon which the increased frequency stimulus differs from the other stimuli in its category. One consequence of this is that stimuli in alternative categories which differ on that dimension from the other stimuli in that category, will also benefit from this withdrawal of attention and will therefore be more accurately responded to, at least during the initial stage of training. This was what was actually observed in average test points 1 and 2 in the UF condition.

The frequency effect discovered in Experiment 3 suggests that the findings of Nosofsky and Palmeri's Experiment 2 were unlikely to be solely the result of an effect of surprise, although one cannot conclude that the element of surprise did not have some confounding effect in their experiment.

Experiment 3 also had a number of shortcomings. Firstly, the recording of participant's responses only started after nine training blocks had been completed, by which time a considerable difference in participants' responses between the conditions was already apparent. Secondly, the effect of the increased presentation frequency for stimulus 2 in the UF condition was not large enough to modify the similarity structure to any significant extent, other than for stimuli 2 and 6. One consequence of this limited effect is that the regularity of the category structure remained the largest determinant of the predictions of the ALCOVE(RT) model. This situation may possibly be remedied by increasing the frequency of the high frequency and/or making the category more complex.

In the following chapter, an experiment is described which seeks to investigate further the effect of stimulus frequency on categorisation accuracy and RT both during the period of training and in a subsequent testing stage while attempting to address some of the shortcomings just outlined.

Chapter 10

Experiment 4

Introduction

Experiment 4 is designed to allow further investigation into the effect of stimulus frequency on categorisation performance and in so doing, address several issues which have been raised by the analysis of the previous three experiments and Nosofsky and Palmeri's Experiment 2. Firstly, in Experiments 1 and 2, the differences found in observed response times to stimuli in relatively homogeneous category structures led to the hypothesis that the salience of particular dimensions or features may play a significant role in the determination of participants' performance. In terms of the current investigation, this salience effect can only be regarded as an unwelcome complication because of its ability to obscure and/or distort any effects due to the manipulation of stimulus frequency. To minimise the possibility of this confounding effect in Experiment 4, therefore, a category structure was devised in which each category contained the same number of values for each of the dimensions. Such a category structure makes it likely that both categories are similarly affected by any variations in feature salience. In Experiments 1, 2 and 3, the relative simplicity of the category structures also placed a severe constraint on the ability of the ALCOVE(RT) model to account for the combined response accuracy and RT data. The category structure employed in Experiment 4, therefore, is a great deal more complex in terms of the similarity structure, the number of stimuli used, and also because the stimuli vary on five dimensions rather than four.

Experiment 4 also attempts to build upon the results of Experiment 3 by investigating in greater detail the effect of stimulus frequency on classification accuracy and RT in both the training and transfer stages. The success of the ALCOVE model in reproducing the observed learning curves for the manipulated frequency stimulus in Experiment 3 suggests that a further test of this ability is required. In addition, the recording of participant's responses in Experiment 3 only commenced after nine training blocks had been completed, by which time a considerable difference in participants' responses between the conditions was already apparent. The design of Experiment 4, therefore, enables the monitoring of the accuracy and rapidity of participants' responses throughout the entire course of the training stage which will provide a more detailed picture of the time course of category learning in relation to stimulus frequency and allow a more thorough test of the ALCOVE model. The design of Experiment 4 also differs from that of Experiment 3 in that it does not consist of alternating sets of training and testing blocks but has a more standard structure of a training stage followed by a testing stage. Therefore the problem of stimulus repetition affecting RT for an increased frequency stimulus during training (which the design of Experiment 3 was intended to combat), does have the potential to occur. However, it is not considered pertinent to the main analysis of this experiment because, although a qualitative comparison between RTs for the frequency manipulated stimulus during the training stage is undertaken, RT data will only be modelled from the testing stage.

Method

Participants

Fifty-one undergraduate and postgraduate psychology students from the University of Birmingham participated in the experiment. The undergraduate students who took part were given credit towards the Psychology department's research participation scheme.

Apparatus and stimuli

The experiment was carried out on an Elonex PC-433 computer with a Vale EC 32 cm SVGA colour monitor, using a display mode with 640 pixels horizontally and 480 pixels vertically. Participants' responses were recorded using two microswitches connected to the computer's parallel port. The stimuli used were drawings of flags which varied on five binary dimensions—*border* (square or pointed), *background* shade (light or dark), centre *symbol* (cross or star), *shape* surrounding the central symbol (circle or square), and the *colour* shared by the centre symbol and surrounding shape (black or white). Two example stimuli showing the full range of dimension values are shown in Figure 10.1. Twelve stimuli were used in the experiment, six per category. The equal distribution of values across dimensions between the categories ensures that any differences in feature salience in a dimension affects the categories equally. This constraint produced a structure in which both categories were composed of three pairs of stimuli, each stimulus in a pair being the mirror image, (i.e. containing the opposite value on every dimension) of the other.

The category structure used in the experiment is shown in Table 10.1. It was envisaged that a potential problem of using such a category structure composed of stimuli varying on five dimensions would be that it may prove to be too difficult for participants to learn. One potential benefit in terms of modelling the behaviour could be expected from the use of this structure, however, in that such a structure could be seen to promote exemplar storage by making the formulation of simple category decision rules more difficult.

Design and procedure

The experiment was designed as a standard categorisation reaction time experiment, consisting of a training stage followed by a speeded transfer stage. In both stages, stimuli were presented in blocks, each block containing all of the twelve stimuli in random order. There were two conditions in the experiment to which participants were randomly assigned. In the *equal frequency* (EF) condition, each training block contained one copy of each stimulus. In the *unequal frequency* (UF) condition, training blocks contained each stimulus once together with an extra four copies of one stimulus (stimulus five) distributed randomly throughout the block. In the learning stage, participants in both conditions were presented with a total of 30 training blocks. Participants in the EF condition, therefore, saw stimulus five 30 times during training, whereas those in the UF condition saw the same stimulus 150 times.

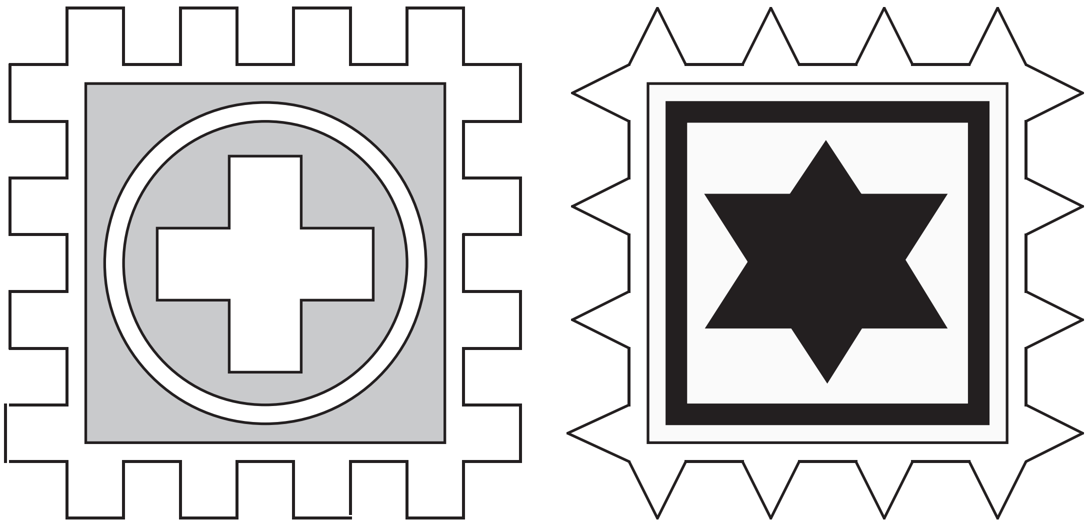

Figure 10.1: Sample stimuli in Experiment 4

On each training trial, a fixation cross appeared in the centre of the screen for 300 ms. After 200 ms of blank screen, one of the stimuli appeared on the screen and remained until the participant registered a category decision by pressing one of the buttons on the button box. Before the training stage commenced, participants were told that they could take as much time as they wished to make a decision and were not informed that their response times were to be recorded. Once a category decision had been made, the computer produced an auditory signal for 500 ms to indicate the accuracy of the response. If a correct category response was made, a 600 Hz (high pitch) tone was produced whereas if the response was incorrect, a 100 Hz (low pitch) tone was given. There then followed a gap of 800 ms before the fixation cross reappeared. Participants were allowed to take a short rest twice during the training session, after nine and nineteen training blocks had been completed.

