THE REPRESENTATION OF CONCEPTS:
AN EVALUATION OF THE ABSTRACTIVE AND EPISODIC PERSPECTIVES

By

BRUCE WILLIAM ARTHUR WHITTLESEA, B.A., M.A.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

THE REPRESENTATION OF CONCEPTS

DOCTOR OF PHILOSOPHY (1983)
(Psychology)

McMaster University
Hamilton, Ont.

TITLE: The Representation of Concepts: An Evaluation of the Abstractive and Episodic Perspectives

AUTHOR: Bruce W. A. Whittlesea, B.A. (University of Toronto), B.A. (Hon. Psych.) (University of Windsor), M.A. (University of Windsor)

SUPERVISOR: Professor L. R. Brooks

NUMBER OF PAGES: xii, 202

ABSTRACT

The field of concept formation has been dominated until recently by the abstraction perspective, which holds that categories are mentally represented by abstract summaries of their members. Two variants of this view are the prototype models, which employ singular, central representation via an abstracted central tendency, and strength models, which represent categories through abstracted counts of the frequency of their members' feature compounds. In conflict with these notions are instance models, which reject summary representation in favour of separate encodings of individual experiences of category members. The three types of models make similar generalization predictions in stimulus domains whose density is greatest near the central tendency, but make importantly different predictions in other domains.

The assumptions of abstraction models regarding the representation of variability and contingency relationships of stimulus features were formalized, and a variety of models differing in the complexity of their assumptions were tested, employing perceptual identification, recognition and categorization tasks. Models based on traditional assumptions of the prototype perspective could not account for the variety of generalization patterns obtained, while the assumptions of models which were successful in accounting for the data were argued to violate the cardinal prototype values of economical and summary representation.

A new instance model, the "episode model", was proposed. This model was found to account successfully for a wide variety of patterns of generalization in a variety of domains of differing density, through parallel processing of

multiple prior encodings. An important aspect of the model is its emphasis on the degree of integration of prior encodings, which is held by the model to determine the breadth of generalization of performance supported by prior episodes. This aspect of the model reflects its concern with the effects of processing differences on performance.

One class of feature-frequency models was also found to be capable of accounting for the patterns of results. However, the instance account was argued to be preferable on grounds of economy and simplicity of representation, sensitivity to processing context and differences in processing, and heuristic utility in directing attention to important adaptive abilities of the organism.

ACKNOWLEDGEMENTS

A thesis is rarely the exclusive product of one individual's imagination, and this one is no exception. Many of the ideas in this paper are the common property of the members of the Brooks and Jacoby labs, having been extensively developed through lab discussions and parallel research programmes. However, I also wish to acknowledge the direct contributions of individuals to the present work. I have particularly to thank Dr. Lee Brooks for his support, his thoughtful guidance and most of all for his friendship. He has been an unfailing source of theoretical insights and methodological suggestions. I have profited greatly from his experienced judgement and enthusiasm for research. Second, I wish to thank Dr. Larry Jacoby, who has been intimately concerned with the present work from its beginning, and whose own research provided many of the tools and intuitions exploited in this paper. I particularly owe to him the breadth of my current perspective on the interrelatedness of memory, perception and concept formation. I also wish to thank my lab mates, including Dr. Gordon Hayman, who anticipated many of the conclusions of this thesis, and provided me with much diverse information; Paul Gabel, to whom I am in debt for the encoding task used in this paper, and for many useful arguments; Dawn Witherspoon, who never yielded a point unnecessarily, and kept me on the straight and narrow; and Dr. John Vokey, whose parallel work sponsored many stimulating discussions.

However, my greatest debt is to Gloria, who brought me balance. She has given me comfort and support, entertainment and love. More than a wife, she has been my friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Following Page

CHAPTER 1

Introduction

I Structure and Representation of Natural Categories

It is generally agreed that the hallmark of "knowing a concept" is the ability to deal with classes and members of classes. People are evidently able not only to deal with events as individual, isolated occurrences, but also able to group events, and to make judgements about membership of events in groups. Considerable controversy exists about the knowledge structures which underlie this ability. At one extreme, explanations emphasize the learning of class knowledge; that is, in forming a concept people learn some information that is generally true or typical of category members, and do not retain (or at least do not use) information about the particular events experienced. The other extreme emphasizes knowledge about the individual events. Under this perspective the knowledge structure representing the concept consists only of the various events as they were experienced; class-level information is thought not to be abstracted during concept learning. The purpose of this paper is to deny the necessity of the class-level explanation of concept representation by demonstrating the sufficiency of event-level explanation. The intention is not to deny that people ever engage in learning summary information, but rather to indicate that the automatic assumption of this level of explanation is unwarranted, and that consideration of the event level of explanation has important heuristic advantages.

Much of the recent work employing the class-knowledge explanation has been motivated by the common observation that people think that the members of many natural concepts differ in how well they fit their category. For example, people generally regard robins as better examples of the class of birds than are ducks and geese, while hawks are of intermediate goodness as examples; similarly deer are rated as better examples of the class of mammals than bears, and bears better than pigs (Rips, Shoben and Smith, 1973). In general, people respond in a graded fashion to members of categories. This phenomenon is extremely reliable, and has been encountered in a wide variety of categories (Rosch, 1973).

Rosch and her colleagues (Rosch 1973, 1975, 1977; Rosch, Simpson and Miller, 1976; Mervis, Catlin and Rosch, 1976) utilized this observation in promoting a perspective that had previously had little impact on the concept formation literature. They pointed out that the world does not present to the concept learner a uniform, unstructured set of stimuli which can only be divided arbitrarily. The objects which a learner encounters can generally be thought of as possessing a cluster structure, consisting of the clustering of the objects into non-arbitrarily separate groups. This structure afforded by the world is abstract: it is not a property observable in any event, but arises in an overview of successive events. Such abstract structure could be a major resource for concept learners sensitive to it, since if the world already contains natural groupings, then the most efficient way for the learner to divide the world may be along the naturally-occurring divisions. In support for the notion that people actually employ this resource in concept formation, Rosch noted that natural concepts, like "bird" and "chair", parallel natural groupings of objects in the world. She speculated that, in general, categories arise to reflect the information made available by groupings implicit in the stimulus set (Rosch, 1977). Her argument appears to be that natural category formation capitalizes on the clustering of encountered objects to evolve maximally discriminable categories, by grouping similar objects together and placing dissimilar objects in separate categories.

i) The Structure of Natural Categories

Rosch's work contains the suggestion that correlation of attributes or dimensions is the principle responsible for the multiple-cluster structure underlying natural groupings. The present author contends that while correlation of attributes and dimensions is an important aspect of structure, as discussed below, it is not the principle responsible[4] for the structural separation of items into non-arbitrarily-defined groups. Instead, this separation is effected by the relative frequency of values on each dimension of the stimuli, or more specifically, the number of modes on the separate dimensions of the stimulus space. Fig. 1 illustrates eight stimulus domains whose constituent dimensions are uniform, uni-modal or bi-modal, and either uncorrelated (Fig. 1a - d) or correlated (Fig. 1e - h). Inspection of the plots demonstrates that multiple clusters of stimuli occur if and only if at least one of the dimensions of the domain is bi-modal, irrespective of whether the dimensions are correlated or not. (It is evident in this figure that "cluster" refers to a local, relatively densely populated area of the space, bounded and separated from other clusters by regions of relatively low concentration.) This disagreement is discussed more fully in Appendix 1. The present discussion of structure employs Rosch's insight into the non-arbitrariness of natural concepts, but explains the structural basis of this phenomenon in terms

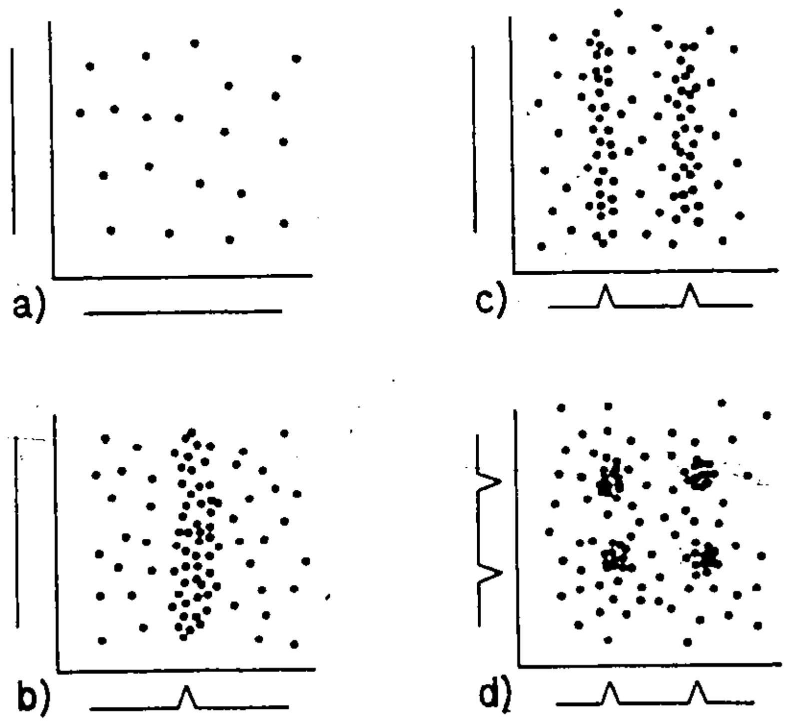


Fig. 1a. Uncorrelated dimensions.
a) uniform - uniform; b) uniform - unimodal; c) uniform - bimodal; d) bimodal - bimodal.

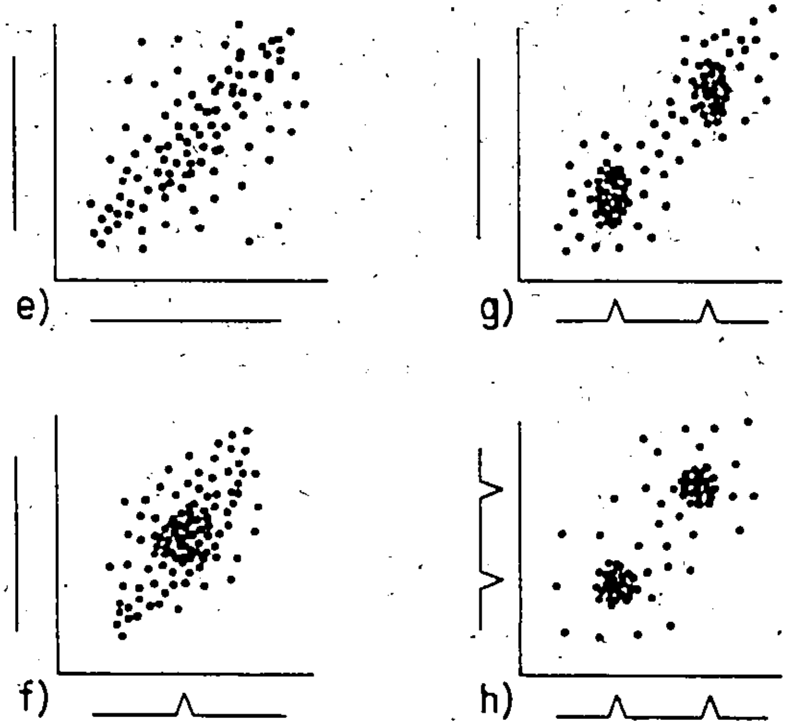


Fig. 1b. Correlated dimensions.
e) uniform - uniform; f) uniform - unimodal; g) uniform - bimodal; h) bimodal - bimodal.

of the distribution of modes of individual stimulus dimensions. It also employs examples from natural domains whose features may be best considered to vary qualitatively rather than quantitatively. As explained in Appendix 1, a different logic applies to similarity spaces defined from qualitatively-differing features, which has impact on the conceptualization of similarity in the experiments to follow. Examples employing features differing both quantitatively and qualitatively are given in this discussion.

Fig. 1h (repeated as Fig. 2 for the convenience of the reader) can be used to illustrate several important points about cluster structure. For purposes of example, the axes of this stimulus space are given as the width and height of a stimulus object. The correlation of the dimensions of this space means that if the value taken by a stimulus on one dimension is known, the value it takes on the other is to some extent predictable. A necessary consequence is that the stimuli are not uniformly distributed throughout the stimulus space; some objects (e.g. short-skinny) will occur with higher probability than others (e.g. tall-skinny). In fact, the correlation constrains items near the major diagonal of the similarity space to be most probable. This non-uniformity, attributable to the correlation of dimensions or features, is a first step toward abstract structure. Thus although correlation of the dimensions does not contribute to multiplicity of clusters, it does affect the structure of the space, and is probably usual in real-world domains. It is illustrated in natural domains by the highly contingent relation across animals between feathers and wings, and between scales and fins. The resultant non-uniformity of the stimulus domain is exemplified by the non-existence of potential feature combinations such as furry fish and flying pigs.

The bimodal distribution of values on each dimension of Fig. 2

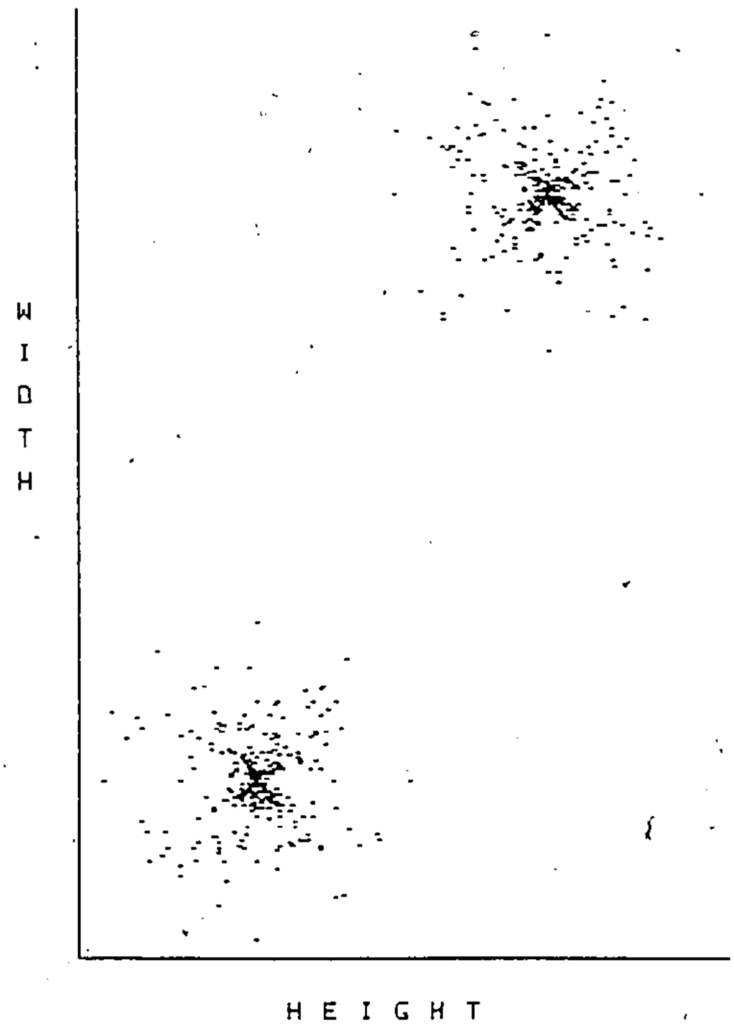


Fig. 2. Stimulus space created via correlated and multi-modal dimensions.

constrains high and low dimensional values to be more probable than intermediate values. The result is a relative lack of items in the centre of the space, dividing the items into two clusters. This relative discontinuity between clusters yields a non-arbitrary cut-point for the inclusion of an item in one category as opposed to the other. It has the effect of maximizing similarity within and dissimilarity between the categories it forms, and is in this sense the best possible cut-point. One may think of the discontinuity as due to the lack of intermediates between legs and wings, or between fur and scales.

Another aspect of structure arises because the items in a cluster are not all identical (i.e. there is within-cluster variability). The result is that items within a single category differ with respect to their centrality in the cluster, and hence with respect to their typicality for the cluster. In Fig. 2, it may be seen that some members of the short-skinny category are close to the centre of their category, while others are marginal. In effect, the members of the cluster differ in goodness of membership in the category. This graded membership is exemplified by contrasting a bat (furred, toothed, but winged) with a dog (furred, toothed, four-legged); the bat shares less features with the members of the mammal category in general than does the dog, and hence is not as good a member of the mammal category.

Rosch made an undoubted contribution to the study of concept formation by pointing out that important domains of the world present the learner with clustered events. She further documented a general correspondence between the structure of category membership and the structure of responses to category members, demonstrating that response gradients frequently parallel the graded membership of events in clusters, and that natural concepts appear to discriminate events into categories along natural discontinuities. This correspondence is important and undisputed. However, Rosch (e.g. 1977) concluded from this correspondence that people are sensitive to the structure of the world, and use it as a resource in creating categories: that in learning categories, people actually abstract the internal structure of the presented clusters. It is this conclusion that is at issue in this paper. The following section presents the conflicting prototype and instance perspectives on the memorial representations underlying the correspondence noted by Rosch.

ii) The Representation of Natural Structure

In order to clarify the contrasting claims of prototype and instance perspectives, the stimulus domain illustrated in Fig. 2 will be examined in greater detail. This space can be described in at least two ways: in detail and in summary. A detailed description would involve specification of the co-ordinates of each stimulus object. In contrast, the space could be described with little loss of information through only two summary parameters, the central tendencies and variances of the clusters. The economy and accuracy of such a summary description increases rapidly with the degree of correlation between dimensions, and as the variability around each mode decreases. Fig. 3 illustrates such a summarization. A discriminant has been dropped through the region of lowest concentration to maximize the degree of clustering within two areas of the space (i.e. orthogonally to the best-fitting regression line). The central tendency of each cluster has been calculated, and is located at the centre of the concentric circles. Further, some index of dispersion of stimuli around the central tendency has been computed,

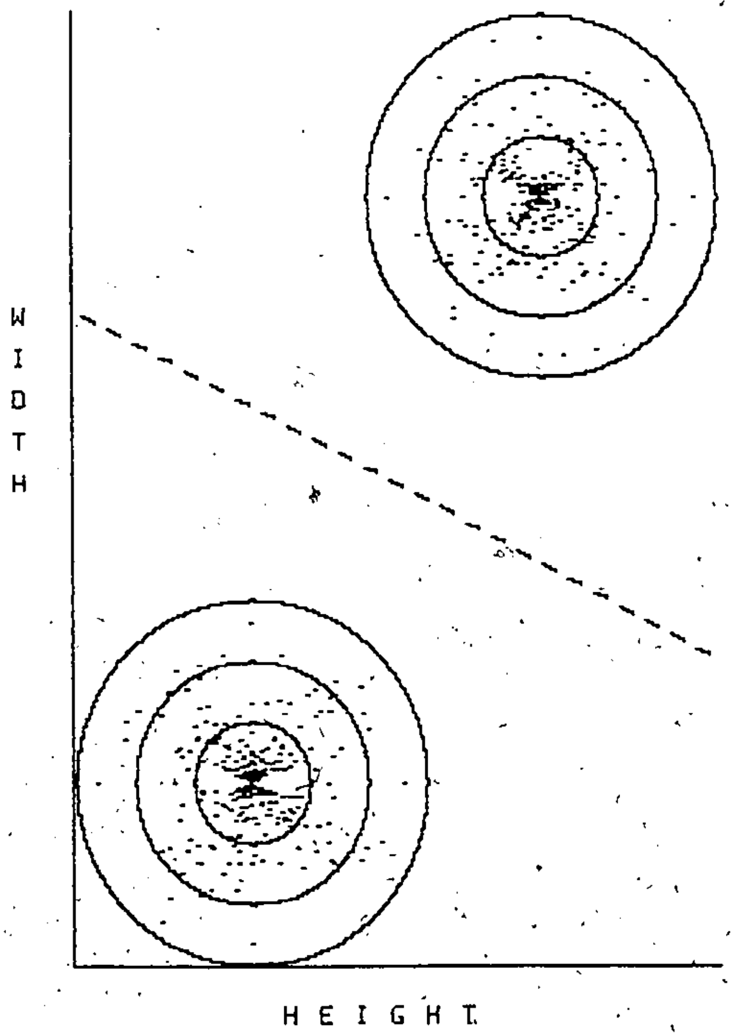


Fig. 3. Illustration of summary representation. Best-fitting discriminant (dotted line) and index of dispersion (circles) have been plotted in the stimulus space shown in Fig. 2.

and is represented in Fig. 3 through the set of circles radiating out from each central tendency. All points on a circle deviate equally from the central tendency. The set of circles thus represents a set of contours forming a gradient of dispersion. Each point in the space can now be described in terms of its deviation from a central tendency. (The amount of deviation can alternately be termed the distance or dissimilarity of a stimulus from the central tendency; these terms are used interchangeably below).