In the transfer stage, participants in both conditions were again presented with each of

Table 10.1: Structure of Stimuli in Experiment 4

Structure	Stimulus Number	Dimension				
		Border	Background	Symbol	Shape	Colour
Category A	1	1	1	0	0	0
	2	1	0	1	0	1
	3	1	0	0	1	0
	4	0	1	1	0	1
	5	0	1	0	1	0
	6	0	0	1	1	1
Category B	7	1	0	1	1	1
	8	1	0	1	0	0
	9	1	0	0	0	1
	10	0	1	1	1	0
	11	0	1	0	1	1
	12	0	1	0	0	0

the twelve stimuli in ten blocks and were asked to respond as quickly as possible without sacrificing accuracy. All aspects of the transfer stage were the same as the training stage except that no auditory feedback was given in the transfer stage.

Results

Training

Because stimulus frequency was manipulated during training, participants were not trained to a particular criterion of performance. Therefore, all 51 participants were allowed through to the transfer stage and a criterion of an average of at least 70% correct responses over the last five training blocks was subsequently used to determine whether participants' data should be included in the analysis. This inclusion criterion is a great deal less demanding than that employed in Experiments 1 and 2 which required perfect classification accuracy for all stimuli before training stopped. It is hoped that this more liberal rule will result in differences in response accuracy between stimuli to remain in the testing stage. The inclusion criterion required the exclusion of 23 participants, (12 from the EF condition and 11 from the UF condition), and left 14 participants per condition for analysis. The mean proportion of correct responses for those participants who did not achieve criterion performance over the last five blocks was .57 (SD = 0.08) for the EF condition and .61 (SD = 0.07) for the UF condition. The mean proportion of correct responses for those participants who did achieve criterion performance over the last five blocks was .85 (SD = 0.11) for the EF condition and .85 (SD = 0.11) for the UF condition. An ANOVA on the error data provided by the successful participants yielded a significant effect of stimulus type, $F(11, 286) = 5.10, p < .001, MSE = .112$, and a significant interaction between stimulus type and condition, $F(11, 286) = 6.93, p < .001, MSE = .152$. An ANOVA on the response time data also produced a significant effect of stimulus,

$F(11, 286) = 3.79, p < .001, MSE = 576909$ and a significant interaction between stimulus and condition, $F(11, 286) = 5.78, p < .001, MSE = 879622$.

The precise nature of the effect of the frequency manipulation on both categorisation accuracy and response time for stimulus 5 can be seen in Figures 10.2 and 10.3. To reduce the number of points on the x axis, training blocks were grouped into three so that each point represents the average score over three blocks. Figure 10.2 shows clearly that participants in the UF condition reached a high degree of accuracy (i.e. above 90% correct) in classifying stimulus five very early on in the training stage and that they continued to improve until accuracy levels off at just below 100%. In the EF condition however, classification accuracy for the same stimulus remains close to 50%, only improving during the last 10 training blocks but never reaching above the 70% correct level.

Figure 10.3 shows that the difference in frequency also has a marked effect on response times from very early on in training. By the third block, participants in the UF condition are more than 1000 ms faster when categorising stimulus five than those in the EF condition. Although the RT for stimulus 5 in both conditions reduces during the course of training, the average difference between them does not significantly improve over time, remaining at about 1000 ms. Such a large and sustained difference in the mean RTs for stimulus 5 between the conditions is unlikely to be due solely to a stimulus repetition effect.

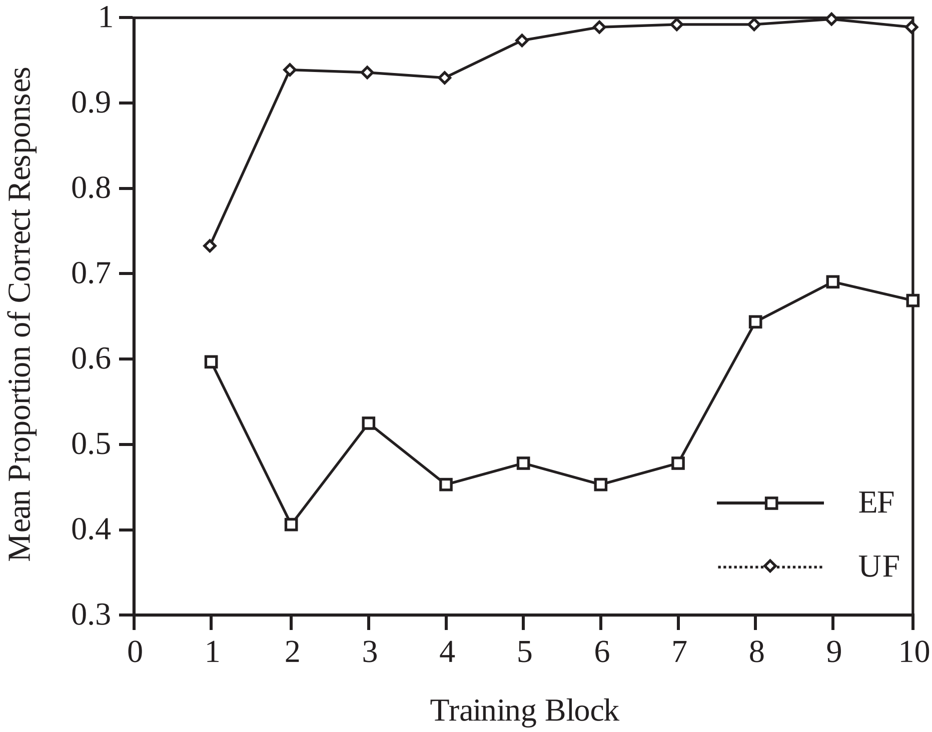

Figure 10.2: Mean proportion of correct responses for Stimulus 5 during training in equal frequency (EF) and unequal frequency (UF) conditions, Experiment 4. Note. Each point on the x axis is the average value over three blocks.

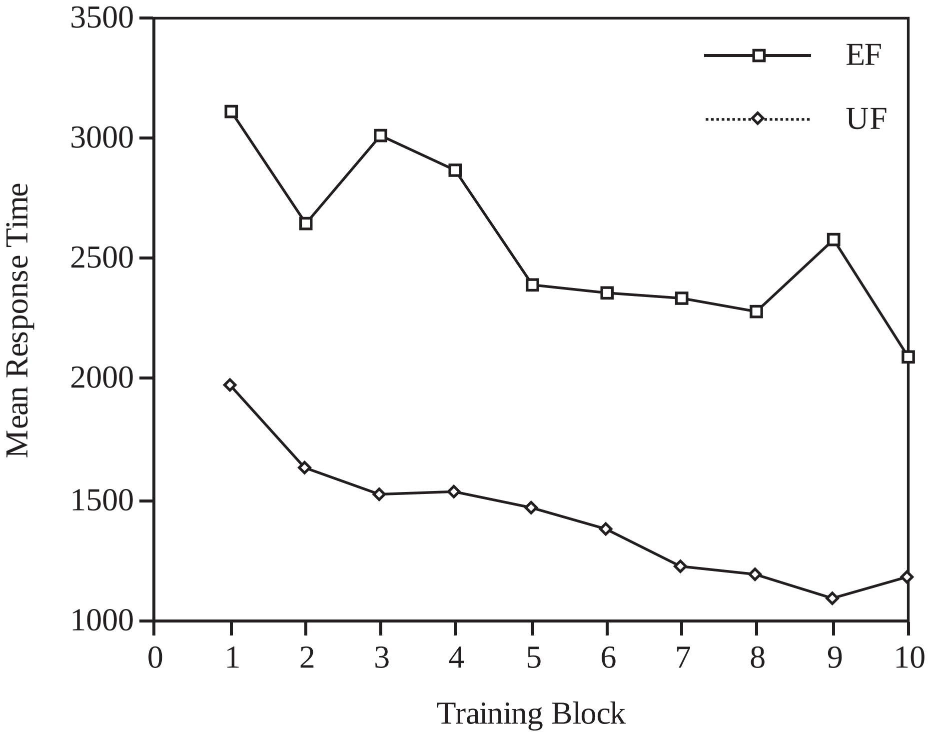

Figure 10.3: Mean response times for Stimulus 5 during training in equal frequency (EF) and unequal frequency (UF) conditions, Experiment 4. Note. Each point on the x axis is the average value over three blocks.

Transfer

The proportions of category A responses and mean RTs for each stimulus are reported in Table 10.2. An ANOVA on the mean proportion of correct responses for each stimulus produced a significant main effect of stimulus, $F(11, 286) = 2.18, p < 0.05, MSE = 0.098$ and significant interaction between stimulus type and condition, $F(11, 286) = 3.24, p < 0.001, MSE = 0.146$. An ANOVA on the mean RTs for each stimulus also yielded a significant main effect of stimulus type, $F(11, 286) = 2.17, p < 0.05, MSE = 245653$ and a significant interaction between stimulus and condition, $F(11, 286) = 1.84, p < 0.05, MSE = 207778$.

The effect of the frequency manipulation on the category responses for stimulus five is quite striking and is consistent with the training data. As can be seen in Figure 10.2, the mean proportion of correct responses for stimulus five at the end of the training session in the EF condition is slightly above the .6 value and this level of accuracy is maintained in the transfer stage. Similarly in the UF condition, from approximately half way through the training stage onwards, stimulus five is categorised correctly as a category A stimulus close to 99% of the time and again, this level of accuracy is retained through to the transfer stage. The effect of the frequency manipulation on the average response times for

Table 10.2: Proportion of Category A Responses (RP) and Response Times (RT) for Each Stimulus, Experiment 4

	EF Condition		UF Condition	
Stimulus	RP	RT	RP	RT
1	0.814	1172	0.714	1258
2	0.800	1357	0.843	1234
3	0.836	1113	0.929	1220
4	0.957	1091	0.807	1328
5	0.614	1485	0.986	1115
6	0.843	1224	0.850	1218
7	0.221	1368	0.371	1407
8	0.143	1266	0.107	1316
9	0.229	1429	0.150	1352
10	0.093	1249	0.179	1176
11	0.229	1221	0.093	1042
12	0.286	1568	0.250	1282

stimulus five is also quite marked. In the EF condition, with the exception of stimulus 12, participants were, on average, slower in categorising stimulus five than for all of the other stimuli. The mean RT for stimulus five in the UF condition, however, is on average, after stimulus 11, the stimulus to which participants most quickly responded. Participants in the UF condition responded on average 370 ms faster to stimulus five than participants in the EF condition.

The speed and accuracy of participants' responses to stimulus 11 are particularly interesting in respect to one of the main findings of Nosofsky (1988) discussed in Chapter 7 which saw classification accuracy and typicality ratings worsen for stimuli very similar to a high frequency stimulus in the opposite category. Stimulus 11 shares four of the five dimension values with stimulus 5 making the stimuli very similar when attention is evenly distributed across dimensions. To classify stimulus 11 as accurately and as rapidly as is observed in the UF condition, therefore, participants were required to place a great deal of attention on the dimension upon which the stimuli differed (i.e. the *symbol*).

Model-based analysis

EGCM

The EGCM was applied to the observed response proportions and reaction times from the two conditions. The model for each condition had fourteen free parameters—five inclusion rate parameters, q, four utility parameters, u, the sensitivity parameter c, a category response bias parameter, β, the residual time parameter, t_{res}, and the Ψ and θ parameters used in the dimension processing time function. The predicted category A response proportions and mean response times produced by the model for each stimulus in both conditions are shown in Table 10.3.

The model for the EF condition accounted for 95.1% of the variance in the response proportions ($RSS = 0.059$) and 80.4% of the variance in the RTs ($RSS = 47039$) whereas the model for the UF condition accounted for 93.8% of the variance of the response proportions ($RSS = 0.088$) but only 22.3% of the variance of the response times ($RSS = 89720$). Note that the predicted response proportions and RTs for the EF condition closely match the similarity relationships between the stimuli, forming two symmetrical patterns around two imaginary lines between stimuli 3 and 4 and stimuli 9 and 10. Note that, in the UF condition, the predicted response proportion for stimulus 5 is considerably less than the observed value. However, the model does predict the difference in mean RT for stimulus 5 across the conditions but captures less well the accuracy and rapid response to stimulus 11 in the UF condition. Note also the large discrepancy between the observed and predicted response proportions for stimulus 12 in the UF condition. This is most likely to be due to the fact that stimulus 12 differs from stimulus 5 on only one dimension—the surrounding *shape*. The EGCM predicts that this dimension is not particularly important in the computation of stimulus-exemplar similarity, nor does it predict a particularly high salience (q) value for this dimension. In contrast, the two dimensions given the highest values of q and u by the EGCM are dimensions upon which stimuli 5 and 12 share the same value. Therefore, given the high strength of stimulus 5, the summed similarity between stimulus 12 and the category A stimuli will also be high.

Table 10.3: Observed (Obs) and EGCM Predicted (Pred) Proportions of Category A Responses (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 4

	EF Condition				UF Condition			
	RP		RT		RP		RT	
Stimulus	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	0.814	0.808	1172	1196	0.714	0.751	1258	1240
2	0.800	0.750	1357	1423	0.843	0.832	1234	1209
3	0.836	0.858	1113	1101	0.929	0.883	1220	1199
4	0.957	0.858	1091	1101	0.807	0.836	1328	1207
5	0.614	0.750	1485	1423	0.986	0.882	1115	1199
6	0.843	0.808	1224	1196	0.850	0.746	1218	1275
7	0.221	0.261	1368	1469	0.371	0.262	1407	1368
8	0.143	0.201	1266	1245	0.107	0.166	1316	1225
9	0.229	0.223	1429	1336	0.150	0.175	1352	1229
10	0.093	0.223	1249	1336	0.179	0.262	1176	1248
11	0.229	0.201	1221	1245	0.093	0.170	1042	1220
12	0.286	0.261	1568	1469	0.250	0.431	1282	1321

The best-fitting parameter values for the EGCM are shown in Table 10.4. Note that the utility value for the *symbol* dimension is, as expected, relatively high in the UF condition. In the EF condition also, the *symbol* dimension is given the highest values for the utility and inclusion rate parameters. Moreover, the model predicts that the *background* dimension had very little significance in the computation of stimulus-exemplar similarity in the UF condition.

Table 10.4: Best-Fitting Parameter Values for EGCM, Experiment 4. Note. The value of the utility parameter for the *symbol* dimension (in brackets) is constrained by the values of the other four.

	Condition	
Parameter	EF	UF
$q(border)$	0.019	0.002
$q(background)$	0.004	0.004
$q(colour)$	0.007	0.119
$q(shape)$	0.016	0.015
$q(symbol)$	0.023	0.033
Ψ	0.496	0.360
t_{res}(ms)	1000	1189
$u(border)$	0.192	0.138
$u(background)$	0.187	0.000
$u(colour)$	0.186	0.200
$u(shape)$	0.205	0.164
$[u(symbol)]$	0.230	0.498
c	10.058	32.261
θ	4.944	4.228
β	0.546	0.584

EBRW

The EBRW model was also applied to the response proportion and RT data from both conditions. In fitting the model to each data set, it was found that, as with Experiment 2, estimating a parameter for each response boundary gave no better fit to the data than using only one. Therefore, a single parameter representing both boundaries was used. Each model had nine free parameters, the sensitivity parameter, c, four dimension weights, w, a category decision boundary parameter, A, the time constant parameter, α, and the slope, k and y-intercept, μ_R, parameters used in the linear regression computation. The predicted response proportions and RTs produced by the models are displayed in Table 10.5. The best-fitting model for the EF condition accounted for 94.2% ($RSS = 0.070$) and 73.7% ($RSS = 63316$) of the variance in the response proportions and response times respectively while that for the UF condition accounted for 80.8% ($RSS = 0.272$) and 48.6% ($RSS = 59386$) of the variance in the response proportions and response times respectively. The best fitting parameter values for the EBRW in each condition are displayed in Table 10.6.

Like the EGCM, the response accuracy and RT predictions of the EBRW in the EF condition closely follow the similarity structure of the stimuli. Also in line with the EGCM, the EBRW places highest weight on the *symbol* dimension in the EF condition.

Although the model was able to provide reasonably accurate joint fits to the data in the

Table 10.5: Observed (Obs) and EBRW Predicted (Pred) Proportions of Category A Responses (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 4

	EF Condition				UF Condition			
	RP		RT		RP		RT	
Stimulus	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	0.814	0.876	1172	1245	0.714	0.811	1258	1249
2	0.800	0.635	1357	1461	0.843	0.840	1234	1250
3	0.836	0.943	1113	1105	0.929	0.985	1220	1240
4	0.957	0.943	1091	1105	0.807	0.982	1328	1279
5	0.614	0.635	1485	1461	0.986	0.961	1115	1105
6	0.843	0.876	1224	1245	0.850	0.812	1218	1247
7	0.221	0.289	1368	1423	0.371	0.078	1407	1279
8	0.143	0.120	1266	1248	0.107	0.160	1316	1260
9	0.229	0.154	1429	1292	0.150	0.156	1352	1260
10	0.093	0.154	1249	1292	0.179	0.225	1176	1244
11	0.229	0.120	1221	1248	0.093	0.445	1042	1192
12	0.286	0.289	1568	1423	0.250	0.142	1282	1264

EF condition, it was unable to do so in the UF condition. In order to determine whether the model could produce better fits to either data set, two versions of the model were optimised on differentially weighted goodness of fit measures. In the first model, the R^2 value for the predicted response proportions was given a weight 100 times greater than that of the response times while in the second model, these weightings were reversed. The first model accounted for 95.8% of the variance of the response proportions ($RSS = 0.060$) and 13.4% of the variance in the response times ($RSS = 100050$) while the corresponding values produced by the second model were 72.2% ($RSS = 0.395$) and 30.1% ($RSS = 80703$). The small increases in the goodness of fit for the more highly weighted predictions and the overall decrease in goodness of fit of the differentially weighted models suggests that the initial model had achieved a maximal fitness level and that no further increase in the accuracy of the RT predictions is likely.