A learning device capable of categorizing new stimuli in this height-width domain could proceed in a variety of ways. It may employ rules or similarity as the basis of classification: that is, it may either deduce an algorithm from its experience with category members, and apply this algorithm to novel stimuli, or alternatively may use an analogy mechanism. The algorithmic alternative will not be dealt with in this paper; it has been extensively criticized elsewhere (e.g. Rosch, 1973, 1975; Brooks, 1978; Vokey and Brooks, Reference Note 6; Brooks, Reference Note 1). Both prototype and instance theories assume that the act of categorization is based on a similarity comparison of the target stimulus with material in memory other than a diagnostic rule.

Assuming an analogy mechanism, the learning device could classify targets using information about all or about only some of the category members to which it has been exposed. Secondly, the device may possess only a single standard for classifying all targets in a category, or multiple standards to select from for particular targets. Thirdly, the information it uses may be an ahistorical summary of events experienced, or consist of unsummarized representations of events which have actually occurred. (This distinction parallels Tulving's (1972) semantic - episodic distinction.) A wrinkle in this latter distinction is that depending on processing assumptions, a summary may be either an average of presented events, in which case it need not be identical to any event which has actually been experienced, or may be a most typical event which the system has actually experienced (Neumann, 1974, 1979). In the former case, the representation of the concept is generated, in the latter case it is merely identified. For introductory purposes, prototype and instance perspectives will be contrasted on all three of these distinctions.

Simple versions of prototype theory (e.g. Rosch, 1977) posit that people are sensitive to the structure of the world, and thus able to discover the clusters in presented stimuli. Each cluster discovered is memorially represented via a single summary, abstracted out of the mass of experiences, utilizing information about all category members experienced. In terms of Fig. 3, this simple prototype view would claim that the short-skinny category is represented via a memorial standard, identified with the stimulus which occupies the central tendency, whether or not that stimulus has been experienced by the subject. The central tendency of the characteristics of all stimuli in a cluster represents the typicality structure of the cluster with great economy and little loss of information, as noted above. Prototype theory asserts that people take advantage of this fact in employing central tendencies as informationally economical representations of the graded membership structures of the world.

Rosch suggested that abstraction of the central tendency may occur in one of two ways. Where particular events are of high <u>perceptual salience</u> (e.g. wavelengths of light for which the visual system is most sensitive) they become focal points around which other events are organized. Alternatively, where events do not differ in perceptual salience, the most <u>typical</u> event in a cluster emerges as the focal point of organization through an averaging procedure. (This latter case suggests evolution of the concept over time, which the former does not.) In either case the memorial representation of the concept is a prototype: that is, the most central (salient or typical) member of the category. In terms of the distinctions outlined above, this prototype represents information generally true of all experiences of category members (being computed as the average of them all), is a singular, summary form of representing this information, and is generated if not presented.

As a cognitive structure representing the concept, the prototype notion accounts for the correspondence between the structure of the world and the structure of categorical judgements in a very direct way. When required to classify an event, one compares it to prototypes in memory, and assigns it to the category of the nearest prototype. The particular experiences a person receives have no direct impact on the classification procedure beyond defining the central tendency. Classification using such a heuristic has a number of virtues: It works for novel items as well as it does for items actually experienced; it is likely to be highly accurate, since in clustered spaces items which are similar are likely to be in the same category; and it requires only a simple and informationally economical representation of the concept in memory. Moreover, it explains the observed gradation of responses to categories, in that speed, accuracy and confidence of the judgement can be expected to covary with the similarity of a probe to the nearest prototype. This explains why people exhibit graded responses like "robins are better examples of birds than are penguins": robins are, in their experience, more typical of birds in general than are penguins, and hence more similar to their prototypical conception of a bird.

The observation of a parallel between response gradients and gradients of membership in presented stimulus clusters has been a basic motivating phenomenon of the prototype tradition. The classic evidence on which the prototype perspective rests consists of variations on this theme. For example, where the most typical event is known to the experimenter (as in artificial domains), response gradients have been shown to peak at the most typical event. Moreover, even when the instance that forms the category centre has not been presented, no instance is rated faster than the central instance (Posner and Keele, 1968; Franks and Bransford, 1971; Rosch, 1977). These findings led Rosch to conclude that the internal structure of categories (i.e. the graded membership structure) is known to subjects and affects their categorical judgements; further, that this knowledge is represented mentally via the prototype of the category; and that if this prototype is not presented, it will be generated by subjects. The general line of evidence for these conclusions is the correlation of response measures with distance of a probe from the prototype. However, while this evidence is certainly consistent with the notion of prototypes, it is indirect, consisting of an observed correspondence between the structure of the world and the structure of people's <u>categorical judgements</u>, not their <u>cognitive structures</u>.

Brooks (1978) denied the necessity of concluding from such evidence that prototypes are abstracted and form the memorial representation of the concept. He pointed out that instance models (e.g. Brooks, 1978, Medin and Schaffer, 1978) can also predict graded reactions to members of a category, based on the same clustered structure of experience. A common instance approach suggests that a concept is represented mentally via encodings of those instances of the category the person has experienced, but that the memor ' standard applied to a given target consists of information about less than all category members. This information consists of particular instances of the category, all of which have actually been experienced. The specific experiences a person receives are thus directly involved in the classification procedure. In a simple instance model, classification might involve comparing a stimulus to the most similar encoded instance: the greater the similarity, the greater the likelihood of classifying the stimulus in the same category, and the greater the speed, accuracy and confidence of the classification. By this view, robins are reacted to more readily than penguins because robins are more frequent in the person's experience; hence the set of encodings accessed to perform the judgement for robins is likely to contain more or closer analogies than the set accessed for penguins.

Under the assumptions of this simple instance notion, the learner simply encodes and stores his experiences of each instance presented, and does not generate any typicality information. In terms of Fig. 3, subjects may have been exposed to and consequently encoded stimuli of varying typicality. When asked to classify a stimulus, the learner might compare the current encoding with the most similar priorly encoded instance. The resultant accuracy and speed of classification would vary with the degree of analogy between current and resurrected encodings. Such a system would account for the relevant data for precisely the same reason as the prototype notion: the stimulus space is clustered. In such spaces, the majority of prior encodings is likely to lie near the central tendency (i.e. near the innermost ring in Fig. 3). As can be seen in the figure, instances not close to the central tendency are likely to be both rarer and farther apart. Thus a target instance which is itself close to the central tendency is also close to many prior encodings. The high similarity of the target to its nearest neighbour will result in highly confident classification, and also in high accuracy, since in clustered domains highly similar items are usually in the same category. Conversely, a target far from the centre is unlikely to be close to any prior encoding. The low similarity of the target to its nearest neighbour leads to low accuracy and confidence. Thus the learner doesn't necessarily directly code the graded structure of the events he experiences. However, this structure biases the availability of stored instances, so that he has more coded near the central tendency. As a result, his responses are graded. Thus the instance perspective predicts exactly the same outcome as prototype theory, namely a high correlation between response measures and distance of target from the central tendency, but the assumed underlying mechanism is utterly different.

Summarizing the two perspectives, prototype theorists base their claim of an abstracted prototype on the correlation of response measures with distance of a target from the category centre. Instance theorists claim this correlation arises only because of the correlation between response measures with distance of a target from instances actually experienced, the average of which is the prototype. The perspectives essentially disagree on whether it is the subject or the experimenter who computes the prototype: whether the general organization of responses is a property of the subject, due to his abstracting the structure of the category, or instead a property of the experimenter, who knew the structure of the category and recognized it in the response gradients. Thus the issue to be examined in this paper is whether the internal structure of the category is actually abstracted by subjects, is represented directly via a singular, central, summary prototype and is evidenced in behaviour as a result of this prototype guiding decisions; or has not been abstracted, is represented via multiple, local, particular experiences of instances, and is evidenced only as an emergent phenomenon.

The evidence thus far described relies upon comparing response measures to probes differing in typicality. As has been argued above, this manipulation is insensitive to the differences between the competing theories. In the experiments to be described below, the basic tactic employed to test these alternatives was to contrast response measures to targets differing not only in typicality, but also in distance to priorly-encoded instances. Fig. 4 exemplifies this tactic, using the dispersion-ring graphic convention adopted in Fig. 3. The various rings represent locations of potential members of the category. The numerals on the second ring indicate stimuli which have been exposed to the subject in training. In this very simple domain these are the only past experiences of the concept the subject has on which to create a basis of classifying novel items. These training stimuli are all two similarity units from the central tendency (symbolized P for prototype). This central stimulus has not been exposed, nor have the probes (A - D). Probe A is located one deviation unit from both the prototype and stimulus 1; probes B and C are both two units from the prototype, but are respectively one and two units from stimulus 1; probe D differs from the prototype by three units, and from stimulus 1 by one unit; and stimulus 1, used as a probe, is two units from the prototype, but zero units from itself.

Prototype and instance theories can now be seen to make contradictory predictions. The simple version of prototype theory suggests that the central tendency will be abstracted out as a prototype during exposure to the training stimuli, and will represent the concept. Thus it must predict that stimuli most resembling the prototype will be reacted to best. This enables us to predict an

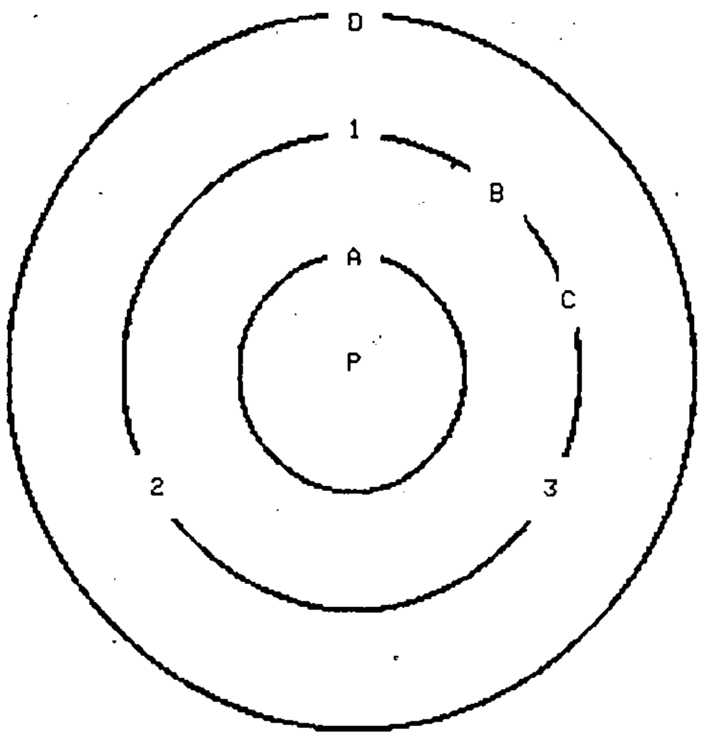


Fig. 4. Example of stimulus space in which prototype and instance predictions differ. Training stimuli are numbered (1 - 3) while novel probes are lettered (A - D).

ordering of responses to probes: the prototype will be responded to best, then A, then 1, B and C in a dead tie (since they are equally typical), followed by D (the least typical). In contrast, a simple version of instance theory claims that during learning trials, the stimuli 1, 2 and 3 would be encoded comparatively literally (as experienced, not in summary), and that classification would involve a comparison of a probe only with the most similar of these encoded stimuli. In the present case, all probes would be compared with the memorial trace of stimulus 1. The predicted ordering of responses to probes is now stimulus 1 reacted to best (since it is compared to an almost exact copy of itself), followed by A, B and D in a dead tie (since all are equally similar to 1), followed by C and the prototype – a pattern very different from that predicted by the prototype view. These sets of predictions are clearly necessitated by the three processing assumptions of each perspective discussed above. Finding one of these patterns of response would place the opposing theory in difficulty.

It may appear surprising that the two theories can so easily be made to disagree, after the long prologue above arguing that in general they predict the same transfer patterns. The difference lies in a confounding factor, density. Ordinarily, in clustered spaces, the prototype is likely to have more and closer neighbours than any other probe (see Fig. 3): that is, the space is densest in the centre, and has least density on the perimeter. A cross-section of such a space yields a roughly normal distribution of density. It is only with density distributions like this that peak at the centre that the two theories make generally similar predictions. Such distributions may be quite common (a plot of frequency by similarity of types of birds probably has a cross-section something like this), and density itself may be thought to be an important impetus to abstraction (in the sense that storing all instances of a populous domain may be a heavy cognitive load). But the distribution of density of the domain is not a part of the assumptions of either theory: a normal distribution of density is unnecessary for good cluster structure. The topic of distribution of density will be discussed more fully later in this paper.

II Notions of Abstract Structure

i) Traditional Abstraction Models

The foregoing section introduced a simple version of prototype theory. The intention was to capture the motivating concerns of abstraction accounts of concept formation rather than any of the particular formulations of the theory, which are various. This section attempts to portray the range of notions subsumed under the rubric of abstract categorical representation. The following section presents a typology of the major alternative abstraction assumptions, attempting to organize the diversity of notions presented in this section.

Although the notion of abstract representation of concepts goes back at least to Plato's pure forms, the modern abstraction tradition can be considered to begin with Bartlett (1932), who introduced the idea of "abstract schema" as an important property of the cognitive system. He was concerned with observations which suggested that people do not remember precise details of their experience, but rather form a summary representation of the gist of their experiences. He emphasized the importance in recall of the production of details consistent with the summary schema, as opposed to retrieval of stored detail. This theme of summary representation is fundamental to the most popular class of notions of abstract representation of concepts, the distance models.

More recently, a second class of models of abstract representation, the strength models, has arisen in opposition to the distance models. Both classes of models have been styled "prototype" theories (e.g. Neumann, 1974; Rosch and Mervis, 1975; Rosch, 1977, 1978) despite the fact that they have little in common save for some idea of abstraction. This usage is more confusing than helpful, since it leads one to expect non-existent commonalities in the theories. In accordance with popular usage, and because the name is apt, this paper will adopt the convention of referring to abstractionist distance notions as "prototype" models, while models appealing to abstracted strength will be called "feature-frequency" models. As will be seen later, both classes are in conflict with instance models.

Posner and Keele (1968) conducted a seminal study in the prototype tradition. They were concerned with the notion that information common to a set of events is abstracted and stored. They were open to the possibilities that only this abstracted schema was stored, or that both the prototype and the events were stored and affected recognition of new items. In transfer tests, they observed that while the training items were well recognized, the schema pattern itself was as well recognized (though it had never been presented in training), and better recognized than other novel items. They cautiously concluded from these data that the prototype pattern is unique, having a special status among novel items. They did not conclude that the abstracted prototype affects the recognition or classification of new items; in fact, they drew no conclusion regarding the relative roles of the prototype and coded instances in the cognitive structure responsible for classifying novel instances. Their sole claim was that stimulus generalization from training instances is insufficient to account for the observed degree of recognition of the prototype pattern. However, in a later paper (Posner and Keele, 1970) they observed loss of recognition over a period of a week for training items, but not for the prototype pattern or other novel items. They concluded that the schema must be abstracted <u>during</u> <u>training</u>, and constitutes a major (and after a short duration the only) component of the cognitive structure representing the category.

The schema considered by Posner and Keele conforms in many ways to the description given in the foregoing section of the "simple form" of prototypes. In such models, the prototype of a category is its central tendency. Each member of the category has associated with it some distance from the central tendency. The decision rule for such models is "place an item in that category whose prototype is least distant from the item". In the Posner and Keele variant, the prototype is central in the sense of being a prime form (in their case an arbitrarily chosen dot pattern) from which the members of the category are derived as statistical distortions. Such a prototype is unlikely to be the average of its distortions unless specially constrained, which the prototypes used by Posner and Keele were not; a surprising outcome of this lack of constraint is discussed below, in Section <u>V</u> <u>iii</u> of the next chapter.

In most prototype models developed since Posner and Keele's pioneering work, the prototype is held to consist of the average values (mean or mode) of those category members which have actually been presented (e.g. Reed, 1972; Homa, Cross, Cornell, Goldman and Shwartz, 1973; Rosch, 1977; Homa, Sterling and Trepel, 1981). This was an assumption of t. "simple prototype" version outlined in the first section. It contrasts with Posner and Keele's conception in that the prototype is clearly thought to be a function of what has been experienced by the learner, rather than being a prime form preceding experience. Such average-value prototypes are clearly non-arbitrary in Rosch's (1973, 1977) sense, since they can be computed at any stage of the subject's experience as those compounds of values which have minimum distance to all presented instances. By contrast, pure-form prototypes are arbitrary at least to some degree: while the instances are derived from them, they are not derivable from the set of instances, which makes it difficult to understand how the learner is supposed to synthesize them. The present author suggests that the success of preparations employing pure-form prototypes has depended on the random creation of distortions of the prime pattern resulting in the prime pattern being not very far distant in general from the average combination of values. The two notions can of course be combined. The presented distortions of a pure original form can be constrained to average to the original form, as in Franks and Bransford's (1971) preparation.

Posner and Keele concluded that subjects learn the central tendency and variability of the patterns. However, they were deliberately vague about both the form and content of the stored information underlying this knowledge. In particular, they did not distinguish between two possible types of variability information, whether the encoded variability information consists simply of knowledge about particular instances, perhaps information about how specific instances depart from the central tendency, or whether it additionally consists of category-level knowledge, information about the typical variability of instances in general. This distinction is not trivial, since it reflects precisely the issue of whether the formation of concepts typically proceeds by the abstraction of class-level information or by encoding instance-level information.

Before discussing this issue, it is necessary to clear up a terminological problem. In many of the papers referenced above, it is frequently unclear when "the prototype" is mentioned whether the author intends to refer to the most central stimulus which could be imagined for a domain, or the substrate of information in the subject's cognitive system which is imagined to be responsible for his performance in concept tasks: that is, whether "the prototype" refers to a stimulus or a construct. The convention will be adopted in this paper of referring to the central stimulus as the "prototype pattern", and to the construct thought to underlie performance as the "prototype". This distinction is rarely clearly made in papers on prototype theory, which in general contain ill-specified representation assumptions. As will be seen shortly, it is frequently claimed that the prototype consists of central tendency and variability knowledge about the domain, without specification of what information is stored that represents this knowledge.