Note, that, like the EGCM, the EBRW predicts that the *background* dimension has no significance in the stimulus-exemplar similarity computation for the UF condition. Also note, however, that the utility value for the *symbol* dimension in this condition is relatively small. This is probably one of the reasons that the model fails to provide a good account of the data because the model predicts a much greater category A response probability and a slower RT for stimulus 11 in the UF condition than was observed.

ALCOVE(RT)

Learning curves

To examine the ability of ALCOVE to account for the effect of the frequency manipulation on the learning curves for stimulus 5, the model was applied to the category A response

Table 10.6: Best-Fitting Parameter Values for EBRW, Experiment 4. Note. The value of the weight parameter for the *symbol* dimension (in brackets) is constrained by the values of the other four.

	Condition	
Parameter	EF	UF
$w(border)$	0.184	0.232
$w(background)$	0.133	0.000
$w(colour)$	0.123	0.296
$w(shape)$	0.105	0.319
$[w(symbol)]$	0.456	0.153
c	7.22	12.196
α	0.398	0.146
A	2	1
k	901.145	7274.173
μ_R(ms)	0.000	0.000

proportion data from the two conditions. As was found in Experiment 3, the increased presentation frequency of stimulus 5 in the UF condition can be expected to result in a more rapid increase in the accuracy of the models' classification for that stimulus during training. Four model parameters were estimated in fitting the model to the data—the hidden unit specificity parameter, c, the association and attention learning rate parameters, λ_w and λ_α, and the category unit response mapping parameter, ϕ. As before, R^2 was used as a measure of goodness-of-fit. In applying the model to the data from each condition, the network was trained for thirty epochs on all twelve stimuli using batch updating. In the UF condition, an extra four copies of stimulus 5 were presented in each epoch.

The model provided good fits to the data, yielding R^2 values of .94 ($RSS = 0.072$) and .95 ($RSS = 0.070$) for the EF and UF conditions respectively. The predicted category A response proportions produced by the model for each stimulus in the two conditions are shown in Table 10.7.

Note that the observed response proportions for stimulus 5 in the two conditions are closely matched by the predictions of the model. Note also, however, the large discrepancy between the observed and predicted response proportions for stimulus 12 in the UF condition. This prediction was also made by the EGCM. An examination of the input unit attention strengths revealed that, like the EGCM, the ALCOVE model gave relatively little weight to the *shape* dimension ($w_{shape} = 0.941$) compared to *symbol* and *colour* dimensions ($w_{symbol} = w_{colour} = 1.33$) causing the model to predict a greater summed similarity between stimulus 12 and the category A stimuli. The best-fitting parameters for the model are shown in Table 10.8. To examine the effect of the frequency manipulation on the models' categorisation responses and compare them to the observed data, the category A response probability for stimulus 5 after each training epoch in both conditions are plotted together with the observed data in Figure 10.4.

To place the models' responses alongside the observed values, they were averaged over

Table 10.7: Observed (Obs) and ALCOVE Predicted (Pred) Proportions of Category *A* Responses for Each Stimulus in Equal Frequency (EF) and Unequal Frequency (UF) Conditions, Experiment 4

	EF Condition		UF Condition	
Stimulus	Obs	Pred	Obs	Pred
1	0.814	0.863	0.714	0.720
2	0.800	0.672	0.843	0.821
3	0.836	0.947	0.929	0.975
4	0.957	0.947	0.807	0.958
5	0.614	0.672	0.986	0.970
6	0.843	0.863	0.850	0.751
7	0.221	0.346	0.371	0.335
8	0.143	0.140	0.107	0.093
9	0.229	0.140	0.150	0.093
10	0.093	0.140	0.179	0.139
11	0.229	0.140	0.093	0.139
12	0.286	0.346	0.250	0.412

Table 10.8: Best-Fitting Parameter Values for ALCOVE, Experiment 4

	Condition	
Parameter	EF	UF
c	4.345	1.259
λ_α	0.023	0.053
λ_w	0.004	0.020
ϕ	2.978	1.943

three training epochs so that each point on the x axis represents the mean predicted category *A* response probability over three epochs. The learning curves produced by the model for both conditions closely resemble those observed in the experiment, indicating that the frequency manipulation does result in model behaviour bearing a number of similarities to human behaviour. It is particularly interesting to notice that in the EF condition, both the human data and the model predictions show a slight reduction in accuracy to below the 0.5 level before rising again toward the end of training and that in the UF condition, both rise sharply to above the 0.9 level early in training and then remain high throughout.

Transfer data

The ALCOVE(RT) model was applied to the combined category *A* response proportion and RT data from the two conditions of Experiment 4 using the same method as was employed in Experiments 1 and 2. Table 10.9 shows the predicted values produced by the model for both conditions while the best-fitting parameters for the model are displayed in

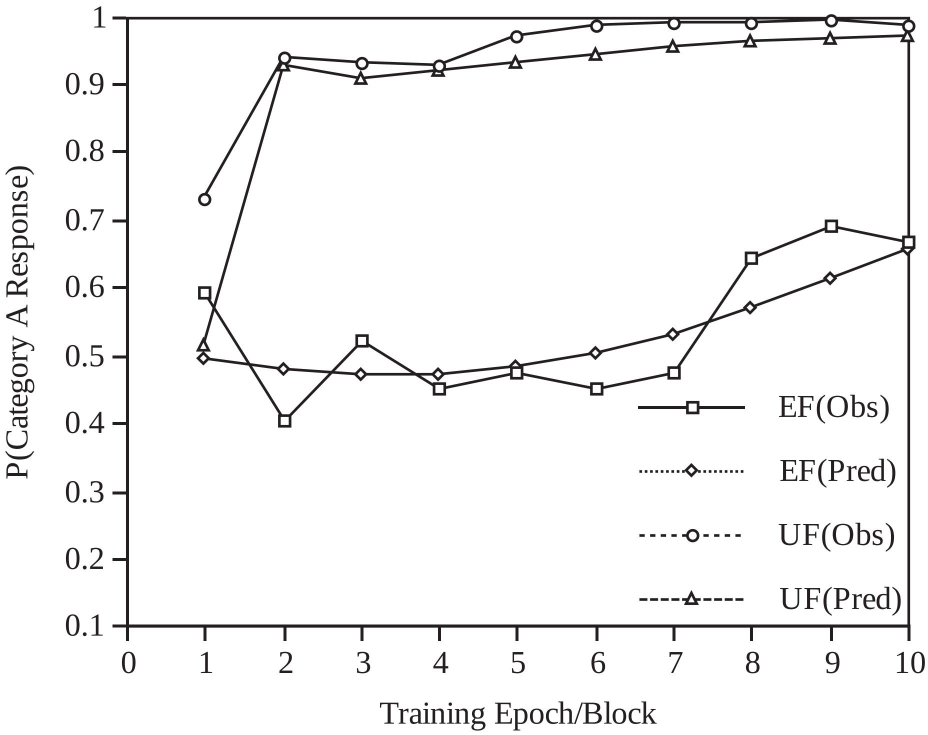

Figure 10.4: Plots of observed (Obs) and ALCOVE predicted (Pred) proportions of category A responses for Stimulus 5 during training in equal frequency (EF) and unequal frequency (UF) conditions, Experiment 4. Note. Each point on the x axis represents the average value over three epochs/blocks.

Table 10.10. The model provides a relatively good fit to the combined data from the EF condition, accounting for 82% ($RSS = 0.212$) and 67% ($RSS = 80608$) of the variance in the response proportions and RTs respectively. The fit to the data from the UF condition is less close. Although the model yielded an R^2 of .927 ($RSS = 0.104$) for the response proportions, the same measure for the RTs was .119 ($RSS = 101780$).

Like the EGCM and EBRW, the response accuracy and RT predictions of ALCOVE(RT) closely followed the similarity structure of the stimuli. Also in line with the EGCM and EBRW models, the ALCOVE(RT) model predicts that attention is withdrawn from the *background* dimension in the UF condition ($w_{background} = 0.0$) and also predicts that in the same condition, most attention is paid to the *symbol* dimension ($w_{symbol} = 0.676$).