Among the few variants of the prototype class which do specify the type of information stored in forming a concept are the prototype-plus-transformation (Bransford and Franks, 1971; Franks and Bransford, 1971) and concept-plus-correction (Reitman and Bower, 1973) models. In the studies supporting these models, as in the Posner and Keele studies, instances were created as distortions from a prime form, variously a compound linguistic proposition, geometric form or four-tuple of letters and numbers. The measure of distance consisted of the number of operations required to transform an instance into the prototype. Franks and Bransford observed that recognition ratings were inversely related to an item's transformational distance from the prototype. They proposed that the mental representation of the concept consists of the abstracted central tendency and additionally abstracted representations of the transformations necessary to produce the exemplars. Knowledge about the transformations was thought to be general: the subject was considered to be learning about possible transformations, not the particular concatenation of transformations that defined an individual presented instance.

However, while most contemporary prototype notions maintain an emphasis on abstraction of knowledge about the central tendency, few of them make any explicit assumptions about the abstraction of variability information. In some papers which do make reference to the learning of variability, it is unclear whether the abstraction of class-level or instance-level variability knowledge is intended (e.g. Rosch, 1977; Homa, Sterling and Trepel, 1981; Omohundro, 1981). Frequently the whole issue of whether and what variability learning takes place is obscured by a blanket reference to the abstraction of "abstract category information" or "general information" (e.g. Homa and Vosburgh, 1976; Robbins, Barresi, Compton, Furst, Russo and Smith, 1978; Omohundro, 1981). The intent of many of these papers is to examine the necessity of class-level explanation. In doing so, they typically attempt to disprove the sufficiency of an instances account, through demonstrating that at least some of the transfer obtained in experiment is accountable in terms of distance from the central pattern. However, they are frequently vague about the nature of the prototype responsible for this outcome; neither indicating clearly whether variability knowledge is learned, nor specifying the type of information by which the prototype represents central tendency and/or variability knowledge. For this reason, a variety of conventional and unconventional prototype models are tested below, employing a variety of assumptions regarding the abstraction and representation of knowledge about the stimuli.

Rosch (1977) distinguished two modes of prototype representation. She typified the average-value prototype as "digital", characterizing digital representations as having the property of storing a set of dimensional values or features. By contrast, "analog" prototypes would be thought to store an abstract image of a typical category member. Such analog prototypes would not code a list of typical item components, but rather consist of a gestalt. Psotka (Reference Note 5) has given an analogy which clarifies this notion: an analog prototype is like the image that would form on the retina if one looked through a stack of photographic transparencies, all of members of some category. For example, if one looked through a stack of 100 transparencies of dogs in profile, the resultant image would be typical of the group. Features common to many of the individuals would be accentuated, while atypical features would tend to be lost. Rosch's motive in suggesting analog representation appears to have been that in some domains, such as dot patterns, stimuli cannot easily be summarized in terms of verbalizable common elememts. Such propositional inefficiency would appear to count against the simplicity and centrality of prototype representation, and would tend to limit prototype representation to domains of easily verbalized commonalities. However, a prototype consisting of an analog image of a central pattern carries the virtues of prototype representation to other domains. Such analog prototypes are apparently not to be taken as literal templates, with the problems that would entail in terms of image justification (see e.g. Neisser, 1967).

Digital and analog prototypes may differ in their similarity (distance) computations. Deviation from a digital representation is usually assessed in terms of an additive combination of the independent differences between component elements of the representation and the probe (i.e. components treated as separable). However, to the extent that analog implies a pattern of elements, such that relations among elements are an integral part of the representation, the deviation from an analog representation must be assessed from the resemblance of whole images (i.e. components treated as integral). To paraphrase the Gestalt motto, the difference between wholes differs from the sum of differences of their parts. The result is that analog and digital prototype models may make conflicting predictions regarding the similarity of a probe to the prototype and its consequent facilitation. The consequences of this point are discussed in Section IV of the next chapter.

Prototype models are to be distinguished from another class also employing distance as the unit of structure, the average distance models. Under the assumptions of this class, the learner assigns a probe to a category if the average distance from the probe to all patterns in the category is less than the average distance from the probe to all instances of another category. This kind of model has been tested as an alternative by Reed (1972) and Hayes-Roth and Hayes-Roth (1977), employing both city block and Euclidean metrics (separable and integral features, in Garner's (1974) terms), but has not been favoured as an account of concept formation. The main value of the average distance notion has been in clarifying distance models of prototypes, from which it differs in several regards. First, prototypes have been traditionally allied with the city-block metric (although this has been criticized by (among others) Robbins, Barresi et al, 1978). Second, even assuming this metric, prototype and average distance models are not identical: distance to the average pattern of a category is not the same as average distance to the patterns of a category, as pointed out by Reed (1972), although (as he failed to point out) they are, with a few exceptions, monotonically related. A more important difference is that the average distance class requires retention of the presented items and on-line computation of the classification criterion; that is, no summary of the category members that could aid classification of a probe is pre-computed. By contrast, prototype notions specify pre-computation of a standard in the form of a synthesized prototype, requiring only a single on-line computation (distance of probe to prototype) at the time of classification. Thus although average-distance and prototype-distance models are occasionally confused, they have importantly different assumptions. In fact, the average distance models can be regarded as closer to the instance than the prototype perspective, since they are based on similarity to the particular patterns experienced, and not to the central pattern.

The second major class of abstractive models, the feature-frequency or strength models, is exemplified by Neumann's (1974, 1977) attribute-frequency, Hayes-Roth and Hayes-Roth's (1973, 1977) property-set and Reitman and Bower's (1973) tag models. The major differences between prototype and strength models lie in their assumptions of what is stored during learning, and in the consequent comparison leading to facilitated classification of a probe. Prototype models in general emphasize the storage of the average (mean) of each dimension on which stimuli vary; the facilitation experienced by a probe is a function of its similarity to this average on each dimension of variation. Strength models, by contrast, assume that the memorial representation consists of registers of the frequency of components of presented stimuli; the facilitation of a probe is thought to be a function of the frequency of its components. Thus, as pointed out by Neumann (1977), one contrast between strength and distance models is in terms of which measure of central tendency is considered appropriate, the mean or the mode of the dimensional values of presented items. Where these values fail to coincide, the models make different predictions. As Neumann put it, ". . . the prototype-distance model predicts that, if the experienced values form a circle in a two-dimensional similarity structure, the prototype will be in the centre of that circle, whereas the attribute-frequency model predicts that the best recognized stimulus must lie on the circumference" (p. 187). While important, this contrast is only relevant to stimulus spaces consisting of continuous dimensions, since in featural domains no dimensional mean can be computed. However, it does point to another difference, the preference under strength models to consider concepts to be bundles of discrete features, whereas prototype notions are more comfortably applied to continuous dimensions.

Another difference is that prototype models may code only the central tendency or additionally variability knowledge, whereas strength models always indirectly code variability information, by virtue of coding the frequency of all presented features.

A greater difference between distance and strength notions lies in the level of feature compounds considered. Traditional distance models count the similarity between two items as the number of individual features in common. For example, the items "ABC" and "ABX" have two features in common, "A" and "B". First-order feature-frequency models of the kind rejected by Franks and Bransford (1971) operate in this fashion. After being presented with these two stimuli, the cognitive system would be thought to have a count of two for each of "A" and "B", and one each for "C" and "X". However, higher-order feature-frequency models like those of Neumann and Hayes-Roth and Hayes-Roth count not only these individual features but also higher-order compounds. Under these assumptions, the cognitive system would be supposed additionally to have counts of one each for "ABC" and "ABX", one each for "BC", "BX", "AC" and "AX", and most importantly, two for "AB". Thus while prototype models typically treat stimulus components as independent, strength models typically treat them simultaneously at all levels of compounds.

The decision rule for classifying items and the basis of recognition confidence are the least worked-out parts of strength formulations. Hayes-Roth and Hayes-Roth (1977) suggest that recognition is a function of the strength (the accumulated frequency) of the whole set of compounds constituting an item, while classification is determined by the single compound most differentially associated with the categories.

Conceptually, prototype and feature-frequency models are far apart. Feature-frequency models generally consider the memorial representation of the concept to consist of multiple stores, and to encode all information presented to the subject. They are uneconomical of storage, and rely on on-line processing. They are abstractive in a sense very different from Bartlett's conception of picking up the gist: rather they extract and store all possible permutations of detail, effecting a thorough analysis of presented information. They appear to be more in the associationist tradition than the cognitive, treating the nominal stimulus as perfectly predictive of the functional stimulus, without consideration of variations in the functional stimulus due to processing differences. They do not directly code the central tendency, and contain no singular, central representation. Rather, the representa ~n of the concept consists of the whole set of feature and feature-compound frequencies. By contrast, prototype notions emphasize singular, central representation, with storage of only a summary of presented information. Thus they are economical of storage, and also of on-line work, since much of the information required to accomplish a classification is pre-computed. They are quite congruent with Bartlett's conception of abstraction.

Prototype and strength notions are similar in some regards; however. Neither is much concerned with repetition of items. Distance models reflect differential item frequency through the correlation of dimensional values within a category, but offer no special status to repeated items per se. The central tendency and variability remain constant under repetition of a whole set of items, and thus the abstracted prototype would be expected to be identical under a single or repeated presentation of the whole set of items. Strength models do have a particular routine for repeated items, which is that only on exact repetition is the counter for the highest-level compound augmented. However, this is but one counter among many, without special prominence. This treatment of repetition will be contrasted later in the paper with that offered by the instance perspective, which, as is indicated below, suggests that repetition, among other variables, may be expected to have an effect on what is encoded when an item is presented. Unlike the abstractionist notions, an instance model to be offered below suggests that encoding is flexible, such that the system may, as a result of factors such as repetition, encode only the highest-level or only the lowest-level compounds of features at a given point in its experience with the set.

Generally, distance and strength notions agree that the best-recognized item may be one that the learner has never seen before. Although this issue is still very much alive (Robbins, Barresi et al., 1978, but contrast Hintzman and

Ludlam, 1981) in terms of when and how a never-experienced prototype emerges, most distance theories (e.g. Omohundro, 1981; Homa et al, 1981) assume that learners can generate the prototype, not merely select it from among presented instances, thus permitting the prototype pattern to be best recognized, even if it is a novel item. Strength theories appear to entail this phenomenon. If a novel stimulus contains the highest-frequency individual and compound attributes to a greater degree than presented items, it must be better recognized. (As an example, if ABCE, ABFD, AGCD and HBCD have been presented, then the multi-dimensional mode ABCD must be better recognized than any presented item, since the total presentation frequency of its n-tuples is greater than that of any actually presented item.) In fact, this false recognition must occur immediately, not merely after a week's delay; current strength theories depend solely on frequency counts to predict recognition. In order to account for delay effects such as those shown by Posner and Keele (1970), Homa, Cross et al (1973), and Robbins, Barresi et al (1978) they would have to include some mechanism which decreases the importance of high-level compounds over time, such that predicted recognition for re-presented training items decreases over time. This might be accomplished through the random loss of registers (a solution similar to that suggested by Hintzman and Ludlam (1980) for instance theory). Since there are fewer higher-order registers, and very few very-high-order registers, loss of registers becomes more influential at higher levels. If a stimulus ABC has been encoded, and the register for A is lost, other information at that level is still available (the registers for B and C). But if the ABC register is lost, all of the highest-order information, the information about particular instances, has been lost.

Probability models form a third general class of potential bases of concept formation. They are particularly appropriate to featural domains, although they can be applied to continuous dimensions. The principle appealed to has been called diagnosticity by Tversky (1977) or cue validity by Beach (1964a, 1964b). Fundamentally the notion is the use of Bayes' theorem to decide the category of an item. For a given feature (cue) one may compute the conditional probability that an item bearing the feature is a member of a particular category, based on the differential rates at which items in the target and contrast categories possess that cue. The process may be repeated for each category, yielding a set of conditional probabilities of category membership given that feature. The validity or diagnosticity of the feature for a category is the ratio of the conditional probability for that category to the total conditional probability of that feature for all categories. This provides an intuitively appealing principle for assigning items to categories depending on the relative number of features they share with various categories.

Many variants of this class are possible. Reed (1972) considered cases where the number of cues actually compared across categories varied from one to the number of features in an item. However, greatest attention has been paid to cue validity not as a model of concept formation in its own right, but rather as an adjunct to strength models. For example, Hayes-Roth and Hayes-Roth's (1977) property-set model contains the premise that classification of an item is determined by the item's most diagnostic feature compound.

The combination of cue validity and strength notions opens up a wide array of unexplored models. For example, if the cue validities for each feature separately are stored, the cognitive system would essentially function like a simple feature-frequency model. If it additionally stores cue validities of higher-order compounds, it would function much like a higher-order feature-frequency model. In both cases the information retained is abstracted across items, but not summarized via a central parameter, and a categorical decision requires on-line access to multiple sources of information. Such a system would yield graded responses, with confidence, speed and accuracy of classification dependent on the magnitude of the ratio of cue validities of a probe for its potential categories. However, the cue validities could also be used to define an optimal discriminant between categories (like the one illustrated in Fig. 2 above). In this case the system precomputes a categorical decision standard, and as a result, categorical decisions require little on-line work, merely a comparison of a probe to the single decision rule represented by the discriminant. Unlike the un-precomputed models, such a system would not yield graded responses: its decision rule is all-or-none (this or that side of the discriminant). However, the discriminant could be combined with central tendency and variability information in a fashion which retains both precomputation and graded responding. The model might posit that creation of categories, the decision about how many categories of what breadth and content should be employed to best capture the diversity of items, is conducted via the computation of optimal discriminants, following which the central tendency and variability of each category so formed is abstracted. This notion would explain not only the performance of subjects in assigning items to categories, but additionally how they derive the categories in the first place, in a manner congruent with Rosch's observation that natural categories appear to have formed to maximize informational cuts made available by the non-uniformity of the world.

Rosch and Mervis (1975) linked cue validity to Wittgenstein's family resemblance notion of categories. Family resemblance, as described by Rosch and Mervis, is essentially a first-order feature-frequency notion, consisting of a count of the number of items in a domain possessing each feature. Each item is assigned a family resemblance score which is the total of the counts of its component features. Under this notion, the set of items forming a category need not share any overall commonality; it is sufficient that members grade into each other, that each member is partially overlapped by another. The set of items "ABC", "BCD", "CDE" and "EFG" possess such a structure. No item can be picked out as the average, but the item "BCD" shares more features with all items in the category (has a higher family resemblance) than any other item. Rosch and Mervis made no representation assumptions about such structures. However, they combined family resemblance with cue validity, such that the effective resemblance of an item for its category was thought to be in part determined by its degree of overlap with a contrast category.

Although Rosch and Mervis did not consider the point (perhaps because they wished to avoid representation assumptions), the family resemblance structure can also be considered to be a prototype (distance) structure. Since the item "BCD" in the example of the last paragraph shares more features with its category than does any other item, it can be considered most typical, and as such may be considered to be the category prototype, with all the entailed implications for representation. Other members of the category would then be considered to bear distance relationships to this item commensurate with their overlap with it. Cue validity considerations could be added in such that the category prototype is the item bearing highest resemblance for its category and least for other categories. In both of these variants, categories would form on the principle "maximize resemblance within category and discriminability between categories", but classification would appeal only to the distance of an item from a formed prototype. In another conjunction of distance and cue validity notions, all of the distance models could be re-worked to include the ratio of distances to prototypes as a principle of category formation or classification, instead of simply appealing to the distance to the closest; however, this would cost distance models some of their simplicity.

Reed (1972) pointed out that distance and cue validity criteria can result in differential predictions. This may be seen by contrasting two stimulus spaces consisting of continuous dimensions differing only in density, in neither of which do categories overlap. In such a case the cue validities associated with the categories of the two spaces are identical, although the distance to the prototype increases with decreasing density. However, Reed failed to point out that this contrast is only true for continuous dimensions. For featural spaces, alteration of the distance of an item from the prototype affects the frequency for the category of those features altered to change the distance. This inevitably results in a corresponding change in the validity of those cues. The problem in continuous spaces pointed out by Reed is essentially due to the distance model treating the stimuli as continuous, and the cue validity model treating them as discrete and non-ordered (see Appendix 1). Thus cue validity and distance formulations are congruent in situations where both apply feature logic.

In addition to the models thus far described, there are a number of hybrid models that combine active abstraction with encoding of the presented instances. One such model considered by Posner and Keele (1968) consists of an abstracted schema responsible for classification of novel items, plus encodings of actually presented items, which are used as the referents when these items are re-presented for classification. Another is the modified ACT model of Elio and Anderson (1981), a version of strength theory which abstracts the single highest-order compound which is generally true of the members of a category, but additionally encodes actually presented items, thus coding both abstract and relatively literal aspects.

A proliferation of other models are created by weighting features (e.g. Reed, 1972 and Hayes-Roth and Hayes-Roth, 1977). Two reasons are generally offered for this procedure. One is that dimensions may differ in terms of perceptual salience: weights may be applied to reflect differential use of dimensions due to differential salience. A second reason is that discriminability between categories may be increased by differential weighting. This appears to have been part of the motivation behind the additive rule (linear regression) models proposed in the decision literature (e.g. Goldberg, 1970; Elstein and Bordage, 1978), although as Dawes (1979) pointed out, equal-weights models are robustly successful in prediction. The issue of differential weighting is discussed further in Appendix 4.

ii) A Typology of Abstraction Models

This paper is intended to assess the sufficiency, necessity and heuristic value of abstraction accounts of conceptual representation. This task is difficult on at least four grounds. First, as indicated above, abstraction accounts of representation fall into separate groups, at least two of which (prototype and strength notions) have little in common. Secondly, within each of these classes, there exists a broad variety of models with slightly differing assumptions, and the potential for the creation of new models with modified assumptions to meet the challenge of disconfirming data. Thirdly, and especially true of prototype models, the representation assumptions of many published models have not been specified in sufficient detail to permit rigourous assessment. Fourthly, many of the models have been identified with particular types of stimulus domains, which suggests that a very broad set of experimental preparations would be required to assess the spectrum of abstraction models adequately. In the face of these difficulties, two steps were taken to reduce the alternative set to manageable proportions.

The first step in this reduction was to note that despite differences in underlying logic, three major types of stimuli alluded to above may be dealt with under the same assumptions. The first type consists of stimuli that take values on continuous dimensions, exemplified by the height-weight domain used above. Such stimuli are generally summarized via a multi-dimensional mean, and have usually been employed to test distance models. The second type consists of channels of discrete values, usually summarized via a multi-dimensional mode. An example is birds, which possess the information channel of "beak", taking discrete, qualitatively different values "blunt", "curved", "sharp", etc. Continuous-valued stimuli may be treated in this fashion, by dividing the continuous scales into ranges, each range being considered a discrete entity. A third type consists of an undimensionalized pool of features. An example of this type is letter strings in which no positional or sequential constraint has been imposed on the construction of the strings. Such stimuli have been used most often in studies employing the feature-frequency or family resemb ce perspectives, under which features are not compared within channels, but simply in terms of the strength of the sum of features in items.

While each of these three types of stimulus is most often associated with and is perhaps most appropriate to a particular type of model, one must beware of too strongly identifying stimuli with models. Most models are capable of treating each type of stimulus, at least with some modification. Thus for example the unordered pool of features can be handled in terms of distance by asserting that the distance between two items varies along the single dimension of number of features overlapping (see Appendix 1). Other stimulus types can of course be thought of, such as dimensions of discrete but ordered stimuli, and a variety of unidimensional cases. However, these have received little attention, and in any case the arguments to be made here generalize easily to these cases. The experiments below employ the second type of stimulus, discrete feature values thought of as organized into separate information channels. Both prototype and feature-frequency models can be held to make strong predictions about the way such stimuli are processed.

The second step in reducing the size of the problem was the creation of a typology of assumptions of abstract representation. This typology organizes the variety of notions of what knowledge is abstracted for both prototype and strength models. A premise of this typology is that for abstraction models in general the feature, the unit of processing employed by the cognitive system, is at a level lower than the whole item. For example, in the stimulus domain ABX, AXC, XBC processing units would not be identified with whole stimuli, but perhaps with the individual letters which make up items. This assumption that the processing unit is lower than the item appears to be commonly, if implicitly, made by abstraction models of concept formation. Thus feature-frequency models assume that presented items are represented memorially via counts of item components and/or compounds of item components (e.g. Hayes-Roth and Hayes-Roth, 1977; Neumann, 1974, 1977). This assumption reflects the central value of the feature-frequency models, that items can be thought of as composed of recombinable elements, a perspective reminiscent of the associationist tradition. Prototype models also typically assume that the processing unit is at a level lower than the whole item (e.g. Posner and Keele, 1968, 1970; Franks and Bransford, 1971; Rosch, Simpson and Miller, 1976; Homa and Vosburgh, 1976; Rosch, 1977; Robbins et al, 1978; Homa, Sterling and Trepel, 1981; Omohundro, 1981). Although to the present author's knowledge this assumption has never been justified, it may be argued from the prototype perspective that the feature generally should be at a level below the item. Just as Rosch (1977) argued that "basic-level categories" exist at the most inclusive level of classification at which members have numerous attributes in common, because this is the most economical information cut, so it might be argued that the "basic level of features" is the greatest level at which the co-variation of their parts is maximized. In most domains this level will be found below the item level. Thus both strength and distance abstraction models typically assume that the level of the unit of processing is lower than the item level, and both have some reasons of internal consistency for this assumption.

The typology of abstract structure to be offered consists of three classes. Each class subsumes a variety of models and stimulus types. They classify structures in terms of variation of only two prominent assumptions. The first is "complexity", which is meant to refer to whether information regarding the dispersion of category members is embodied in the cognitive structure. Most simply, the cognitive structure might consist only of the most typical member of the category. Using the concept of bird as an example, the structure might consist of a picture of a typical bird, which might look midway between a robin and a sparrow. In this case it contains no information about the range of values taken by category members in general. An analogy is knowing the central tendency of a distribution while having no information about the variance or shape of the distribution. In terms of birds, no information would be stored by models of this class about the way beak shape varies across the population: only the most typical beak is represented. This appears to be the popular version of prototypes, mentioned above. More sophistocated versions might contain information about the frequency or variance of dimensional values. In the example, information about the typicality of buzzard- and hawk-type beaks would also be directly represented. Such versions entail an additional processing cost, but might increase the classifying power of the prototype through refining the system's expectations.

The second dimension used to organize prototype notions is "dependence", by which is meant the amount of information encoded in the cognitive structure regarding the co-occurrence of stimulus values. At the least sophistocated level, the cognitive structure encodes no information about how often aspects of members co-occur. For example, the frequency with which black plumage and sharp beaks are found to-gether would be unrecorded in the cognitive structure. More sophistocated versions might directly code such information. "Dependence" is used to describe these possibilities in the sense of "informative about". It refers to the amount of information the system can generate about one aspect of a stimulus given knowledge of a second aspect. It may be noted here that such contingency information can be effectively coded only if the system already codes a range of values of each aspect. If the system has only encoded "black" and "sharp", it can have no information regarding the correlation between colour and shape in general.

This fact reduces the complexity – by – dependence matrix to three cells. The first will be referred to as the "<u>simple independent</u>" class. Prototype models of this class assume that only the most typical aspects of the domain are encoded, and no information regarding co-occurence of aspects is represented. This class is exemplified by a list of the features of a typical bird. Such a list contains no information about the range of beaks available to birds, nor the frequency with which birds prefer green plumage given that they have blunt beaks. This appears to be the most popular view of prototypes. It is the only class of models in which the cognitive structure representing the category can be isomorphic with an actual member of the category. This class subsumes most of the models described above as prototype (distance) models, which share the idea that the prototype is an ideal category member. The only strength model belonging in this cell would be a model claiming that only high-frequency (i.e. most typical) single attributes are coded in the memorial representation. None of the strength models discussed above would be included. Even the first-order feature-frequency model encodes information relevant to the dispersion of items. However, it would include the family resemblance model in its simplest distance version, since that variant does not specify dispersion, but only the central tendency.

The second class of abstraction models is called "<u>complex independent</u>". This class of models assumes that not only typical but also atypical aspects of stimuli are directly represented in the cognitive structure, and perhaps additionally the frequency with which the various values occur. However, these models assume that the correlation between values is not represented. Such a structure might contain the information that although brown is the most common colour for a bird, they also come in green, black and red. The structure might also contain more precise information, for example that 60% are brown, 20% green, 15% black and 5% red. Such a structure would thus be sensitive to the novelty of a creature which had mostly very bird-like characteristics, but was blue. However, the structure contains no information on the frequency with which brown birds have blunt beaks, or red birds sport sharp beaks.

This class, to-gether with the simple independent class, exhausts the variety of traditional prototype models, which appear to accept as an unexamined assumption that the unit upon which people make similarity comparisons is the single aspect, unencumbered with considerations of the joint probability of two aspects. That is, prototype models implicitly assume that in comparing two birds, people separately assess the similarity of beaks, colour, and legs, rather than attending to higher-order units. This issue of the level of the unit of comparison is dealt with below in detail. For the moment it is sufficient to note that because of this unit assumption traditional notions of prototypes appear to fall into either the simple or complex independent classes.

First-order feature-frequency models also fall into the complex independent class. Although these models do not directly code either the central tendency or the particular items departing from the central tendency, they do code the dispersion of attributes, via differential frequency counts for attributes. They could thus be sensitive to the novelty of items containing novel attributes, although they are insensitive to the novelty of an item containing a novel combination of high-frequency attributes, for example the multi-dimensional mode, as discussed above.

The third class of abstraction models is called "complex dependent". As in the complex independent models, the cognitive structure is assumed to consist of information about not only typical aspects of stimuli, but also deviations, and perhaps even the frequency or variance of deviations. Additionally, the mental representation contains correlation or contingency information encoding the co-occurrence of stimulus aspects. Not only are the variety of beaks and colourations open to birds explicitly represented, but also the frequency with which red birds have blunt beaks. In dimensional terms, models of this class code not only the central tendency and variance, but also the covariance of stimulus aspects. Despite the statistical language used to decribe the information making up such a cognitive structure, however, knowledge at these levels need not have been achieved through procedures like multiple regression. It is possible that the system achieves such information and stores it through analog procedures.

The only traditional models falling in this class are the higher-order strength models, which count not only frequencies of single attributes but also frequencies of compounds. These models represent the dispersion of items through the differential frequency of orders of compounds and the differential frequency of compounds within an order. Thus they are sensitive to the novelty not only of items containing novel attributes, but also of items containing novel combinations of high-frequency attributes or compounds (although they may make an error: novel items near the multi-dimensional mode may have greater total strength than old items, as illustrated in part i of this section, above).

Summarizing this typology of abstract structure, the simple independent class consists of all those models in which the cognitive structure representing the concept consists solely of typical values, that is of a multidimensional mean or mode, or of a set of most frequent values. The complex independent class contains all those models that code not only the typical, but additionally a set of frequencies of atypical features in the discrete case (for each channel in the multidimensional featural case), or the range or variance of each dimension in the continuous case. As indicated above, a vital characteristic of these two classes of models is the assumption that the unit of subjective similarity or strength is the single feature or dimensional value, as defined by the experimenter. Higher units of similarity can only be accounted for in one of two ways: either the cognitive structure codes inter-dimensional correlation information, or the stimuli are being coded relatively holistically, in which case the experimenter has simply made a wrong guess about the level of the unit of subjective similarity. As will appear shortly, in the section below on cue validity, it has become a strong value of the prototype perspective that the features of stimuli are processed independently rather than interdependently.

The third class of abstraction models, the complex dependent class, consists of those models which code not only typical values and information about departures, but also contingency or correlation information summarizing the co-occurrence of features or dimensional values. This co-occurrence information might be at any level from bivariate (e.g. probability of a feature given that a second has occurred) up to the dimensionality of the space (e.g. probability of a feature given knowledge of all other elements of the stimulus). At higher levels of co-occurrence information subjects would give the impression of holistic encoding, inasmuch as the elements of the stimulus become bound to-gether through mutual prediction. That is, if no co-occurrence information is stored, it is quite possible for a subject in reproducing a stimulus to make errors on some elements quite independently of his probability of an error on another element; however, at high levels of coded co-occurrence information, error production for the components of a stimulus would tend toward all-or-none.

Medin and his associates (Medin and Schaffer, 1978; Medin and Smith, 1981; Medin and Schanenflugel, 1981) have proposed a dichotomy similar to the independence - dependence dimension used in this paper to typify model assumptions. They class as "independent cue models" all those which assume that the information used to make categorical judgements is an additive combination of the component dimensional elements. This class consists of all models (including traditional prototype notions) which imply that categories can be separated through a linear discriminant function of the component cues. They contrast this class with "relational coding models" which involve combinations of attributes as the functional processing unit. This class implies non-linear separability of categories. Medin's classification system differs from the present one in two regards. First, it does not include the "complexity" issue, the degree to which intra-dimensional variability is directly represented in the cognitive structure representing a category. Secondly, the "independent - relational" distinction is treated as dichotomous, specifying either independent (additive) or interactive (multiplicative) processing of attributes. By contrast, the "independent - dependent" distinction drawn in this paper is a continuous dimension, permitting attributes to be interdependent in processing at any of a wide range of levels. Conceptualization of higher-order feature-frequency models is little affected by which of these distinctions is drawn, since they effectively act like analysis-of-variance models, summing the separate frequencies of single attributes (main effects) and combinations (interactions). However, it will be seen below that the difference between Medin's independent cue - relational coding and the present processing independence - dependence distinctions is reflected in differences in central parameters of instance models proposed by Medin's group and by the present author.

The three classes of models identified in the present typology capture the central assumptions of digital distance and strength models. The strategy of this paper is to test these classes successively, assessing the force of the assumptions. The early experiments test the simplest class of models, while later experiments add or alter assumptions until an abstraction model can be made to account for all the data. However, the more assumptions which must be added to the simplest class, the further the resultant model is from the cardinal values of prototype theory, because the first and simplest class most clearly embodies the spirit of automatic, economical representation by purely summary information, while later classes sacrifice the summary nature, or the economy of representation, or both. The more complex the assumptions become, the more the values of the prototype perspective are violated.

This slippery slope does not apply to the higher-order strength models, which do not share the values of simple, summary representation. Those models will be evaluated instead in terms of the rigidity of their assumption of automatic abstraction of components and compounds of components. In pursuit of the simplest model which suffices, hybrid models, containing assumptions from both the abstraction and instance perspectives, are also assessed, and the implications of analog representations are investigated. Probability models are not directly tested, but the concerns of probability models are incorporated in the construction of stimulus domains, as indicated in the following section. In its conclusions, this paper will deny the sufficiency of the simple and complex independent classes and of hybrid classes of abstraction models to explain categorical performance; and will argue that the complex dependent class is unnecessary theoretically, cognitively uneconomical, and of questionable heuristic value.

Thus far little mention has been made of the formal properties of current instance models. This is not intended to suggest that the present research exists in isolation: in fact, a variety of instance notions have been proposed (Smith and Medin, 1981, provide a review), including such sophistocated models as Hintzman and Ludlam's (1980) "MINERVA" simulation and Medin and Schaffer's (1978) "context model". However, the strategy taken in this paper is to commence with the simple nearest-neighbour notion discussed in Section I ii, above, and allow successive experiments to inform it of necessary changes in its assumptions, in parallel with testing the assumptions of abstraction models. This strategy culminates in the proposal and testing of a new instance model, the "episode model", whose major concerns are then integrated or contrasted with those of prominent models in the literature.

III General Methodology: The Domain, Design and Dependent Measures

The stimuli for the experiments to be described were designed to incorporate Rosch's insights regarding the non-arbitrary structure of natural categories. Table 1 shows a set of 20 stimuli, similar to those employed in the experiments to follow. These strings have been drawn from two categories (labelled I and II) which exhibit non-arbitrary structure, consisting of relatively

Table 1

Example of Structured Domain

Category 1		Category 2	
FURIT	KURIT	NOKAL	NEKAL
FUKIG	FEKIG	NYBAL	NYBAP
FUREG	FUTEG	NOBAT	NOKAT
PURIG	PURYG	POBAL	POBYL
FYRIG	FYRIP	NOBEL	NOTEL
FURIG	FURIG	NOBAL	NOBAL

Neutral: T, E, K, Y, P

discontinuous clusters possessing graded internal membership. The first aspect of this structure, as illustrated in Table 1, is that each category has a central tendency. As can be seen in the first two columns, if the stimuli are treated as consisting of five dimensions, each identified with a position of the string, the modal values of Category I strings are "F", "U", "R", "I", and "G", such that "FURIG" is the multi-dimensional mode of the category, while (from columns 3 and 4) "NOBAL"-is the multidimensional mode of Category II. Such modal items or strings of modal letters are considered the category prototype patterns. They are average, most typical items, the "best" members of the categories. The central tendency of Category 1, "FURIG", has been plotted in Fig. 5, using the deviation-ring plot introduced in Figs. 3 and 4.

A second aspect of structure also evident in these stimuli is that the items are graded in their membership in or typicality for their category. The items in the first column of Category I all differ by exactly one letter-in-position (one dimensional value) from their central tendency; they are plotted in Fig. 5 on the first deviation ring. The items in the second column differ by two dimensional values, and are plotted in Fig. 5 on the second deviation ring. The prototype, as has been indicated, is the best member of the category, since it shares more features with all members of the category than any other member does. (In fact it shares four dimensional values with each 1st-ring item and three with each second-ring item.) The items on the first ring are not as typical, sharing fewer features with members in general than does the prototype (three dimensional values with each other first-ring item and an average of 2.6 with each second-ring item). The items on the second ring are still less typical (sharing an average of two dimensional values with each first-ring item and 1.5 with each other

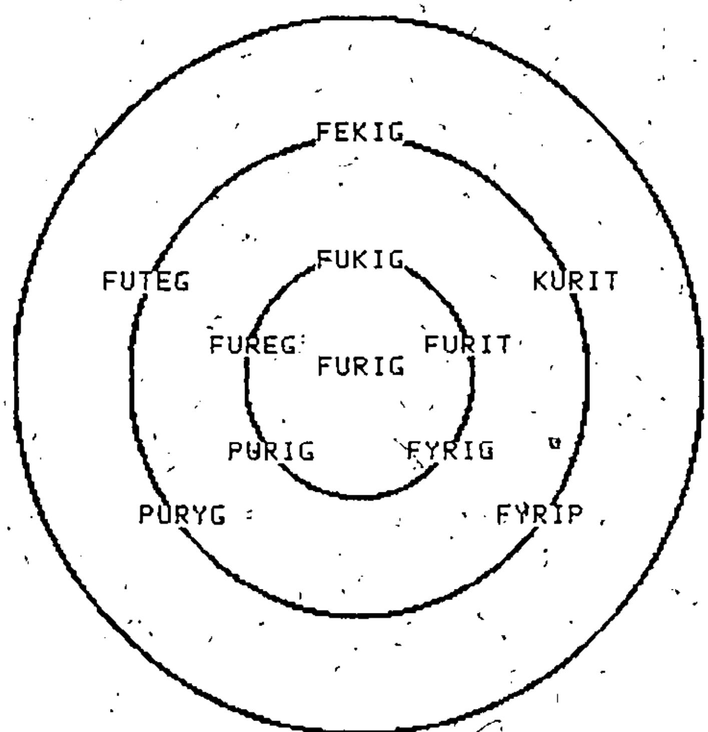


Fig. 5. Illustration of category structure. (Category 1 from Table 1).

second-ring item). The whole set of items thus exhibits graded membership, with the prototype the best member and items on peripheral rings being the worst members of the category.

A third aspect of structure evident in these materials is a relative discontinuity between categories. It may be recalled from the foregoing that without such a discontinuity there would in effect be only a single cluster of stimuli: the discontinuity splits the stimuli into two non-arbitrary categories. The discontinuity is evident in Table 1 from the fact that the letters "F", "U", "R", "I" and "G" occur only with items of Category I and the letters "N", "O", "B", "A" and "L" only with items of Category II, and the letters of both sets occur with equal and high frequency. The letters "T", "E", "K", "Y" and "P" occur with items of both categories equally, but with low frequency. The result is that any item of Category I shares very few features with any item of Category II and vice versa. This issue is taken up in more detail below, in a discussion of cue validity.

The clusters of stimuli comprising the two categories are thus well-structured. All sets of stimuli employed in the experiments to be described bore this structure. In any experiment, groups of stimuli were selected ten at a time, five from each category, all ten deviating from their prototype by the same number of features, and so selected that all sets of five stimuli from one category had the same mode. In terms of the ring structure illustrated in Fig. 5, this meant that a set of five stimuli selected from a category would all be on the same ring, and symmetrically located around the prototype. (The stimuli were actually located in a five-dimensional space, consisting of the five locations on the strings at which a letter could appear. The categories can be thought of as two contiguous hyperspheres, each consisting of a centre point (the prototype) surrounded by five nested shells. Each shell consists of many stimuli having in common their distance from the central tendency. The ring structure shown here is a two-dimensional representation of this 5-space; sets of five stimuli were actually located symmetrically around the prototype in five dimensions, not in the two illustrated. As a result, distances between stimuli which differ only in similarity to the protoype can be read directly from the plot, whereas distances between stimuli differing in similarity to another item cannot be directly represented.)

Performance on many sets of stimuli, selected to form interesting contrasts, is discussed in the studies described below. These sets can be conceived of as being drawn from a larger stimulus space clustered around two prototypes. One of the four counterbalanced stimulus spaces from which these sets were drawn is illustrated in Table 2. The relationships among the various types of stimuli in the domain from which experimental sets were drawn are illustrated for one category in Fig. 6a. Each stimulus plotted is to be understood as representing not only itself, but four other stimuli arrayed symmetrically around the same ring, and five more in the other category, similarly arrayed. For example, the item "FUKIG" represents all the stimuli in column Ia of Table 2; that is, a balanced set of ten items each deviating from its prototype ("FURIG" or "NOBAL") by one feature. The item "FUTIG" represents another set of ten items also deviating from the prototype by one feature, and additionally differing from the first set of items (the "FUKIG"-type items) by one feature. Items of the "FEKIG" type differ from the prototype by two, and so on through the space. Items like "GIKEF" differ from the prototype on all five dimensional values.