Discussion

Experiment 4 provides strong evidence for the claim that stimulus frequency can affect the course of category learning and subsequent performance during a subsequent speeded categorisation task. It was found that the frequency manipulation had a significant effect

Table 10.9: Observed (Obs) and ALCOVE(RT) Predicted (Pred) Proportions of Category A Responses (RP) and Response Times (RT in ms) for Each Stimulus, Experiment 4

	EF Condition				UF Condition			
	RP		RT		RP		RT	
Stimulus	Obs	Pred	Obs	Pred	Obs	Pred	Obs	Pred
1	0.814	0.724	1172	1240	0.714	0.818	1258	1280
2	0.800	0.706	1357	1385	0.843	0.844	1234	1239
3	0.836	0.669	1113	1095	0.929	0.825	1220	1199
4	0.957	0.669	1091	1095	0.807	0.844	1328	1199
5	0.614	0.706	1485	1385	0.986	0.827	1115	1159
6	0.843	0.724	1224	1240	0.850	0.827	1218	1280
7	0.221	0.318	1368	1530	0.371	0.167	1407	1280
8	0.143	0.271	1266	1240	0.107	0.157	1316	1199
9	0.229	0.271	1429	1240	0.150	0.157	1352	1199
10	0.093	0.271	1249	1240	0.179	0.164	1176	1199
11	0.229	0.271	1221	1240	0.093	0.164	1042	1199
12	0.286	0.318	1568	1529	0.250	0.175	1282	1280

on both the levels of classification accuracy and mean RTs for the manipulated frequency stimulus in both the training stage and the later transfer stage.

In the training stage, it was found that the frequency manipulation had a significant effect on both response accuracy and RT to stimulus 5 in the UF condition and the effect on RTs was so large that it was considered unlikely to be due solely to a stimulus repetition effect. The difference in the learning curves for stimulus 5 in each condition also clearly showed the effect of the frequency manipulation. As with Experiment 3, the ALCOVE model was able to predict learning curves for the frequency manipulated stimulus in both conditions very similar to the observed curves.

Table 10.10: Best-Fitting Parameter Values for ALCOVE(RT), Experiment 4

	Condition	
Parameter	EF	UF
c	4.359	4.267
λ_w	0.032	0.079
λ_α	0.006	0.120
ϕ	2.998	3.744
τ_{hid}	0.492	0.099
τ_{out}	0.343	0.093
ϑ	0.102	0.200
υ	144.834	40.327
ω	950.530	836.145

The EGCM and EBRW models gave relatively similar accounts of the data from Experiment 4. In particular, both models were unable to provide accurate predictions for the RT data in the UF condition but did model the frequency effect by predicting a relatively short RT for stimulus 5. Similarly, the ALCOVE(RT) model could not give accurate predictions of RTs for individual stimuli in the UF condition but did predict a shorter RT for stimulus 5.

As with Experiment 3, the strong frequency effect found in Experiment 4 also lends weight to the claim that the results of Nosofsky and Palmeri's (1997) Experiment 2 are unlikely to be due entirely to a surprise effect.

In Chapter 11, the results of this experiment and the other earlier experiments are assessed and the implications for the various themes of this research are drawn.

Chapter 11

General Discussion

Introduction

This research has been undertaken to achieve three primary goals. The first of these goals is to provide a definite answer to the question concerning the effect of stimulus frequency on two measures of categorisation performance—response accuracy and RT. The second aim is to test the two most promising current models of categorisation RT—Lamberts' EGCM and Nosofsky and Palmeri's EBRW—in terms of their ability to predict combined accuracy and RT data from experiments in which stimulus frequency is manipulated. The third goal of the research is to develop a connectionist account of categorisation RT and to determine the extent to which a network model based on the principles of accumulation of activation over time can predict data from standard categorisation RT experiments as well as those from experiments in which stimulus frequency is a dependent variable. In addition to these three main goals, it was expected that a further outcome of the research would be the provision of greater insights into the nature of the information required for accurate and rapid classification. The three models are each based upon a different assumption concerning this information and so it was hoped that differences in performance between the models and individual patterns of response would indicate strengths and weaknesses in these various assumptions.

The research has been successful in supplying clear answers to a number of the questions posed at the outset. In particular, the results of two experiments (3 and 4) provide strong evidence for the claim that category learning and subsequent classification accuracy and response times are significantly affected by the frequency of stimulus presentation during training. The models' varying success in accounting for the data revealed by this research also clearly shows that certain assumptions concerning the information required for accurate and rapid classification cannot be justified. More specifically, it shows that a highly constrained model such as ALCOVE(RT), in which behaviour is determined by the category structure alone, is too inflexible to capture the variance in the response data. The differences in performance found between the EGCM and EBRW were not sufficiently great or consistent to permit any strong inferences to be drawn concerning the relative merit of their particular processing assumptions. Although the EGCM did occasionally account for the experimental data better than the EBRW, there are several possible factors which may be seen as underlying this superior performance.

In this chapter, I summarise the results of this research and discuss several conclusions which may be drawn concerning the distinct but related themes that have been examined in the previous chapters and outlined above. In discussing these results, I will address each of the issues separately, although inevitably there will be some overlap between the topics. The summary will cover each of the three research aims mentioned above and will also include a discussion of the ability of the ALCOVE model to account for the frequency effects on the learning curves for individual stimuli in Experiments 3 and 4.

The effect of stimulus frequency on categorisation

In Chapter 7, I discussed the relationship between stimulus frequency and categorisation performance and described a number of early experimental findings for and against the idea that presentation frequency affected graded category structure. I then summarised two recent empirical investigations into frequency effects on various measures of categorisation performance carried out by Nosofsky (1988) and Ashby, Boynton and Lee (1994) which had conflicting conclusions. The former suggested that stimulus frequency did affect categorisation accuracy and typicality judgements whereas the latter found that there was no significant effect of stimulus familiarity on categorisation response times. The effect of stimulus frequency on categorisation RTs was also the subject of Chapter 8, which discussed a more recent experiment by Nosofsky and Palmeri (1997) in which significant effects of frequency were found. One of the primary aims of this research, therefore, was to investigate this relationship further and to gather additional evidence in order to assist in the resolution of the issue.

The results of Experiments 3 and 4 gave strong support for the claim that stimulus frequency affects the rate of category learning for individual stimuli and categorisation performance for individual stimuli in subsequent transfer stages. In Experiment 3, a significant difference in mean RT for stimulus 2 (the frequency of which was manipulated) was observed between the conditions. It was noted that the mean RT for stimulus 2 over the period of training was 118 ms shorter than that for the same stimulus in the EF condition. It was also discovered that the frequency manipulation affected the RT for stimulus 2 in relation to the other stimuli in the same condition. Whereas the RT for stimulus 2 over the training period in the EF condition was not particularly shorter than the other stimuli in the same condition, that for stimulus 2 across training in the UF condition was on average 248 ms faster than the next fastest stimulus.

The frequency manipulation was also seen to affect the response accuracy for stimulus 2 across conditions. The effect of the frequency manipulation in the experiment is easily observed in the response proportions for stimulus 2 in both conditions. Whereas the probability of a correct response for stimulus 2 in the EF condition rose from .833 in average test point 1 to reach a maximum of .942, that for the same stimulus in the UF condition started at .983 in average test point 1 and remained at a high level throughout training. In respect to the effect of frequency on stimulus 2 response accuracy relative to other stimuli in the same condition, the level of accuracy for stimulus 2 in the EF condition was not significantly better than that for other stimuli in the same condition. Classification accuracy for stimulus 2 in the UF condition, particularly in the early stages of training, was on average 15% higher than the next best classified stimulus in the same condition.

As with Experiment 3, the results of Experiment 4 indicate clearly the effect of frequency on classification performance. In the training stage, the frequency manipulation was observed to have a significant effect on the accuracy of the responses to the frequency manipulated stimulus (stimulus 5). It was seen that responses to stimulus 5 in the UF condition reached above 90% correct on average from early on during training and continued to rise to just below 100% accuracy throughout the rest of the training stage. Those for stimulus 5 in the EF condition, however, stayed close to 50% accuracy during the larger part of the training stage and never reached above the 70% level. The rapidity of responses to stimulus 5 across the conditions was also significantly affected by the frequency manipulation. The mean RT for stimulus 5 in the UF condition started and remained consistently 1000 ms shorter than that for the same stimulus in the EF condition throughout the entire course of training. It was considered unlikely that this large difference in the mean RTs for stimulus 5 between the conditions is unlikely to be due solely to a stimulus repetition effect.

The results of the testing stage followed those of the training stage by indicating a significant difference in levels of response accuracy and RT between the conditions. Participants were 37.2% more likely to classify stimulus 5 correctly in the UF condition than they were in the EF condition. Participants in the UF condition were also on average 370 ms faster in responding to stimulus 5 than participants in the EF condition.

The results of Experiments 3 and 4 suggest that the frequency effects found by Nosofsky and Palmeri (1997) in their Experiment 2 were unlikely to be entirely due to an effect of surprise, although whether some element of surprise did have some confounding effect in their experiment cannot be decided.

Experiments 3 and 4 both provide strong evidence for the claim that models of categorisation RT must include a mechanism for incorporating stimulus frequency into their account. In this respect, the RT-D model of Ashby et al. can be seen to be incomplete as it considers RT to be purely a function of distance from decision boundary.

ALCOVE, stimulus frequency, and category learning

In Experiments 3 and 4, analysis of the learning curves for the frequency manipulated stimuli in each experiment showed marked differences in the rates of classification learning. The category labels of stimuli with higher presentation frequency were learned far more rapidly than the same stimuli in the equal frequency conditions.