In Fig. 6b the items have been replaced by symbols indicating the

Table 2

Example of Domain

from which Test Items Were Drawn

Category I

Prototype: FURIG

Ia	Ib	IIa	IIb	IIc	III	V
FUKIG	FUTIG	FEKIG	FYKIG	PUTIG	PEKIG	GIKEF
FUREG	FURYG	FUTEG	FUTYG	FURYK	FYTEG	GETUF
PURIG	KURIG	PURYG	PUREG	FYREG	PURYK	GYRUP
FYRIG	FERIG	FYRIP	FERIP	FUKIP	FYTIP	PIRYF
FURIT	FURIP	KURIT	PURIT	TERIG	KURET	TIRUK

Neutral: P,K,T,Y,E

Category II

Prototype: NOBAL

Ia	Ib	IIa	IIb	IIc	III	V
NOKAL	NOTAL	NEKAL	NYKAL	POTAL	PEKAL	LAKEN
NOBEL	NOBYL	NOTEL	NOTYL	NOBYK	NYTEL	LETON
POBAL	KOBAL	POBYL	POBEL	NYBEL	POBYK	LYBOP
NYBAL	NEBAL	NYBAP	NEBAP	NOKAP	NYTAP	PABYN
NOBAT	NOBAP	KOBAT	POBAT	TEBAL	KOBET	TABOK

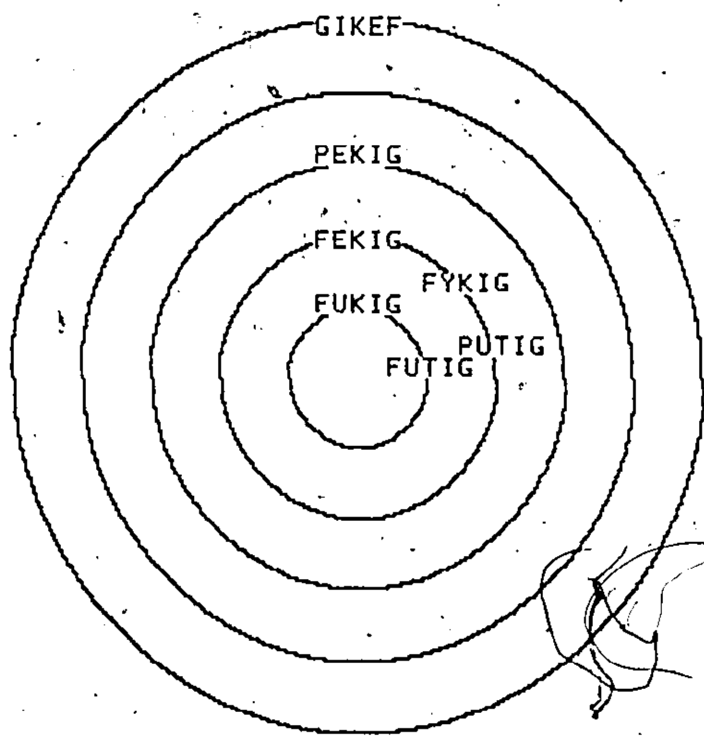


Fig. 6a. Relationships among stimulus types. Stimuli taken from Table 2.

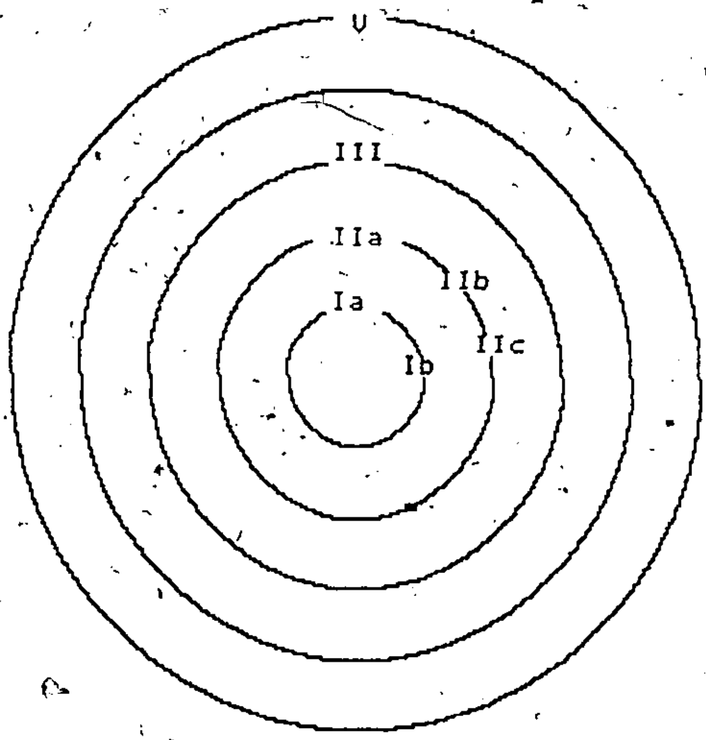


Fig. 6b. Symbolic representation of relationships among stimulus types.

TRAINING	TRANSFER
P	P
Ia Ib	$\boxed{\text{Ia}}$ Ib
$\boxed{\text{IIa}}$ IIb IIc	$\boxed{\text{IIa}}$ IIb IIc
III	$\boxed{\text{III}}$
IV	IV
V	V

Fig. 6c. Schematic represention of relationships among stimulus types.

relations of types of items to the prototype and to each other; these symbols head the columns of Table 2. The Roman numerals are to be read as indicating the number of features by which an item of that type differs from the prototype; for example, IIa items differ from the prototype by two letters in position. The subscript, where it occurs, is an aid to determining the smallest distance between two items on the same ring. Thus IIa items differ by one feature from the most similar IIb item and by two features from the most similar IIc item. These symbols are employed in discussion of the experiments below. In addition, a schematic of the space, like Fig. 6c, which contains a reduced set of the stimuli illustrated in Fig. 6b, is presented with each set of results, to remind the reader of the relative locations of sets of stimuli being contrasted in training and test. More information about the stimulus space is given below, following the introduction of the dependent variables and experimental design.

The most generally employed dependent measure in prototype- and strength-model research is classification. This is a fairly obvious choice, since one of the most salient aspects of concept formation is the ability to classify items by their category affiliation. However, classification is by no means the only psychological function of interest. Concept learning should have implications for performance on other tasks as well. Limited attention has been paid to other functions. For example, Rosch and her colleagues (Rosch, 1973, 1975, 1977; Mervis, Catlin and Rosch, 1977; Rosch, Simpson and Miller, 1976) emphasized that category learning has effects upon the subjective typicality of category members, while Hayes-Roth and Hayes-Roth's (1977) strength model was intended also to account for differential recognition of previously-experienced and novel category members. Classification and recognition judgements are important dependent variables in the arguments advanced experimentally below.

Another function which category learning might be expected to affect is the perception of individual items. The interest of this function for the study of concept formation is less obvious than that of the tasks listed above, since the overt task is concerned with items as individuals, not as members of categories. However, it is widely accepted that perception is influenced by the past experience of the system (e.g. Neisser, 1967), such that priorly-developed structure guides perception and consequent performance. For this reason, differential perception of stimuli can be used to draw conclusions about the nature of priorly-developed structure. For example, the classic Miller and Isard (1963) study presented semantic and grammatical, grammatical only, and ungrammatic sentences to subjects; their finding that meaningfull stimuli were easier to perceive than merely grammatical stimuli, and grammatical stimuli easier than ungrammatical, was interpreted to indicate the impact of a priorly-learned structure (in this case a general syntactic and semantic structure) on the perception of individual events. This strategy of deducing cognitive structure from differential perceptual fluency has been used in studies of a variety of areas, including information processing (e.g. Lindsay and Norman, 1977) and schema formation (e.g. Friedman, 1980). Of particular relevance to the present paper, which is concerned with the representation of categories in memory, are studies in the mainstream of memory research which have employed differential perceptual fluency to make inferences about the level of abstraction of memorial representation. For example, Jacoby and his associates (e.g. Jacoby and Dallas, 1981; Jacoby and Witherspoon, 1982; Jacoby, 1983b) have used this strategy to argue against the necessity of a semantic memory system (with which prototype and feature-frequency models are closely associated, as discussed above), and for the sufficiency of a memory system relying on unsummarized processing episodes (with which instance models may be identified).

Perceptual facilitation has also been employed in studies of word perception which share with concept formation a concern for determining the basis of generalization of performance. Thus, for example, Murrell and Morton (1974) used it to assess Morton's (1969) notion that words are memorially represented via abstract summaries (logogens), a strength notion which bears marked similarity to the feature-frequency models of concept representation discussed above. They measured the relative fluency of targets overlapping a previously presented word (e.g. "bored") either in shared letters (e.g. "born") or additionally in a shared root morpheme (e.g. "boring"), and concluded, on the basis of greater transfer to the latter target type, that abstracted morphemes form an important aspect of the memorial representation of words. In contrast, Feustel, Shiffrin and Salasoo (in press) elaborated on Murrell and Morton's experiment, including both non-words and overlap at the ends of items as well as at their beginnings, and concluded that encoded letter configurations are a better explanation than morphemes for the observed facilitation transfer. These papers raise the issue of the fundamental processing units of the memorial representation supporting generalization, and demonstrate the power of the perceptual fluency measure to make inferences about these units. The nature of these units is a major issue of this paper, as indicated above in the discussion of the complexity – dependence assumptions implicit in prototype and feature-frequency models. For these reasons, many of the studies to be discussed below employ perceptual identification rather than more traditional measures of concept formation.

Miller and Isard, in the study referred to above, employed the accuracy of shadowing sentences presented auditorily as their measure of perceptual ease. Their measure relied on the fact that the task was sufficiently difficult (because of speed of presentation or because of masking with noise) that resources differentially applicable to different types of sentences could evidence themselves in differential facilitation of perception. Had the sentences been presented unmasked and at a low rate, it is unlikely that errors would have been made in any condition. In the studies presented below, perceptual ease was measured via the accuracy of identification of letter strings presented visually. Task difficulty was increased beyond the trivial level by presenting the strings for a very short duration, followed by a pattern mask, in order that priorly-developed structure could be evidenced through differential success rates for different types of stimuli.

Because the methodology used to assess perceptual ability is a little complicated, and because a standard design was employed in the perception experiments below, the task and design will be explained in advance. Fig. 7 illustrates the identification task used to assess ease of perception. At the beginning of each trial, subjects were confronted by a left- and right-caret on a computer monitor (Fig. 7a). These carets were orienting stimuli, and remained constant on the screen throughout the test. Trials were subject-initiated. When a subject depressed a key, a five-letter string appeared between the carets (Fig. 7b) and remained for 30 milliseconds, being terminated by a mask (Fig. 7c). The subject was then required to produce on paper the string he thought he had seen on the screen. Subjects were required to produce five letters on each trial, guessing if necessary.

The design of the perceptual identification experiments was simple,

>   <

Fig. 7a. Fixation stimuli for perceptual identification.

> FURIG <

Fig. 7b. Stimulus presented for about 30 milliseconds.

Fig. 7c. Stimulus display terminated by pattern mask.

but somewhat awkward to describe. It will be described first from the point of view of the subject, following which the rationale of the manipulations will be discussed. First, subjects were passed through 30 perceptual identification trials of the type outlined above, these trials constituting a pre-test phase. In a second (training) phase, subjects were exposed to a set of 30 stimuli, and required to work with them in some fashion. Lastly, subjects were passed through a post-test phase, identical in all respects to the pre-test.

To explain the rationale of these manipulations, the middle (training) phase will be discussed first. In a typical experiment, five stimuli from each category were presented three times each in random order, for a total of thirty trials. Each set of five was selected, as described above, to average to the category prototype. All were equidistant from their prototype and symmetrically arrayed around it. In a typical experiment, the subject's training task consisted simply of copying the strings off the screen, being given all the time he wanted to do the job. This training task is the subject's opportunity to pick up the categorical structure of the domain, either by abstracting out the prototype, doing frequency counts of letter combinations or encoding the presented instances. The stimuli presented in this task are called Old, to reflect the fact that subjects have had a good opportunity to look at them. (The schematic diagram accompanying each set of results represents Old stimuli by means of a box around the relevant symbol in the Training portion (see Fig. 6c, above: IIa are Old). For referring to Old stimuli in text, the convention will be adopted of symbolizing them by means of a subscript "O", for example "IIa_O".)

The stimuli in the pre- and post-tests were of three types. Typically ten were the items exposed during training (the Old items). The other twenty, not exposed during training and consequently called Novel, consisted of two sets of ten items (five from each category) selected to make an interesting contrast. For example, the Old stimuli might be type II, and the Novel items of types I and III. (The schematic accompanying each set of results represents the sets of transfer stimuli via boxed symbols in the Transfer portion: see Fig. 6c, above.) The various models under examination make different predictions about the relative perceptibility of strings of these differing types after training.

The pre-test was a baseline measure of the subject's ability to perceive the stimuli without training. The thirty stimuli were presented once each, in random order. The actual measure of perceptual fluency used was the number of letters correct in position. As indicated above, the post-test was identical to the pre-test, and was scored in the same fashion. Pretest scores were then subtracted from post-test scores, giving gain scores for each item, which were then averaged for each type of item in the experiment (e.g. Old and two Novel types), yielding a gain score for each type. These average gain scores reflect changes in the learner's ability to perceive types of items over the course of the experiment. They represent the amount of extra perceptual fluency exhibited by the learner after he has had an opportunity to absorb the structure of the domain. While a variety of factors (e.g. practice effects) could cause a general gain for all types of items, differential gain of one type compared with another must be due not to a general factor, but to some factor differentially associated with the two types. Differences in gains in perceptual ease were therefore interpreted as due to the differential applicability of whatever structure the learner had picked up in the task for processing the various stimulus types. Employing the strategy described above, these differential gains were used to assess the ability of the various possible conceptual structures to account for performance. Gain scores were used in preference to the post-test scores simply to reduce the variance in response due to differential extra-experimental familiarity of the stimuli.

To eliminate gross ceiling effects and other contaminants, subjects whose average score on any type of item (any one of the sets of 10 parallel items) on the <u>pre-test</u> was four letters correct in position or greater (i.e. less than one letter-in-position from ceiling), or whose average <u>pre-test</u> scores on any two types of stimuli differed by one or more dimensional values (another ceiling effect), were excluded from further analysis. The number of subjects so excluded was small. No subject was eliminated from analysis on grounds of his post-test scores.

Now that the experimental design has been explained, characteristics of the stimulus space can be examined more closely. The actual letters occupying each feature position were varied considerably between experiments, in order to minimize any confounding effects of orthographic regularity. One constraint maintained across experiments was that the letters D, H, J, M, Q, S, V, W, X, and Z were never used, owing to their visual confusability with other letters, absolute rarity, extreme high or low frequency in particular positions of words and/or difficulty of pronounciation in arbitrary association with other letters. Another constraint was that the first, third and fifth positions of all stimulus strings were occupied by consonants, while the second and fourth positions were occupied by vowels ("Y" was always employed as a vowel). This C-V-C-V-C constraint made all strings pronounceable; the implications of this pronounceability are discussed below. Additionally, where possible, stimulus sets were counterbalanced between subjects within experiments. For example, in an experiment employing the stimulus types $\underline{Ia}_O$ – $\underline{Ib}$ – $\underline{V}$, four sets of stimuli were rotated through the three stimulus types, such that each set of stimuli was used for each type one time in four.

Table 3 illustrates one of the stimulus spaces actually used for the comparison of stimulus types $\underline{Ia}_O$ – $\underline{Ib}$ – $\underline{V}$. (These stimuli are drawn from Table 2.) It shows two categories, having prototypes "FURIG" and "NOBAL". All the items belonging to these two categories were created as deviations from these prototypes, by replacing letters of the prototype by one or more of the letters "P", "K", "T", "Y", and "E". Such replacement was symmetrical between categories, such that if the "F" of "FURIG" was replaced by "P", so too would the "N" of "NOBAL". As a result, these deviation-creating letters are equally represented in both categories, and hence non-discriminating. Thus the only features with above-chance diagnosticity, and hence the only features useful for classification, are those which make up the prototype, which should be an optimal situation for the emergence of prototypes as the representation of concepts.

Stimuli for all experiments discussed below were created such that for all sets of training items, not only are the experimenter-defined prototypes the average (modal) strings of the categories, but also the experimenter-defined categories have the highest cue validity of all possible ways of cutting the domain into categories. This is an important consideration for both prototype and feature-frequency models; as indicated above, Hayes-Roth and Hayes-Roth's property-set model (1977) assumes that categorization is determined by the feature compound having greatest diagnosticity, and Rosch (1977) contended that the categories from which learners derive their prototypes are those which maximize cue validity. Categories for which cue validity is maximized are those with optimal structure, those for which classification is most easily and accurately

Table 3

Example of Stimulus Sets Used in the Experiments

Category I

Prototype: FURIG

Ia_0	Ib	V
FUKIG	FUTIG	GIKEF
FUREG	FURYG	GETUF
PURIG	KURIG	GYRUF
FYRIG	FERIG	PIRYF
FURIT	FURIP	TIRUK
FURIG		

Neutral: P,K,T,Y,E

Category II

Prototype: NOBAL

Ia_0	Ib	V
NOKAL	NOTAL	LAKEN
NOBEL	NOBYL	LETON
POBAL	KOBAL	LYBOP
NYBAL	NEBAL	PABYN
NOBAT	NOBAP	TABOK
NOBAL		

accomplished. Thus the issue of cue validity can be thought of as at what level of inclusiveness the most informative cuts can be made in a stimulus domain, so that similarity is maximized within groups and minimized between groups; and as a result at what level of inclusiveness non-arbitrary categories will emerge. The outcome of this issue determines the number of functional categories that learners would be thought to evolve, the number and identity of their members and the identity of the functional prototype representing those categories.

While these considerations are most clearly directed at classification, they also have implications for performance on recognition and perceptual identification tests. Both of these tasks would be considered by prototype models to be accomplished through accessing the representations of categories encoded in memory as a result of the automatic process of abstracting commonalities. This process of abstraction is thought to occur even when the task is not overtly categorical; this is the abstraction explanation of the formation of concepts in the world at large, in which many of one's experiences of category members take place when one is not explicitly attempting to classify them. A simple independent prototype model applied to perceptual identification would seem to claim that the elements of an item would be perceptually facilitated to the extent that they match the separate dimensional values coded in the prototype. Similarly, a purely abstractive, simple independent prototype model would predict that recognition is a function of the overlap between a target and prototype. But the prototype is determined as the central tendency of a category whose membership is dictated by cue validity considerations, so that perceptual identification of an item is an indirect function of the diagnosticity characteristics of the whole domain. Thus cue validity is an important factor in designing a stimulus space in which to assess feature-frequency and prototype models of categorization, and prototype predictions of recognition and perceptual identification. However, it is less important for feature-frequency predictions of recognition (and probably of perceptual identification) since feature-frequency models assert that recognition (and probably perceptual identification) performance is mediated by the total strength of the target, not simply by those compounds with high diagnosticity.

For the stimulus domain illustrated in Table 3, cue validity might be maximized for the individual training items themselves as separate categories, or for the items grouped into the experimenter-defined categories shown in the table. If cue validity is maximized for items, then the domain consists of ten non-arbitrary categories, each consisting of the three repetitions of an item. In this case, the domain would be thought to be represented by ten prototypes, each identical with a training item. Alternatively, cue validity might be maximized for the experimenter-defined categories. In this case, there are only two non-arbitrary categories in the domain, and they are at a relatively high level of inclusiveness, each encompassing the three repetitions of five items. If these categories are most valid, then the functional prototypes should be the two experimenter-defined prototypes, for example "FURIG" and "NOBAL" in Table 3. It is necessary to determine which level of inclusiveness has maximal cue validity, in order to determine what are the functional categories and prototypes, so that appropriate predictions can be made under the prototype perspective.