The mean proportion of correct responses for stimulus 2 for each of the nine test blocks in Experiment 3 was plotted in Figure 9.2. Examination of these plots shows that the probability of a correct response for stimulus 2 in the UF condition was approximately .95 in the first test block and remained close to 1 throughout the greater part of the experiment whereas that for the same stimulus in the EF condition started at .8 in test point 1 and rose to an average of approximately .9 across the remaining test points.

When the ALCOVE model was applied to the category A response proportion data from both conditions, the overall predictions for the complete set of stimuli were not particularly close, although the model did account for 88.7% and 93% of the variance in the data from the EF and UF conditions respectively. However, the model was able to produce learning curves for stimulus 2 in both conditions which were in general agreement with the observed curves. The curve produced by ALCOVE for stimulus 2 in the EF condition traced a smooth arc from approximately .76 in test point 1 to approximately .9 in test point 9 while the corresponding curve for the UF condition started at approximately .96 and ended at approximately .99.

The learning curves for the frequency manipulated stimulus observed in Experiment 4 were plotted in Figure 10.4. The difference in the curves for the frequency manipulated stimulus between the conditions was more marked in Experiment 4 than in Experiment 3. Whereas the probability of a correct response for stimulus 5 in the EF condition started at approximately .6 and remained between .5 and .7 throughout the course of training, that for the same stimulus in the UF condition started at approximately .7 and then immediately rose to above .9 to stay close to 1 from about half way through the training stage onwards.

As with Experiment 3, while the R^2 produced by ALCOVE for the EF and UF conditions was .94 and .95 respectively, the predicted response proportions produced by the model were not particularly close to the observed values of individual stimuli. However, the fits to the observed learning curves for stimulus 5 in the two conditions produced by ALCOVE were closer than those to the curves from Experiment 3. The predicted probability of a correct response for both conditions started at .5. Whereas the curve for the EF condition traced a smooth arc to eventually finish at a value of .6, that for the UF condition rose rapidly to approximately .9 and then remained between .9 and 1 for the greater part of the training stage to finish at approximately .95

The success of ALCOVE in reproducing the learning curves from both experiments suggests that by simply increasing the number of stimulus presentations to a network by the same ratio as that used in the experiment, one can produce behaviour in the model which is very close to that produced by humans. The close fit between the observed and predicted learning curves in Experiments 3 and 4 provides strong evidence for ALCOVE as a model of learning. One further conclusion to be drawn from these results is that significant aspects of performance can be accounted for by a model which simply uses information about the abstract category structure. This latter result lends weight to the claim that it is worthwhile to study the abilities of models to predict behaviour using relatively constrained information about the task domain and provides some justification for the use of ALCOVE(RT) as a hypothetical model of categorisation RT.

EGCM, EBRW and ALCOVE(RT) models evaluated

In Chapter 4, the different underlying assumptions held by the EGCM, EBRW and ALCOVE(RT) models concerning the information used and processes involved in categorisation and their relationship to response times were discussed. It was explained how, because of the learning procedure used to modify the attention and association weights, the ALCOVE(RT) model relied exclusively on the abstract category structure to generate predictions of category learning behaviour and that categorisation RT was based on the gradual accumulation of activation according to the ratio of summed exemplar activations of alternative categories. It was also discussed how, although the EBRW also assumed that categorisation was determined by the category structure, the method of obtaining dimension attention weights in the EBRW was not constrained by this structure, allowing the model an additional degree of flexibility over ALCOVE(RT). Finally, it was observed that the EGCM was more flexible still by assuming that, in addition to the category structure, perceptual information such as the salience of stimulus dimensions is to be taken into account when modelling RTs. A central aim of this research was to determine whether these three alternative sets of assumptions were equally able to account for combined classification accuracy and RT data or whether one or more were inadequate to do so. It was hoped that the success or failure of individual models to account for the data would allow inferences to be made as to the nature of the information utilised in speeded classification tasks.

In Experiment 1, the EGCM accounted for 93.1% of the variance in the RT data and 98.8% of the variance in the response proportion data and provided close fits to the values for individual stimuli. The relative sizes of the best-fitting values of the inclusion rate parameters produced by the model were as expected given the error data in the training stage and the choice data in the testing stage. The model was forced, however, to predict higher response proportions for the category B stimuli than were observed because of the relative rapidity of the responses to these stimuli compared to those from category A. The EBRW accounted for 66.7% of the variance in the RT data and 88.1% of the variance in the response proportions and the predicted response proportions and RTs for individual stimuli were less close than those produced by the EGCM. Like the EGCM, and for the same reason, the EBRW predicted higher response proportions for the category B stimuli than were observed. It was concluded that the worse performance of the EBRW was due, in the main, to the lack of flexibility which was required to account for the large differences in dimension salience indicated by the EGCM. Although the ALCOVE(RT) model accounted for 0% of the variance in the RT data it did account for 99.7% of the variance in the response proportions but was unable to provided close fits to the values for individual stimuli. It was found that the model predicted an average response proportion for each category and one mean RT for all stimuli. The high R^2 produced for the response proportions predicted by the model was due to the small differences between the observed response proportions for the stimuli in each category.

The EGCM also fared best in Experiment 2, accounting for 82% of the variance in the RT data and 92% of the variance in the response proportion data. Although the model gave a relatively close fit to most of the stimuli, it did predict accuracy values and RTs for some stimuli which were quite far from the observed values. The EBRW accounted for 66.1% of the variance in the RTs and 79.7% of the variance in the response proportions and, as with Experiment 1, provided worse fits to the individual stimulus data in general than the EGCM. However, the ALCOVE(RT) model was the least successful in accounting for the combined accuracy and RT data, accounting for 0% of the variance in the RTs and 98% of the variance in the response proportions. As with Experiment 1, the high R^2 produced for the response proportions predicted by the model was due to the relatively small differences between the observed response proportions for the stimuli in each category. Again, the model predicted one average RT for all of the stimuli and, apart from the two irregular stimuli for which a lower level of accuracy was produced, an average response proportion for each category.

In Chapter 7, Nosofsky and Palmeri's (1997) Experiment 2 was discussed and the previous applications of the EGCM and EBRW (Lamberts, submitted) models to the data were outlined. The ALCOVE(RT) model was also applied to the data. The abilities of the EGCM and EBRW to model the combined data and the frequency effect are very similar. The EGCM accounted for 85% of the variance in the RT data and 99% of the variance in the category A response proportion data and predicted the crossover interaction for stimuli 7 and 8 across the conditions. The EBRW was able to account for 92.6% of the variance in the RT data and 99% of the variance in the choice data and also predicted the crossover interaction for stimuli 7 and 8 across the conditions. The ALCOVE(RT) model was applied to combined choice proportions and RTs from the two conditions separately. The model accounted for 99% of the variance in the choice proportion data but only 2.2% of the variance in the RT data from the U7 condition and for 99% of the variance in the response proportions but only 5.2% of the variance in the RT data in the U8 condition. Although the model did not provide close fits to the choice and RT data from individual stimuli, it did produce the crossover interaction between stimuli 7 and 8 across the conditions.

In Experiment 3, the EGCM and EBRW accounted equally well for the data in the EF condition. The EGCM accounted for 52.2% of the variance in the RT data and 89.6% of the variance in the response proportions whereas the EBRW accounted for 46.6% of the variance in the RT data and 90% of the variance in the response proportion data. In the UF condition, the EGCM provided a better fit to the combined data overall, accounting for 73.4% of the variance in the RTs and 92.8% of the variance in the response proportions. The EBRW accounted for 58.1% of the variance in the RT data and 76% of the variance in the response proportion data. Both models captured the essential effect of the frequency manipulation, predicting higher response accuracy and shorter RT for the increased frequency stimulus (stimulus 2) than for the other stimuli in the UF condition and for the same stimulus in the EF condition. The ALCOVE(RT) model was applied to the data from the average test points individually. Although the model yielded R^2 values of .722, .950 and .957 for average test points 1, 2 and 3 in the EF condition respectively, it did so by predicting a mean proportion for each category and one mean RT for all stimuli in each average test point. The model for the UF condition accounted for 90%, 89.9% and 98.7% of the response proportions and 46.3%, 54% and 77.8% of the variance in the RTs in average test points 1, 2 and 3 respectively. The predictions of the model in the UF condition were also affected by the frequency manipulation, the model predicting more accurate and rapid responses for stimulus 2.