Table 4 illustrates the computation of cue validity for the categories in Table 3, assuming as in Table 3 that the Ia items have been used as training items. The first section of Table 4 lists the five training items from each experimenter-defined category and demonstrates that "FURIG" and "NOBAL" are

Table 4

Example of Calculation of Mean Cue Validity

I Training Items

Prototype #1	Prototype #2
FUKIG	NOKAL
FUREG	NOBEL
PURIG	POBAL
FYRIG	NYBAL
FURIT	NOBAT
FURIG	NOBAL

II Validity of Old Items as Separate Categories

P(FUKIG/F)=1/4
P(FUKIG/U)=1/4
P(FUKIG/K)=1/2
P(FUKIG/I)=1/4
P(FUKIG/G)=1/4
1.5

Mean cue validity = 1.5 / 5 = .3

III Validity of Formal Prototypes as Categories

P(FURIG/F)=4/4
P(FURIG/U)=4/4
P(FURIG/K)=1/2
P(FURIG/I)=4/4
P(FURIG/G)=4/4
4.5

Mean cue validity = 4.5 / 5 = .9

indeed the modal strings for those categories. The second section demonstrates the computation of cue validity for a single old item, "FUKIG", that is, the validity of "FUKIG" as a category in its own right, consisting of its repetitions. Of the ten presented items, four have an "F", only one of which is "FUKIG". Thus the probability that a presented string belongs to the "FUKIG" category given the information that an "F" has been presented is 1/4. The same computations hold for the conditional probability of "FUKIG" given "U", "I" and "G". However, of the ten presented strings, only two possess a "K". Thus the conditional probability that a presented string belongs to the "FUKIG" category given that "K" occurred is 1/2. The mean cue validity for "FUKIG", that is, the mean probability that the item presented belongs to the "FUKIG" category given knowledge of any single feature, is the mean of these individual conditional probabilities, in this case .3 . All of the other nine presented items have the same mean cue validity. Cutting the space into ten categories each consisting of one item thus results in fairly poor diagnosticity for the categories: the accuracy of classification from this basis is little better than chance (which is .1 in this case).

By contrast, the third section of Table 4 presents the calculation of cue validity for the higher-level categories represented by the two prototypes in Table 3. This section examines the conditional probability of membership of the item "FUKIG" in its experimenter-defined category given each of its features separately. Of the ten training items, four possess an "F". All four of these belong to the category represented by the prototype "FURIG". Thus the conditional probability that a presented item belongs to the "FURIG" category given that it possesses an "F" is 4/4. The same probability obtains given that a "U", "I" or "G" has occurred. However, only two of the ten training items have a "K", and only one of these is a "FURIG"-type item: the conditional probability given a "K" is thus 1/2. The mean validity of the features of the string "FUKIG" for the "FURIG" category is thus .9 . The same is true of the other four FURIG-type stimuli, and of the "NOBAL"-type stimuli for their category. These two categories therefore represent a high-information cut of the domain, in fact the most informative possible cut: no other way of grouping items achieves as good a basis for classifying. Thus the experimenter-defined categories possess optimal non-arbitrary structure, and should be the functional categories. Under the prototype perspective this means that subjects should be sensitive to this partition of the domain, should abstract out the prototypes "NOBAL" and "FURIG" and should represent the domain memorially through these two strings.

This condition is true for all the sets of stimuli presented as training items in all experiments presented below. In each case the cue validity of the experimenter-defined categories is greater than that of the items as categories. Thus in each case the experimenter-defined prototypes should also be the functional prototypes, under the assumptions of the prototype perspective. The cue validities of experimenter-defined categories are given with each set of results, to provide the reader with a check on this assertion.

While this discussion disposes of the cue validity issue for first-order models, it is possible that some will wish to entertain the notion of a prototype structure which codes higher-order compounds of the stimuli: that is, a structure for which single features consist of compounds of letters. At some point the highest-information partition will switch from the experimenter-defined categories to categories consisting of the three repetitions of the items. Using the same pool of training items as an example, that is the items in the top section of Table 4, we may examine the diagnosticity of one-, two-, three-, four- and five-tuples of features for categories at the item and experimenter-defined levels. The validity of one-tuples has already been given above as .3 . The conditional probability of a presented string belonging to the item-category "FUKIG" given any two-tuple of the letters "F", "U", "I" and "G" is 1/3, since there are two strings besides "FUKIG" possessing each of these compounds; the conditional probability given any two-tuple which contains "K" as well as one of these four letters is 1/1, since such compounds occur once only in the set of ten items. There are six of the former type of compounds and four of the latter. This yields a mean diagnosticity of ((6 x 1/3) + (4 x 1)) / 10 = .6 . Similar computations yield mean cue validities for items as categories of .8 for the three-tuples and 1.0 for the four- and five-tuples. In contrast, for the experimenter-defined categories, the cue validities are .9, 1.0, 1.0, 1.0, and 1.0 for one- to five-tuples respectively. This means that it is not until the learner is processing items in units of four or five letters (i.e. the functional feature is a four- or five-letter chunk) that features become as diagnostic for item-categories as for experimenter-defined categories. In practical terms this means that the prototype perspective must continue to insist that the functional prototypes will be the experimenter-defined prototypes "FURIG" and "NOBAL" until it is willing to grant that the level of the functional feature is the four-tuple or higher. Moreover, the grand mean of cue validities for items as categories across all levels of compounds is .68, while for experimenter-defined categories it is .98 . If learners simultaneously attend all levels of feature compounds (as most feature frequency notions insist) then the experimenter-defined categories retain their superiority in diagnosticity, and hence would still be thought to be the func al categories; and thus the experimenter-defined prototypes would still be thought to be the functional prototypes representing the domain.

CHAPTER 2

The Mental Representation of Concepts

I Encoding Appropriateness and Perceptual Identification

As indicated above, the rationale of the experiments conducted in this paper is that different conceptual structures predict different patterns of performance, at least for some distributions of density of the domain. These differential patterns can be used to eliminate hypothetical conceptual structures making incompatible predictions. This experimental strategy requires a training condition which permits development of cognitive structures to a point where they can evidence themselves through differential performance. Thus in preparation for the studies contrasting abstraction and instance predictions, a series of experiments was run to determine what amount and type of exposure to training stimuli was sufficient to produce performance differences among types of items in the post-test. The experiments described in this section trace out the effects of varying training tasks, including letter-search, pronounciation, spelling and writing, on degree of differential transfer in perceptual identification. Each successive task was found to lead to more differential performance, leading to the writing task being adopted as the encoding condition for all later experiments.

All four experiments to be reported here employed the stimulus sets $\underline{Ia}_0$ - $\underline{Ib}$ - $\underline{V}$. As indicated above, subjects were first passed through a baseline perceptual identification task consisting of the 30 $\underline{Ia}$, $\underline{Ib}$ and $\underline{V}$ strings. They were then exposed, under one of a variety of encoding conditions, to the 10 Ia strings, repeated three times each in random sequence. Lastly they were passed through the post-test perceptual identification task, identical to the baseline task. As indicated above, the dependent variable was the difference between transfer and baseline perceptual accuracy for each stimulus type.

In the first experiment the training requirement consisted of a letter-search task which incidentally required subjects to scan each training string. This scanning activity was the only basis on which subjects could develop some sort of cognitive structure representing the domain. The gain scores for the three types of stimuli are shown in Fig. 8a: the mean gain scores are Ia_O: .34, Ib: .23 and V: .22. These gain scores are all very small, suggesting that training has affected performance on the identification task very little. None of the differences between these gain scores approaches significant levels. The stimuli in Experiment 1 consisted solely of consonants; it was the only study run for which this was true. In all subsequent studies the stimuli were C-V-C-V-Cs, as indicated above.

In Experiment 2, shown in Fig. 8b, the sole training requirement was that subjects pronounce the stimuli. In this case the pronouncing activity is the only basis upon which subjects could develop a cognitive structure. The mean gain scores are Ia_O: 1.09, Ib: .97 and V: .71. There is an evident increase in the magnitude of the gain scores from the first experiment, perhaps attributable to the change to pronounceable strings. However, although there appears to be a tendency toward differential gains (Ia strings gain more than Ib strings, and Ib gain more than V strings), none of the differences between gains is significant. In a third experiment, all conditions were retained, except that subjects were required to

TRAINING	TRANSFER
P	P
$\boxed{\text{Ia}}$ Ib	$\boxed{\text{Ia}}$ $\boxed{\text{Ib}}$
IIa IIb IIc	IIa IIb IIc
III	III
IV	IV
V	$\boxed{\text{V}}$

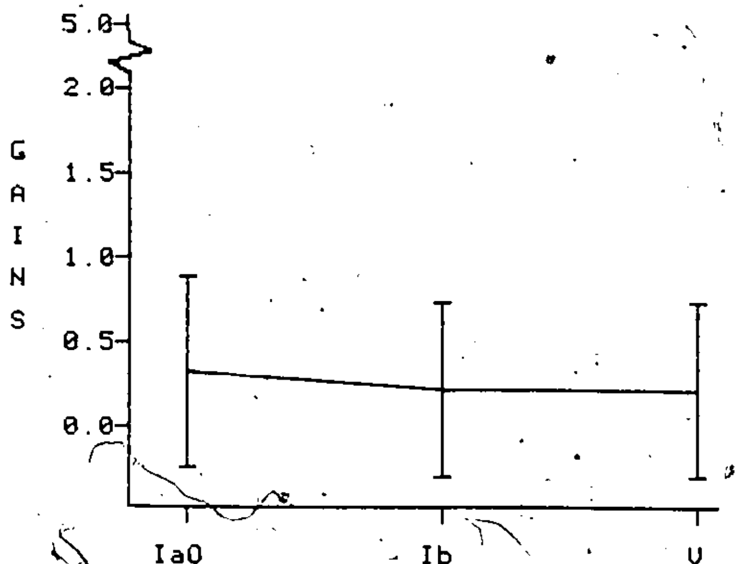


Fig. 8a. Perceptual identification gain scores from Experiment 1 (scanning). Gain scores are in mean letters-in-position per five-letter item.

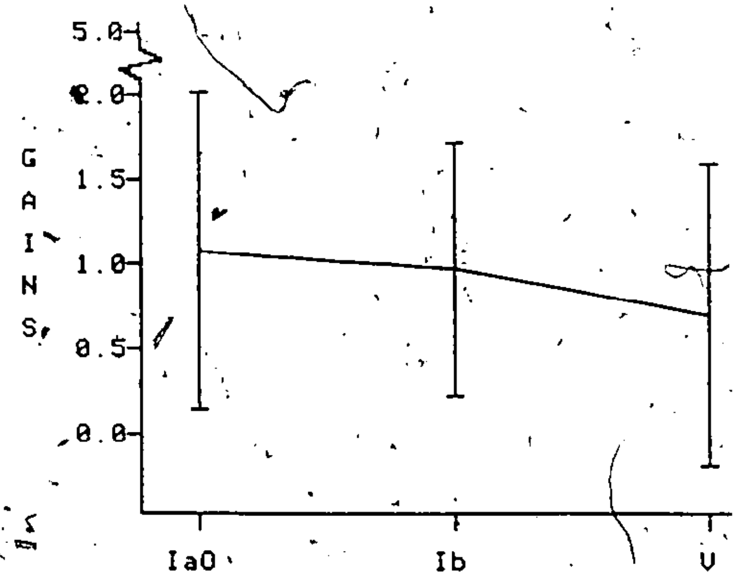


Fig. 8b. Perceptual identification gain scores from Experiment 2 (pronouncing).

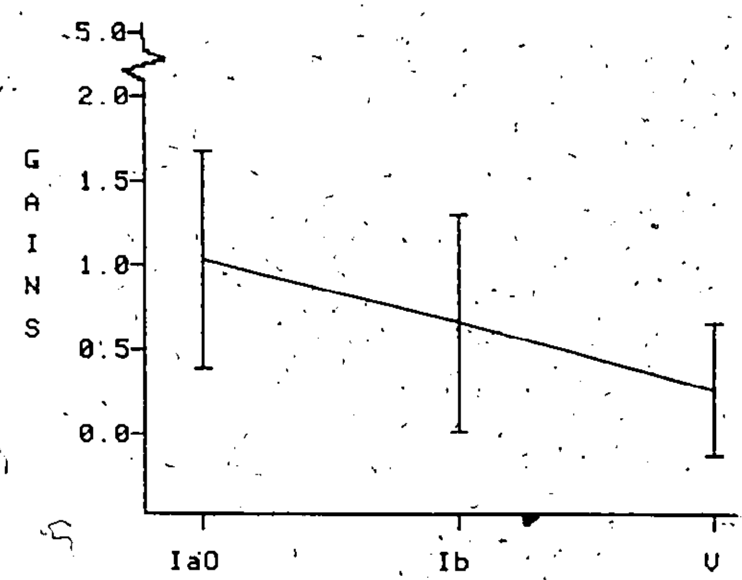


Fig. 8c. Perceptual identification gain scores from Experiment 3 (spelling).

spell the training strings orally. The mean gain scores, seen in Fig. 8c, are $\underline{Ia_O}$: 1.04, $\underline{Ib}$: .66 and $\underline{V}$: .27. These results evidence the same trend seen in Experiment 2, but more strongly; pairwise contrasts between gains are marginally significant ($p < .05$).

Finally in a fourth condition subjects were required to copy the strings during training; again, this activity is the sole basis for developing a cognitive structure representing the domain. The resultant gains (Fig. 8d) demonstrate an even more pronounced effect: mean gain scores are $\underline{Ia_O}$: 1.36, $\underline{Ib}$: .78 and $\underline{V}$: .17. All pairs of gain scores differ greatly ($p < .01$). This training condition thus seemed to satisfy the requirement for a type of exposure to training stimuli that would result in differences between stimulus types on transfer, and so was employed for the studies to follow testing predictions of the prototype and instance perspectives.

This series of studies strongly suggests that differential success on the transfer task depends on the manipulation of encoding during training. Processing at retrieval is not implicated since the retrieval task was not varied in this set of experiments; no argument is made that output task would not affect differential transfer if manipulated. The results suggest a gradient of "appropriateness" of encoding to transfer. This "appropriateness" might be thought of as encoding elaboration (Craik and Tulving, 1975) and/or distinctiveness (Jacoby and Craik, 1979) such that the effective property of the training in the latter experiments compared to that in earlier experiments is a richer or more differentiated encoding which is less confusable with everything else encoded, and hence of greater benefit to old items at transfer, and which is of greater benefit to old than novel items because it makes the encoding of old items not only

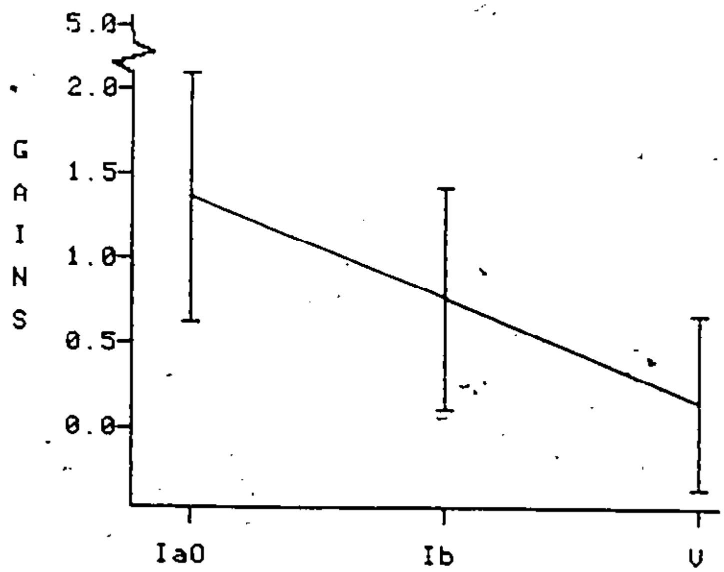


Fig. 8d. Perceptual identification gain scores from Experiment 4 (copying).

distinct from other old items, but from experiences in general. Alternatively, this "appropriateness" could conceivably be thought of as encoding specificity (Tulving and Thompson, 1973; Tulving, 1979), such that what matters is the degree of similarity between encoding and transfer conditions. The training requirement of writing the items puts the subject through a series of operations that are nearly identical to the operations he goes through for old items in the transfer task, but less similar to the operations required in the transfer task for novel items. For other training conditions there might be less similarity between training and transfer operations for old items, and hence perhaps less contrast between the similarity of old and novel item operations on transfer and encoding operations. The experiments offered in this paper do not provide a critical test of these explanations; however, the importance of task-induced differences in encoding for later conceptual performance will be discussed below as a major difficulty for abstraction models, which in general ignore what the subject is doing when he encounters a stimulus.

A second conclusion from these experiments was that the baseline task (pre-test) could safely be discarded as a source of differential gain. Looking across the four experiments, it appears that the slope of the transfer gradient is affected principally by the encoding condition employed in training. It is particularly evident in Experiment 1, which has a flat transfer gradient, that the baseline task has no significant differential effect on gains. As a result, all differential gains in transfer are assumed below to be due primarily to experience of training conditions.

II Simple and Complex Independent Abstraction Models

Subjects in Experiment 4 evidently learned something. The question of the nature of the representation of that learning now arises. In this section, a variety of experiments are reported that test the sufficiency of the first two classes of abstraction models to predict performance. The strategy of this and succeeding sections is to test successively more complicated abstraction models, indicating the inadequacy of each, and showing the additional assumptions required to repair the model. The intent of this process is not, of course, to show that abstraction models cannot in principle account for performance, but rather that in order to do so, they require such strong and complex additional assumptions that they lose the simplicity and economy which were to be their chief virtues.

One of the four counterbalanced stimulus spaces actually used for subjects in Experiment 4 was illustrated in Table 3, above. Simple independent abstraction models must predict that subjects exposed to this domain form two summary structures, each consisting of the modal values of whatever groups of items have highest cue validity. Subjects will develop these structures despite the fact that in this experiment they were not informed they were in a concept task, and were never required to attend to possible ways of classifying the stimuli. Such models must predict the formation of summary structures under these circumstances because of the prime value of the abstraction perspective that abstraction of structure is an automatic activity, accounting for the formation of concepts not only while people are explicitly trying to learn categories, but also when they are simply operating on the stimuli of the world with no explicit intention of learning to classify. (Again, this paper does not dispute that people can elect to abstract, only that abstraction is an automatic and ubiquitous activity.)

Table 4, above, illustrated that the experimenter-defined groups of training items maximize cue validity; the modal values of these groups, shown in Table 3, must therefore constitute the summary structures that subjects are thought to abstract through experience with the domain. A simple independent prototype model must predict that items consisting of exactly these values should be best perceived, since they consist of precisely that information by which the categories are thought to be represented. All other items are predicted to be perceived with accuracy proportional to their similarity to the nearest prototype, where similarity is defined by the number of dimensional values overlapping between the prototype pattern and probe. Taking all this into account, the simple independent prototype model can be held to predict that Ia and Ib items will be perceived with equal, high accuracy, despite the fact that Ia items are old and Ib novel, since they are all one letter-in-position different from their prototypes. Items of the V type will be poorly perceived, since they are all completely different from the nearest prototype. The pertinent results have already been shown in Fig. 8d, but are repeated here as Fig. 9a for the convenience of the reader. It is evident that the prediction for V items is correct, but the prediction of equal accuracy of identification of Ia and Ib items is disconfirmed: the old Ia items are better perceived than the novel Ib items ($p < .01$).

The simple independent strength model, consisting of the set of high-frequency features, predicts that accuracy of perception will be a function of the number of high-frequency features possessed by a probe. Since the set of high-frequency features consists of the modal features in this preparation, the predictions of this model are identical to those of the simple independent

TRAINING	TRANSFER
P	P
$\boxed{\text{Ia}}$ Ib	$\boxed{\text{Ia}}$ $\boxed{\text{Ib}}$
IIa IIb IIc	IIa IIb IIc
III	III
IV	IV
V	$\boxed{\text{V}}$

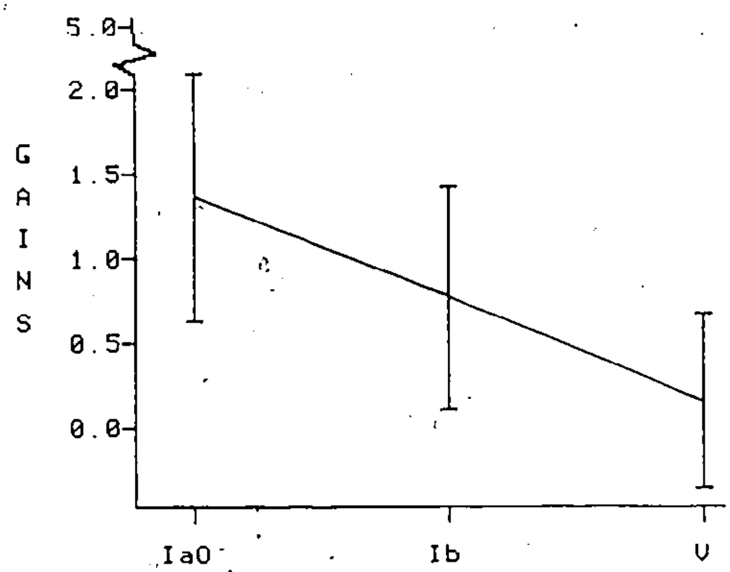


Fig. 9a. Perceptual identification gain scores from Experiment 4 (repeated from Fig. 8d).

prototype model, and hence the prediction of this model regarding the Ia and Ib items is also disconfirmed.