As with Experiment 3, the EGCM and EBRW accounted equally well for the data in the EF condition in Experiment 4. The EGCM accounted for 80% of the variance in the RTs and 95% of the variance in the response proportions while the EBRW accounted for 74% of the variance in the RT data and 94% of the variance in the response proportions. In the UF condition also, the models fared equally, the EGCM accounting for 22% of the variance in the RT data and 93% of the variance in the response proportion data and the EBRW accounting for 49% of the variance in the RTs and 81% of the variance in the response proportions. The models were unable to provide accurate predictions for the RT data in the UF condition but were both able to take the frequency information into account by predicting a shorter RT for the frequency manipulated stimulus (stimulus 5). The ALCOVE(RT) model accounted for 67% of the variance in the RT data and 82% of the variance in the response proportion data from the EF condition and for 11.9% of the variance in the RTs and 92.7% of the variance in the response proportions in the UF condition. Like the EGCM and EBRW, ALCOVE(RT) could not give accurate predictions for the RTs in the UF condition but was able to predict the shorter RT for stimulus 5.

Implications for theories of categorisation RT

One strong conclusion which may be drawn from the overall poor performance of the ALCOVE(RT) model is that the prediction of detailed response accuracy and RT data requires more information than is provided by the abstract category structure alone. The constraints upon the adaptation of attention and association weights imposed by the learning mechanism are so strong as to restrict the possible weight values obtainable by the model too narrowly. It is the additional flexibility achieved by the attention weight adaptation mechanisms in the EGCM and EBRW which allows them to provide closer fits to the observed data.

An alternative explanation of the models' varying degrees of success in accounting for the experimental data is that the ability of each model to fit a given data set is simply a function of the number of free parameters contained in the model. According to this analysis, the reason for the relative inability of the ALCOVE(RT) model to account for the data and the slightly better fits sometimes demonstrated by the EGCM over the EBRW is that the EGCM has additional degrees of freedom with which to capture any noise in the data whereas the ALCOVE(RT) model has far fewer degrees of freedom than either the EGCM or EBRW. It is undoubtedly true that the inclusion of additional parameters can result in a model which is overgeneral in the sense that it is able to provide a reasonably close fit to almost any pattern of data. Unfortunately, apart from the obvious proviso that the number of model parameters should not exceed the number of data points being modelled, there are very few hard and fast rules governing the construction of models. Therefore the inclusion of model parameters must be justified by reasonable assumptions about the cognitive structures and processes taken to exist in the system being modelled. An individual parameter in a model may be tested by applying two versions of the model to a data set—one with the extra parameter and one without—

in order to determine whether a significant increase in model fit justifies its inclusion. Additional justification for a parameter may also be obtained by the carrying out of further experiments in which factors affecting the value of the parameter are manipulated. This is the course of action taken by Lamberts to provide evidence of perceptual processing mechanisms in the EGCM (Freeman & Lamberts, 1997; Lamberts, 1995, 1997; Lamberts & Brockdorff, 1997; Lamberts & Freeman, 1997a, 1997b). In the context of this research, the inclusion of parameters in all of the models under investigation has been justified by reference to the various cognitive structures and functions assumed to take place during the categorisation process. It is also doubtful whether any of the models can be regarded as being overgeneral because none of them was able to provide consistently close fits to all of the data sets.

The results of the comparison between the EGCM and EBRW remain unclear because no significant and consistent superiority of one of the models over the other was observed. If the EGCM had been found to provide consistently better fits to the data, there would have been solid evidence for the role of perceptual processes as determinants of RT. As this is not the case, however, one conclusion which may be drawn is that either model may provide a correct account of the data from these experiments. Lamberts (submitted) has demonstrated a consistent superiority of the EGCM over the EBRW in terms of accounting for combined accuracy and RT data from speeded classification experiments in which response deadlines are imposed (Lamberts, 1995, 1997) or where stimuli are presented very briefly (Lamberts & Freeman, 1997b). In all of these studies, however, Lamberts also found that the models gave virtually equivalent accounts of RT data from experiments when no response deadlines were imposed. The similar results of the experiments reported here suggest that simple speeded categorisation RT experiments may be inadequate to discriminate between the two models. A possible direction for further research which may provide a method of distinguishing between the EGCM and EBRW in the current context, therefore, is to incorporate the deadline conditions or brief stimulus exposure durations used in Lamberts' experiments into experiments in which stimulus frequency is manipulated.

An alternative conclusion which may be drawn is that the research has shown that aspects of both models are useful in providing an accurate account of the data and that, therefore, rather than attempting to discriminate between these accounts, an attempt should be made to combine the most important aspects of each model into a more general account out categorisation. Such a combined model would include a perceptual mechanism for accommodating differences in dimensional salience and a time-dependent category decision mechanism.

References

Anderson, J. A. (1991). Why, having so many neurons, do we have so few thoughts? In Hockley, W. E. & Lewandowsky, S., (Eds.), *Relating Theory and Data: Essays on Human Memory in Honour of Bennet B. Murdock*, (pp. 477–507). Hillsdale, NJ, Erlbaum.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception and probability learning: Some applications of a neural model. *Psychological Review*, 84, 413–451.

Ashby, F. G. (1992). Multidimensional models of categorisation. In Ashby, F. G., (Ed.), *Multidimensional Models of Perception and Cognition*, (pp. 449–483). Hillsdale, NJ, Erlbaum.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorisation response time with multidimensional stimuli. *Perception & Psychophysics*, 55, 11–27.

Ashby, F. G. & Gott, R. E. (1988). Decision rules in the perception and categorisation of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 33–53.

Ashby, F. G. & Lee, W. W. (1991). Predicting similarity and categorisation from identification. *Journal of Experimental Psychology: General*, 120, 150–172.

Ashby, F. G. & Lee, W. W. (1992). On the relationship among identification, similarity, and categorisation: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General*, 121, 385–393.

Ashby, F. G. & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Percepion and Performance*, 16, 598–612.

Ashby, F. G. & Maddox, W. T. (1991). A response time theory of perceptual independence. In Doignon, J. P. & Falmagne, J. C., (Eds.), *Mathematical Psychology: Current Developments*, (pp. 389–414). New York, Springer-Verlag.

Ashby, F. G. & Maddox, W. T. (1992). Complex decision rules in categorisation: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Percepion and Performance*, 18, 50–71.

Ashby, F. G. & Maddox, W. T. (1993). Relations between prototype, exemplar and decision bound models of categorisation. *Journal of Mathematical Psychology*, 37, 372–400.

Ashby, F. G. & Maddox, W. T. (1994). A response time theory of perceptual separability and perceptual integrality in speeded classification. *Journal of Mathematical*

Psychology, 38, 423–466.

Ashby, F. G. & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, 95, 124–150.

Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154–179.

Ashcraft, M. H. (1978). Property norms for typical and atpical items from 17 categories: A description and discussion. *Memory & Cognition*, 6, 227–232.

Atkinson, R. C. & Estes, W. K. (1963). Stimulus sampling theory. In Luce, R. D., Bush, R. R., & Galantner, E., (Eds.), *Handbook of Mathematical Psychology*, (pp. 121–268). New York, Wiley.

Barsalou, L. W. (1981). *Determinants of graded category structure in categories.* PhD thesis, Stanford University.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 629–654.

Barsalou, L. W. (1992). *Cognitive Psychology: An Overview for Cognitive Scientists.* Hillsdale, NJ, Erlbaum.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford, Oxford University Press.

Boring, E. G. (1950). *History of Experimental Psychology.* New York, Appleton-Century-Crofts.

Bullinaria, J. A. (1995). Modelling reaction times. In Smith, L. S. & Hancock, P. J. B., (Eds.), *Neural Computation and Psychology.* Berlin, Springer.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalisation and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 638–648.

Carroll, J. D. & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In Krantz, D. H., Atkinson, R. C., Luce, R. D., & Suppes, P., (Eds.), *Contemporary Developments in Mathematical Psychology*, volume 2, (pp. 57–105). San Francisco, W. H. Freeman.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332–361.

Donders, F. C. (1969). Over de snelheid van psychische processen (On the speed of mental processes). In Koster, W. G., (Ed.), *Attention & Performance, Vol. 2*, (pp. 412–431). Amsterdam, North-Holland. (Original work published 1868).

Eckman, P. (1982). *Emotion in the Human Face.* Cambridge, Cambridge University Press.

Eckman, P. & Oster, H. (1979). Facial expression of emotion. *Annual Review of Psychology*, 30, 527–554.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57, 94–107.

Estes, W. K. (1972). An associative basis for coding and organisation in memory. In Melton, A. W. & Martin, E., (Eds.), *Coding Processes in Human Memory*, (pp. 161–190). Washington, D. C., Winston.

Estes, W. K. (1973). Mmory and conditioning. In McGuigan, F. J. & Lumsden, D. B., (Eds.), *Contemporary Approaches to Conditioning and Learning*, (pp. 265–286). Washington, D. C., Winston.

Estes, W. K. (1986a). Array models for category learning. *Cognitive Psychology*, 18, 500–549.

Estes, W. K. (1986b). Storage and retrieval processes in category learning. *Journal of Experimental Psychology: General*, 115, 155–174.

Estes, W. K. (1994). *Classification and Cognition.* New York, Oxford University Press.

Estes, W. K., Campbell, J. A., Hatsopoulis, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 556–571.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Vol. 1.* New York, Wiley.

Freeman, R. P. J. & Lamberts, K. (1997). Salience and categorisation. *Manuscript submitted for publication.*

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition.* San Diego, Academic Press, Second edition.

Garner, W. R. (1974). *The Processing of Information and Structure.* Hillsdale, NJ, Erlbaum.

Gluck, M. A. & Bower, G. H. (1988a). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227–247.

Gluck, M. A. & Bower, G. H. (1988b). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27, 166–195.

Green, D. M. & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* New York, Wiley.

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.

Hinton, G. E. & Sejnowski, T. J. (1983a). Analysing cooperative computation. In *Proceedings of the Fifth Annual Conference of the Cognitive Science Society.*

Hinton, G. E. & Sejnowski, T. J. (1983b). Optimal perceptual inference. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace model. *Psychological Review*, 93, 411–428.

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in multiple-trace memory. *Psychological Review*, 95, 528–551.

Hintzman, D. L. & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 88, 297–306.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalisation and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 418–439.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79, 2554–2558.

Hume, D. (1739). *A Treatise on Human Nature.* Oxford, Oxford University Press.

Jacoby, L. L. & Brooks, L. R. (1984). Nonanalytic cognition: memory, perception and concept learning. In Bower, G. H., (Ed.), *The Psychology of Learning and Motivation, Vol. 18*, (pp. 1–47). New York, Academic Press.

James, W. (1890). *The Principles of Psychology: Volume I.* New York, Dover.

Knapp, A. G. & Anderson, J. A. (1984). Theory of categorisation based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 616–637.

Kohonen, T. (1982). Self-organised formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.

Kruschke, J. K. (1990a). *ALCOVE: A connectionist model of category learning.* PhD thesis, University of California at Berkeley.

Kruschke, J. K. (1990b). ALCOVE: A connectionist model of category learning (Cognitive Science Research Rep. No. 19). Technical report, Bloomington: Indiana University.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, 5, 3–36.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional Scaling.* London, Sage Publications.

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorisation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1003–1021.

Lamberts, K. (1995). Categorisation under time pressure. *Journal of Experimental Psychology: General*, 124, 161–180.

Lamberts, K. (1997). Information-accumulation theory of speeded categorisation. *Manuscript submitted for publication.*

Lamberts, K. (in press). The time course of categorisation. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Lamberts, K. & Brockdorff, N. (1997). Fast categorisation of stimuli with multivalued dimensions. *Memory & Cognition*, 25, 296–304.

Lamberts, K. & Chong, S. (1997). Dynamics of dimensional weighting in categorisation. *Manuscript submitted for publication.*

Lamberts, K. & Freeman, R. P. J. (1997a). Building object representations from parts. *Manuscript submitted for publication.*

Lamberts, K. & Freeman, R. P. J. (1997b). Categorisation of briefly presented objects. *Manuscript submitted for publication.*

Lewandowsky, S. (1995). Base-rate neglect in ALCOVE: A critical reevaluation. *Psychological Review*, 102, 185–191.

Locke, J. (1706). *An Essay Concerning Human Understanding.* London, Dover, Fifth edition.

Logan, G. D. (1988). Toward an instance theory of automatisation. *Psychological Review*, 95, 492–527.

Luce, R. D. (1963). Detection and recognition. In Luce, R. D., Bush, R. R., & Galantner, E., (Eds.), *Handbook of Mathematical Psychology*, (pp. 103–189). New York, Wiley.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organ-*

isation. Oxford, Oxford University Press.
Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorisation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 288–301.
Maddox, W. T. & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorisation. *Perception & Psychophysics*, 53, 49–70.
Maddox, W. T. & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 795–817.
Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 539–555.
Malt, B. C. & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*, 10, 69–75.
McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
McClelland, J. L. (1991). Retrieving general and specific information from stored knowledge of specifics. In *Proceedings of the Third Annual Meeting of the Cognitive Science Society*, (pp. 170–172).
McClelland, J. L. (1993). Towards a theory of information processing in graded, random, and interactive networks. In Meyer, D. E. & Kornblum, S., (Eds.), *Attention & Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence and Cognitive Neuroscience*, (pp. 103–189). Cambridge, MA, MIT Press.
McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, part I: An account of basic findings. *Psychological Review*, 88, 357–407.
Medin, D. L. (1975). A theory of context in discrimination learning. In Bower, G. H., (Ed.), *The Psychology of Learning and Motivation, Vol. 9*, (pp. 263–314). New York, Academic Press.
Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 37–50.
Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 333–352.
Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Human Learning and Memory*, 9, 607–625.
Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
Medin, D. L. & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 241–253.
Mervis, C. B. (1980). Category structure and the development of categorisation. In Spiro,

R., Bruce, B. C., & Brewer, W. F., (Eds.), *Theoretical Issues in Reading Comprehension.* Hillsdale, NJ, Erlbaum.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, 7, 283–294.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104–114.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, 38, 415–432.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorisation relationship. *Journal of Experimental Psychology: General*, 115, 39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorisation of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87–109.

Nosofsky, R. M. (1988). Similarity, frequency and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 54–65.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorisation. *Perception & Psychophysics*, 45, 279–290.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 3–27.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43, 25–53.

Nosofsky, R. M. & Kruschke, J. K. (1987). Investigations of an exemplar-based connectionist model of category learning. In Medin, D. L., (Ed.), *The Psychology of Learning and Motivation, Vol. 28*, (pp. 207–250). New York, Academic Press.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 211–233.

Nosofsky, R. M. & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300.

Nosofsky, R. M. & Palmeri, T. J. (in press). Comparing exemplar-retrieval and decision bound models of speeded perceptual classification. *Perception & Psychophysics.*

Posner, M. I. (1978). *Chronometric Explorations of Mind.* Hillsdale, NJ, Erlbaum.

Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipies in C: The Art of Scientific Computing.* Cambridge, Cambridge University Press, Second edition.

Pylyshyn, Z. W. (1979). Do mental events have durations? *Behavioural and Brain Sciences*, 2, 277–278.

Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA, MIT Press.

Ratcliff, R. & van Zandt, T. (in press). Comparing connectionist and diffusion models of reaction time. *Psychological Review.*

Reed, S. K. (1972). Pattern recognition and categorisation. *Cognitive Psychology*, 3, 382–407.

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.

Robbins, D. (1970). Stimulus selection in human discrimination learning and transfer. *Journal of Experimental Psychology*, 84, 282–290.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In Moore, T. E., (Ed.), *Cognitive Development and the Acquisition of Language*, (pp. 111–144). New York, Academic Press.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7, 532–547.

Rosch, E. (1978). Principles of categorisation. In Rosch, E. & Lloyd, B. B., (Eds.), *Cognition and Categorisation*. Hillsdale, NJ, Erlbaum.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 491–502.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage in the brain. *Psychological Review*, 65, 386–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In McClelland, J. L. & Rumelhart, D. E., (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, (pp. 318–362). Cambridge, MA, MIT Press.

Rumelhart, D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception, part II: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.

Schwanenflugel, P. J. & Rey, M. (1986). The relationship between category typicality and concept familiarity: Evidence from spanish- and english-speaking monolinguals. *Memory & Cognition*, 14, 150–163.

Shanks, D. R. (1990). Connectionism and the learning of probabilistic concepts. *Quarterly Journal of Experimental Psychology*, 42A, 209–237.

Shepard, R. N. (1958). Stimulus and response generalisation: Tests of a model relating generalisation to distance in psychological space. *Journal of Experimental Psychology*, 55, 509–523.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210, 390–398.

Shepard, R. N. (1987). Toward a universal law of generalisation for psychological science. *Science*, 237, 1317–1323.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorisation of classifications. *Psychological Monographs*, 75, 1–41.

Smith, E. E. & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA, Harvard University Press.

Sorabji, R. (1972). *"De memoria et reminiscentia", Aristotle on Memory*. Providence, RI, Brown University Press.

Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donders' method.

In Koster, W. G., (Ed.), *Attention & Performance II: Synergies in Experimental Psychology, Artificial Intelligence and Cognitive Neuroscience.* Amsterdam, North Holland Press.

Sternberg, S. (1969b). Memory-scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 57, 421–457.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, 27, 1–32.

Thorndike, E. L. (1913). *Educational Psychology, Vol. II, The Psychology of Learning.* New York, Teachers College, Columbia University.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.

Uhl, C. N. (1964). Effect of overlapping cues upon discrimination learning. *Journal of Experimental Psychology*, 67, 91–97.

Vosniadou, S. & Ortony, A. (1989). *Similarity and Analogical Reasoning.* Cambridge, Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical Investigations.* Oxford, Blackwell.

Woodworth, R. S. (1938). *Experimental Psychology.* New York, Holt.

Woodworth, R. S. & Schlosberg, H. (1954). *Experimental Psychology, (Revised Edition).* New York, Holt.