In contrast, a nearest-neighbour instance model claims that during the course of training, representations of the old items themselves were encoded in memory. It predicts that items will be accurately perceived to the extent that they are similar to the closest item in the training set. The Ia items are closest to items in the training set, since they comprised the training set, and hence are predicted to be most accurately perceived. Ib items are less similar to items in the training set than Ia items, but more so than V items; in consequence they will be less accurately perceived than the Ia items, but better than the V items. This is precisely the pattern of results illustrated in Fig. 9a. However, the formulation of the instance model provided here is very crude: the similarity metric is not specified, nor the precise nature of the representation of encoded old items, nor is the method by which current presentations recruit prior encodings. These issues will be dealt with later in this paper, as relevant evidence arises. In particular, evidence will emerge which will lead to dissatisfaction with the nearest-neighbour assumption of this model, leading to its being supplanted by a multiple-resources assumption.

Experiment 5 was nearly a replication of the foregoing experiment, except that it employed the stimulus sets $\underline{IIa}_O$ - $\underline{IIc}$ - $\underline{V}$. (One of the four counterbalanced sets of stimuli used is illustrated in Table 2, above.) The major rationale for running this experiment will become clear shortly, in the section below on letter-by-letter analyses. However, the experiment is also valuable as a check on the generality of the results of the previous experiment. In this manipulation the training stimuli differ from the prototype by two features, and are also in consequence farther apart from each other. This stimulus space has a lower density than the previous one, and the peak density has moved farther out in the space. The IIc items are also two features different from the prototypes, and two features different from the most similar old item. The V items differ from the prototypes and also from the nearest old item in nearly all of their features. The cue validities for the experimenter-defined categories averaged .8, and represented maximal cue validity for the domain. As in the foregoing experiment, therefore, under the assumptions of a simple independent prototype model the experimenter-defined prototypes should emerge during training to represent the categories.

The lower density of this space may be held to affect the predictions of the simple independent strength model, in that the high-frequency features in this study have lower frequency than they did in the foregoing study. The result should be an overall poorer performance on all types of items, although the predictions of relative success among the types of items remain as they were in Experiment 4. The lower density might also be thought to affect the simple independent prototype, at least in that there might be a higher probability of error in determining the modal values, since the modal values have lower frequency. The result again would be poorer overall performance; however, this is a very weak prediction, since the mechanism by which the prototype is created is ill-specified. More importantly, simple independent prototype models predict the same pattern of relative success as they did for Experiment 4, namely equal performance on IIa and IIc items, since they are equally similar to the prototype. The instance model also makes the predictions it did for Experiment 4, that old items (IIa in this case) will be best perceived, since they are closest to what it claims was encoded in training, and that V items will be worst perceived, since they are farthest from the training items. Unlike the strength notion, it does not predict poorer performance on IIa items in this experiment than on Ia items in the last experiment: the reduced frequency of letters does not affect the similarity of old items to themselves.

The results are shown in Fig. 9b. The mean gain scores are IIa_O: 1.49, IIc: .75 and V: .13 . All pairwise comparisons are significantly different ($p < .01$). The fact that IIa_O items are better perceived than IIc items is consistent with the instance model, but again disconfirms the predictions of the simple independent prototype and strength models. Moreover, comparing Figs. 9a and 9b, the decrease in overall perceptibility of the items that is to some extent predicted by the strength model is not evident.

These results are difficult for simple independent models to explain. However, they present no obstacle to the complex independent models, which code both modal and departure letters in their prototype version, or the frequency of all dimensional values in their strength version. These models actually predict that old strings will be better perceived than novel ones. This is because, in their dimensional versions, these models claim that in the case of an old string like "FUREG", in which the "E" is a departure from the prototype, it is encoded as a departure in the fourth position of the string. When "FUREG" is later presented as a probe, there is a representation in memory for each of its components, both modal letters (F,U,R, and G) and departure letters (E), so all letters are well recognized. However, when a novel item like "FERIG" is presented as a probe, the "E" is in the second position, a position in which it was never seen in training, and hence a position for which it was never coded. Hence when "FERIG" is presented

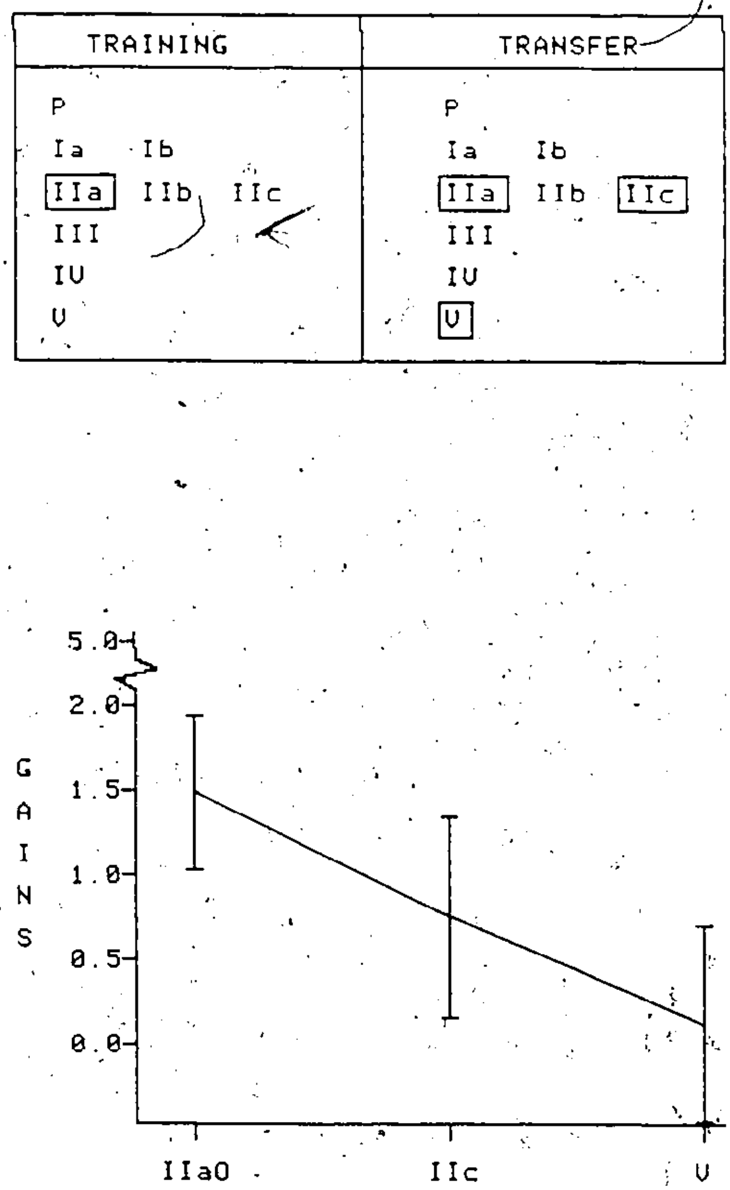


Fig. 9b. Perceptual identification gain scores from Experiment 5 (dispersed training items).

as a probe, the "E" is poorly dealt with, leading to poorer performance for novel than old items.

Two things must be noted about this argument. First, it applies only to stimuli treated as dimensional entities, and hence does not protect the subclass of dimensionless complex independent models (e.g. family resemblance models). This subclass fails along with the simple independent models, since, being dimensionless, it should not matter for these models where the "E" is located: it should be perceived with accuracy commensurate with its frequency, regardless of its position. This failure leaves only dimensional complex independent models as candidates.

Secondly, for this explanation to work even for the dimensional complex independent models, *all* of the difference between old and equally prototypical novel items must be due to poorer perception of the departure letters in novel items. Since all of the modal letters *have* been seen during training in the positions they occupy in both old and novel probes, the prototypes should assist the perception of these modal letters to equal degree in both types of items.

Table 5 presents observed letter-by-letter probabilities of accurate perception for Experiment 4, the first copying study. Column I presents data for *Ia*$_O$, *Ib* and *V* stimuli, giving probability of correct identification of departure letters and average probability of correct identification of modal letters for each stimulus type for the *pre-test*. Column II gives similar data for the *post-test*. Column III shows the gain in probability of identifying each. Thus the first entry in Column III can be read as indicating that due to training, the probability of correctly identifying a modal letter in an item increased by .27 . Since in this

Table 5

Experiment 4 - Letter-by-letter Data

	I Pre-test probability of errors	II Transfer probability of errors	III Reduction in probability of errors (I - II)	IV Gain in letters
$\underline{Ia}_0$ modal	.45	.18	.27	1.08
$\underline{Ia}_0$ departure	.54	.29	.26	.26
Ib modal	.43	.23	.20	.80
Ib departure	.50	.52	-.02	-.02
V modal	.38	.36	.02	.08
V departure	.47	.50	-.03	-.03

preparation there are four modal letters per string, this led to a (4 X .27 =) 1.08 letters-per-string increase in accuracy. The total gain from this source is noted in Column IV. Similar computations in the second line indicate that training led to an increased probability of correctly identifying the departure letter in old strings of .26; since there is only one such departure letter in each string, this amount represents the total gain in accuracy due to this source for the item. Adding the gains from both sources yields a total gain for old stimuli of 1.34 letters-per-string, which was the value plotted for $\underline{Ia}_O$ items in Fig. 9a. The remainder of the table presents the same type of information for Ib and V modal and departure letters. The data from Column III are plotted in Fig. 10, which presents the data from Fig. 9a analyzed into the contributions of departure letters and average modal letters.

Two attempts will be made to interpret these data. First, from the abstraction perspective, it may be seen that the probability of identifying a departure element in a $\underline{Ia}_O$ item increased by .26, whereas the probability of identifying a departure letter in a Ib (novel) item increased not at all as a result of training. The dimensional complex independent models predict just such a pattern, since, as argued above, departures in training items are coded in position, and hence will be better perceived in old items, in which they appear in their training position, than in novel items, in which they appear in a novel postition. However, this is the only evidence in favour of the abstraction predictions; the rest of the data is inconsistent with this perspective. As noted above, the gain for modal letters of $\underline{Ia}_O$ strings is 1.08 letters per item, while that for Ib modal letters is only .80 . Thus .28 letters-per-item of the difference between the perception of old and near-novel items is due not to poorer

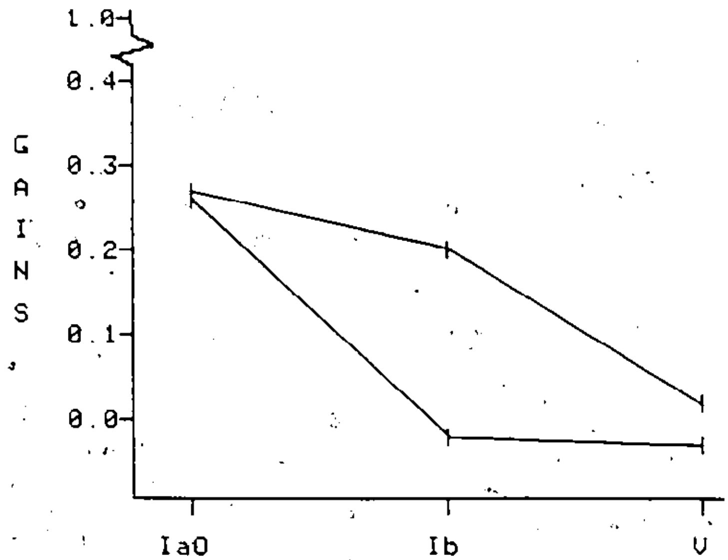


Fig. 10. Modal (above) and departure (below) gain scores from Experiment 4. Data taken from Column III of Table 5.

perception of the departure letters but to poorer perception of modal letters in novel items; and this .28 difference is approximately half the total difference between old and near-novel items (.28 / (1.34 - .80) = .52). This is a serious violation of the complex independent abstraction argument, which, as noted above, requires that <u>all</u> of the difference between old and equally prototypical novel items be due to poorer perception of the departures in novel items, since modal letters <u>have</u> been seen in training in the positions they occupy in both old and novel items, and hence the summary cognitive structure should assist perception of these letters to equal degree in both kinds of items.

A second interpretation of these data depends upon abandoning the assumption that the letters of the training strings are processed independently. This is a serious step for the prototype perspective, which, as discussed above, has an implicit value that items in the world are treated by learners as bundles of independently coded, recombinable elements. The abstractive orientation of the prototype perspective has led it to assume that subjects process items in units smaller than the whole item, units upon which various summary statistics can be calculated across items. The usual assumption has been that the experimenter-defined element (e.g. the letter in a string (Rosch and Mervis, 1975), the elementary transformation (Franks and Bransford, 1971) or the deviation of a dot in a pattern (Homa, 1981)) is the unit of processing, and that items are constructions of such elements (in the tradition of Bartlett (1932). The interpretation of the foregoing data to be made here assumes that the unit of processing is closer to the whole item, and consists of a bundle of the experimenter-defined elements bound into an integral processing whole. In the extreme case, under which the encoding unit is the whole item, the cue validity principle would indicate that prototypes (and incidentally categories) are as multiple as instances, and in fact are the instances. If encoding integrates the experimenter-defined elements to this degree, the prototype view collapses onto the instance view. It is a major argument of this paper that encoding at a level of integration higher than the level of the experimenter-defined element, closer to the level of the whole item, is common, and that to this extent the prototype view is not as general an explanation of performance as is the instance perspective.

The data of Table 5 is consistent with a relatively high level of integration of the elements, that is, relatively holistic encoding of the items. These data appear to show that perception of each element of a stimulus is dependent on the presence of other elements in the context of which that element was originally encoded. (This notion of dependency on context is formally identical to the notion of inter-dimensional dependency introduced above.) For example, the reader may recall that Ib departure letters are presented in a novel position along the string, and adjacent to letters to which they were not previously adjacent, while Ia_O departures are presented in their old position, still adjacent to their old neighbouring letters. Thus while departures of both types of strings are exposed during training, Ib departures are presented in the post-test in a novel context, while Ia_O departures are presented in a completely reinstated context. Under these conditions, the finding that Ib departures do not show any gain in perceptibility after training, while departures in Ia_O items do, suggests that departure letters gain only if they are presented in a familiar context.

More generally, *any* letter is perceived accurately only to the extent that it is in a familiar context. Ib departures are in a quite novel context, and gain nil; $\underline{Ia}_O$ departures, in an entirely familiar context gain .26 letters per string. $\underline{Ia}_O$ modal letters are also presented in an entirely familiar context and gain about the same amount, .27 letters per string on average; while Ib modal letters are in a context which is four-fifths reinstated (the novel departure in Ib strings accounts for the missing fifth) and gain .20 letters per string on average. From these data, the degree of reinstatement of the context of a letter from the training to the post-test appears to be highly correlated with the degree of gain in accuracy of perception of the letter. $\underline{Ia}_O$ modal and departure letters, both in completely familiar contexts, gain large and approximately equal amounts; Ib modal letters, with less context reinstated, gain less ($p < .05$); while Ib departures, in a completely novel context, gain much less ($p < .01$), in fact approximately zero. Modal and departure letters in type-V strings have almost no context reinstated, and also have gains near zero.

However, it cannot yet be concluded that the familiarity of the context of a letter, in the sense of the association of a letter with particular other letters of the string in a particular order, is responsible for the perceptibility of elements. An objection may be raised from the prototype perspective that not only was the context of Ib departure letters not maintained, but also they were presented in novel string positions; and that this latter factor of novel position alone accounts for poor perception of Ib departures. This objection attempts to rescue the independence assumption, by denying the import of the association between elements of presented items. It is a weak objection, since it fails to account for Ib modal letters being less well perceived than $\underline{Ia}_O$ modal letters; it appears that integrative encoding of the elements must be invoked to explain this difference. However, Experiment 5 (already partly explained above) permits investigation of whether this novel-position hypothesis has any merit, or whether the context-reinstatement argument is the better explanation of the data.

In Experiment 5, there were two departure letters per training item. Each particular letter that was used as a departure appeared in two strings of each category (see Table 2 - training stimuli for Experiment 5 are $\underline{IIa}_0$). These departure letters were of course presented in their training positions in old transfer items, and with completely reinstated context. In novel transfer items, the departure elements were also presented in their training positions. The only difference between the $\underline{IIa}_0$ and the equally prototypical IIc transfer items was that the inter-dimensional co-occurrence of departure letters was not maintained for the novel items. Thus if "P" and "Y" are departure elements in the training string "PURYG", they will occur in the novel items "PUTIG" and "FURYK" in precisely the same position, but in a rather novel context. In fact, no departure letter ever occurred in a novel IIc item in a position in which it had not been exposed during training. It is only the combination of elements which is novel in IIc transfer items (see Table 2, above). In consequence, if the positional objection is valid, and what matters is the familiarity of an element in a position independent of the familiarity of its context, IIc departures in this preparation should be perceived as well as $\underline{IIa}_0$ departures.

Table 6 illustrates the observed letter-by-letter probabilities of accurate perception for Experiment 5. The data from Column III of this table, the separate contributions of perception of departure and average modal letters to perception of the whole item, are plotted in Fig. 11. (These data were originally plotted in unseparated form in Fig. 9b.) Departure elements of IIc items evidently contribute much less to the accurate perception of the item than do departures in

Table 6

Experiment 5 - Letter-by-letter Data

	I Pre-test probability of errors	II Transfer probability of errors	III Reduction in probability of errors (I - II)	IV Gain in letters
$\underline{IIa_0}$ modal	.60	.25	.35	1.08
$\underline{IIa_0}$ departure	.55	.32	.23	.46
IIc modal	.57	.36	.21	.63
IIc departure	.55	.46	.09	.18
V modal	.53	.49	.04	.12
V departure	.56	.53	.03	.06

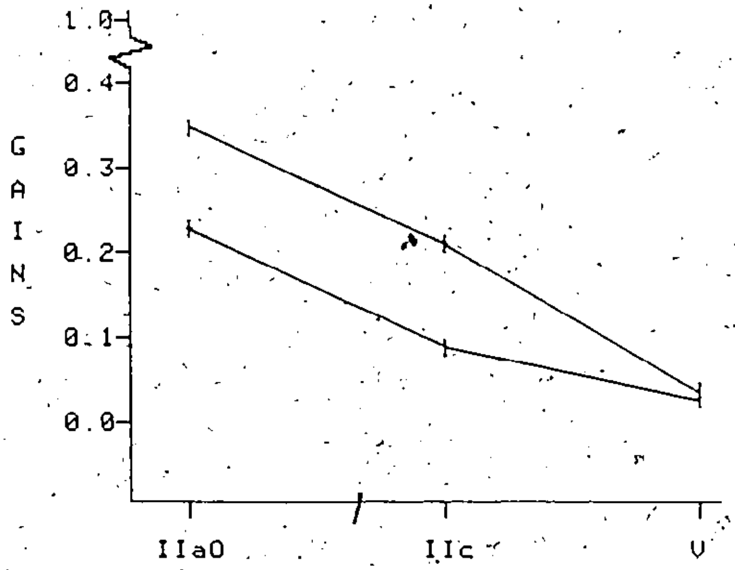


Fig. 11. Modal (above) and departure (below) gain scores from Experiment 5. Data taken from Column III of Table 6.

$\underline{IIa_O}$ items ($p < .01$); since these departure elements do not differ in reinstatement of training position, the positional objection is contradicted. The table and figure also illustrate that, as in the last experiment, a modal letter presented in a novel item is not as well perceived as when it is presented in an old item ($p < .01$), although the only difference between these two presentations of the modal letter is that the identity of <u>other</u> letters of the string has been altered. Both findings again demonstrate that accurate perception of any letter is dependent upon prior experience of that letter <u>in its present context</u>.

In terms of the typology of models introduced above, these findings illustrate inter-dimensional dependency, and eliminate all independent classes of models, of both the strength and prototype varieties, as potential modes of explanation. Moreover, the association of a departure element with a modal element is specific to particular items. Each departure element co-occurs with any modal element or combination of modal elements in precisely one stimulus string in the training phase of both Experiments 4 and 5. Thus the finding of dependency of the perception of modal letters on re-instantiation of their associated departure letters is indicative of the impact on perception of <u>particular</u> information about <u>specific</u> old items. Thus the failure of independent models destroys the notion that only summary information is retained, despite the abstraction perspective's emphasis on summary representation.

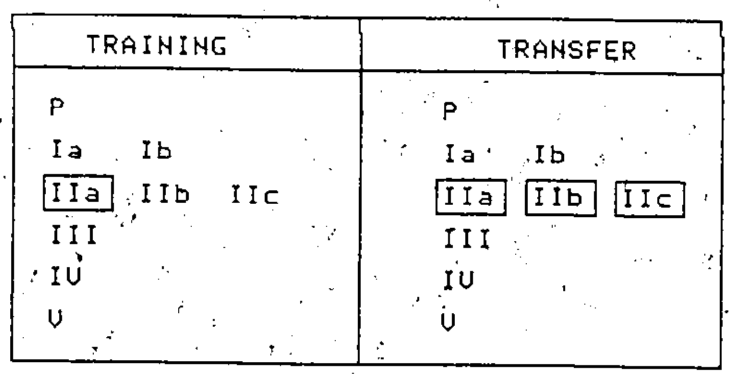
More importantly, perhaps, these findings can be interpreted as indicating relatively holistic coding of the items. The importance of this interpretation will be stressed below, in comparing the instance and abstraction perspectives in terms of heuristic value. For the moment it is sufficient to note only that the failure of independent classes of models forces the abstraction perspective into one of two modes of explanation: hybrid models, which combine abstraction and relatively literal encoding of items, or dependent models. The hybrid models will be dealt with first.

III Strong Hybrid Models and Analog Representation

In their 1968 paper, Posner and Keele rejected the idea that only the abstracted prototype is stored, stating that on the basis of their experimental evidence the presented exemplars themselves must also be encoded and affect later judgements. This claim has largely been ignored by the prototype tradition. The review by Rosch (1977), for example, never mentions it. But in view of the failure of purely abstractive models to account for the foregoing data, hybrid models consisting of a prototype plus special case information may be considered. In the strong form of such models, as suggested by Posner and Keele (1968) and by Homa, Sterling and Trepel (1981), transfer to re-presented items would be accomplished by means of generalization from stored representations of their earlier presentations, as in the instance formulation, whereas transfer to novel items would be accomplished primarily by means of generalization from the abstracted prototype. (This recalls the tradition within the reading literature that words (parallel to old items) are pronounced by retrieving a pronounciation from memory, while pseudowords (parallel to novel items) are pronounced by resort to abstract spelling-to-sound rules. Glushko (1979) attacked this notion, arguing that pronounciation of both orthographically regular words and pseudowords is accomplished by retrieving similar words.) A we.'' form hybrid model is also conceivable, in which generalization from an abstracted prototype plays a lesser but significant role in the recognition of novel items. This section assesses the strong hybrid model; the implausibility of the weak hybrid model will be demonstrated in the next section.

Experiment 6 employed the stimulus set $\underline{IIa_O}$ - $\underline{IIb}$ - $\underline{IIc}$. All of these stimuli differ from their prototypes by exactly two features. The cue validities of the training items for the experimenter-defined categories are the same as they were in the last experiment (.8), and are still maximal for the space. The strong form of the hybrid model predicts that $\underline{IIa_O}$ items will be well perceived, and better perceived than any novel items, by virtue of having been presented in training and in consequence coded and stored. The instance view agrees that $\underline{IIa_O}$ items will be best perceived: after all, this portion of the hybrid model's prediction is purely based on instance representation. The interest in this preparation centres on the predictions for relative perceptibility of the two classes of novel items, $\underline{IIb}$ and $\underline{IIc}$. A strong-form hybrid model must insist that these two types will be perceived about equally, since both are novel and so should be perceived with assistance from the prototype only, and since the two types are equally prototypical. By contrast, an instance model predicts that $\underline{IIb}$ items will be better perceived than $\underline{IIc}$ items, since $\underline{IIb}$ items are each closer to a stored item, one of the $\underline{IIa_O}$ items.

The results are shown in Fig. 12. The mean gain scores are $\underline{IIa_O}$: 1.07, $\underline{IIb}$: .80 and $\underline{IIc}$: .51 . The overall F is significant ($p < .01$); pairwise, $\underline{IIa_O}$ items gain more perceptual accuracy than $\underline{IIb}$ items ($p < .01$) and $\underline{IIb}$ items gain more than $\underline{IIc}$ items ($p < .05$). (When tested non-parametrically, this latter difference is significant at $p < .01$.) It is evident from these results that differential similarity of novel items to old items is an effective predictor of



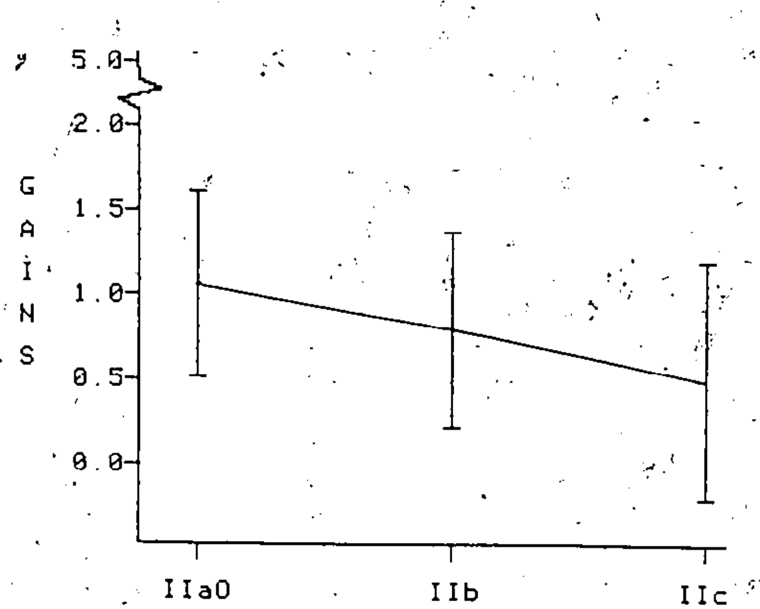


Fig. 12. Perceptual identification gain scores from Experiment 6 (all items equally prototypical).

differential perceptibility of novel items. This disconfirms the strong form of hybrid model. However, it does not deny that a prototype may play a weak role in the perception of novel items. To test this weak form of hybrid model, which suggests that both similarity to instances and also similarity to the prototype exercise a simultaneous effect upon perception of novel items, one must examine stimulus sets in which distance to old items is held constant and distance to the prototype is manipulated. Experiments relevant to this issue are presented in the next section.

The strong hybrid model, as discussed so far, is essentially a compromise version of a complex independent model, in terms of the typology introduced above. For this reason the data from Experiment 6 was analyzed letter-by-letter, as in the last section, to examine yet again evidence relevant to the assumption of independent coding of experimenter-defined stimulus elements. Table 7 presents the letter-by-letter probabilities of accurate perception for Experiment 6, and Fig. 13 plots gains per element for each stimulus type, taken from Column III of Table 7. (Fig. 13 may be compared with Fig. 11: stimulus types $\underline{IIa}_O$ and $\underline{IIc}$ were employed in both.) These data show the same trends evident in the letter-by-letter analyses of Experiments 4 and 5. It is again apparent that the probability of perception of a modal element declines with dissimilarity of the item as a whole from the nearest old, despite the modal element itself remaining completely unchanged, which is in contradiction of the complex independent predictions. Moreover, as in Experiment 5, departure letters were presented in familiar string positions in both novel and old items, so that the positional objection raised in the last section is irrelevant for these data as well. Instead, the context-dependent processing explanation appears to be the only

Table 7

Experiment 6 - Letter-by-letter Data

	I Pre-test probability of errors	II Transfer probability of errors	III Reduction in probability of errors (I - II)	IV Gain in letters
IIa_0 modal	.52	.29	.23	.69
IIa_0 departure	.56	.37	.19	.38
IIb modal	.53	.32	.21	.63
IIb departure	.58	.50	.08	.16
IIc modal	.49	.32	.17	.51
IIc departure	.55	.52	.03	.06

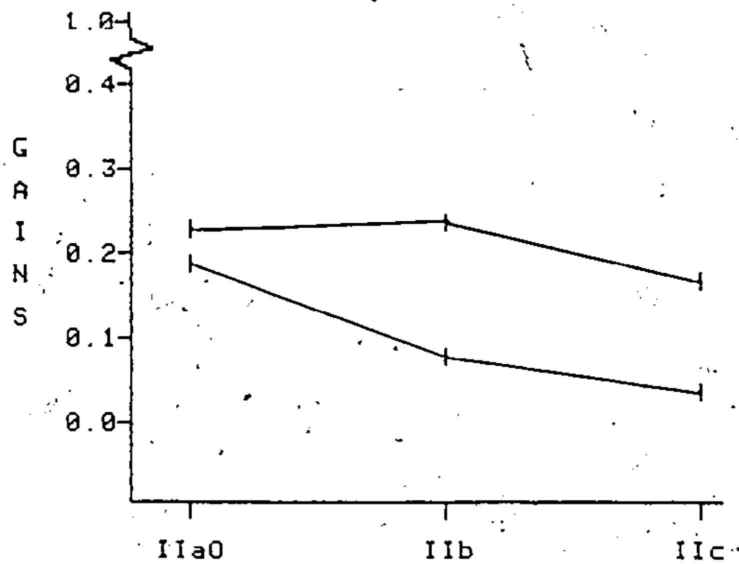


Fig. 13. Modal (above) and departure (below) gain scores from Experiment 6. Data taken from Column III of Table 7.

interpretation which fits the data. As in the earlier experiments, these data demonstrate a progressive decline in the probability of accurately identifying a letter as the amount of reinstated context declines. This is particularly evident in examining the accuracy of perception of departure elements: $\underline{IIa}_O$ departures are presented in the post-test in a completely familiar context (all five letters reinstated in order), IIb departures have three letters of prior context (e.g. the "Y" originally presented in the $\underline{IIa}_O$ item "PURYG" is presented in the IIb item "FUTYG", in which only "U", "Y" and "G" are reinstated), and IIc departure elements only two (e.g. the IIc item "FURYK", of which only the "U" and "Y" are reinstated for the "Y"). The elbow in the departure curve in Fig. 13 coincides with change in rate of decrease of context reinstated across the three conditions. The same trend is evidenced less dramatically in the plot of contributions of modal letters to the accuracy of perception of the whole item. Reinstatement of context is complete for $\underline{IIa}_O$ modal elements, four-fifths for IIb and three-fifths for IIc modal elements. This linear decline is reflected in the relatively smooth downward trend of the contribution of modal elements with decreasing similarity to the nearest old item. These data thus provide additional evidence that letters-in-position are not coded independently of each other, again demonstrating the inadequacy of the independent classes of models to account for performance. The following section investigates the implications of dependent classes of models, linking inter-element dependency to integrated encoding of items and to a multiple-resource model of concept representation.

There is an additional trend evident across the three sets of letter-by-letter data presented above, suggesting that modal elements of old strings are perceived better than departure elements of old strings. The effect is not large, but is present in the stimulus set $\underline{\text{IIa}}_O$ - $\underline{\text{IIc}}$ - $\underline{\text{V}}$ and to a lesser extent the set $\underline{\text{IIa}}_O$ - $\underline{\text{IIb}}$ - $\underline{\text{IIc}}$ (Figs. 11 and 13). This effect is not predicted by a nearest-neighbour instance model. It will be recalled that such a model does not distinguish between modal and departure elements, since these are only defined for cognitive systems which summarize across the set of items. Moreover, the context of a departure letter in an old item is as well reinstated as the context of a modal letter. This effect is the first difficulty encountered with nearest-neighbour instance models, but more will be encountered in later experiments. The effect is accounted for by multiple-resource models, and will be discussed below where these models are presented.

Experiment 6 also provides evidence regarding the utility of analog prototype representation. Analog representation appears to imply that the information maintained by the cognitive system about a category is relatively holistic, that is, that it retains some information about the interdependencies of stimulus elements. Such representations thus escape the disconfirmation which overtook independent models in the last section. A composite image such as that of the typical dog (as described above under Traditional Abstraction Models) maintains information about the typical value of each dimension and the variability of each, and may also be thought to code typical values of different dimensions interdependently. However, it cannot retain information about the co-occurrence of atypical values on different dimensions, nor of an atypical value on one dimension with a typical value on another. A mushy composite photograph of 100 dogs would not indicate to what degree abnormally short legs were associated with normally or abnormally short tails. Thus such a representation could yield dependencies between modal features, resulting in poorer performance on one modal element of a stimulus if a second modal element is not reinstated; but should not yield dependencies between departure features and other features. Yet the results of Experiment 6, described above, indicate that the probability of perceiving a typical (modal) feature is a function of the reinstatement of the particular departure elements with which it was associated in a particular old item. Thus the analog prototype model is also disconfirmed.

IV The Metric Issue

i) Initial Evidence on Performance Gradients

Thus far the strategy of this paper has been to push the prototype perspective into making assumptions that are further and further away from its initial position, through examining selected contrasts that bear on its assumptions. This section begins a similar process for the instance perspective, attempting to find informative contrasts that will force a more precise formulation of instance notions than has been given heretofore. It demonstrates that local domains of influence around particular old items are insufficient by themselves to account for transfer patterns; it suggests that when the domains of influence of two old items overlap, they reinforce each other; and it raises the issue of how close such domains must be before they reinforce (i.e. the extent and slope of the generalization gradients around each old item).

Experiment 7 employed the stimulus set Ia - IIa_0 - III, with cue validities of .8 for the experimenter-defined categories. The resultant performance gradient is illustrated in Fig. 14. The abscissa plots the three types

TRAINING			TRANSFER		
P			P		
Ia	Ib		Ia	Ib	
IIa	IIb	IIc	IIa	IIb	IIc
III			III		
IV			IV		
V			V		

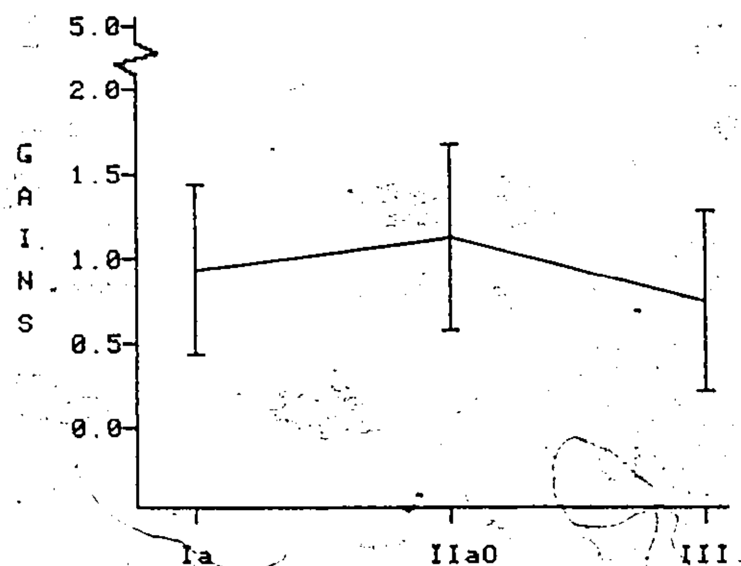


Fig. 14. Perceptual identification gain scores for Experiment 7 (items decreasing in prototypicality).

of items from left to right in decreasing order of typicality. The mean gain scores are Ia: .95, $\underline{IIa_0}$: 1.13, III: .75. The overall F is significant ($p < .01$); pairwise, the training items $\underline{IIa_0}$ are better perceived in transfer than the Ia items ($p < .05$) and the Ia items better than the III items ($p < .05$). These results contradict one variant of instance models, the deterministic nearest-neighbour model. This model, which is the one usually tested by prototype researchers (e.g. Reed, 1972; Hayes-Roth and Hayes-Roth, 1977; Homa, Sterling and Trepel, 1981), and which was descibed in the Introduction of this paper as a simplistic version, states that whichever stored instance is in objective fact the most similar to a probe item will be employed as the sole resource for the perception of the probe. If this were the case, Ia probes in this preparation would have been equally well perceived as III items, since each type deviates from a nearest old item by exactly one element. However, this deterministic nearest-neighbour model has never been seriously proposed by instance theorists; instead, for example, Brooks (1978) proposed a probabilistic nearest-neighbour or concert-of-instances notion, and Medin and Schaffer (1978) proposed multiple-exemplar access. These notions are described in Section V iib under the heading of multiple-access models. Section V iia conducts an analysis of models which access a single memorial resource for facilitation of a probe.

Experiment 6, described in the last section, traced out a performance gradient for items of equal prototypicality (two deviations from the prototype) but of differing similarity to the nearest training item (zero to two deviations from the nearest). Experiment 7 (just described) provided this gradient for stimuli differing simultaneously in similarity to the prototype and to the nearest old item, deviating respectively one, two and three elements from the prototype and one, zero and one deviations from the nearest old item. Both experiments employed the IIa items as training items and consequently as old items in transfer. Taken to-gether, the results of Experiments 6 and 7 trace out local transfer surfaces around the IIa_0 items in two dimensions: around the prototype and out from the prototype. These surfaces peak at old items and taper off in both dimensions. They taper off symmetrically from old items in the around-the-prototype dimension (transfer to items equally prototypical but less similar to the old). However, they taper off more steeply toward the periphery of the space than toward the centre. As indicated above, this assymetry indicates that these transfer surfaces are not properties only of the old item they surround, since in that case they would be symmetrical in all dimensions.

The data from Experiments 6 and 7 trace out a plot of probability density of perception of items. The problem is to determine what processes underlie this probability density plot. It has already been shown that it is most unlikely that each local transfer surface (the surface consisting of the probability density of the identification of transfer items surrounding each old item) is a function only of the old item embedded in it; and also that the prototype representations so far considered could not account for such surfaces. An attempt will now be made to account for these probability density surfaces in terms of the average or simultaneous effects of a set of hypothetical underlying generalization gradients which do belong to old items individually. These underlying gradients are assumed to be symmetric in all dimensions (in 5-space). Supposing that where these underlying gradients overlap, they reinforce each other, they would lead to distortions of the transfer surface away from the symmetry of the underlying local distribution. Fig. 15 illustrates this notion, employing

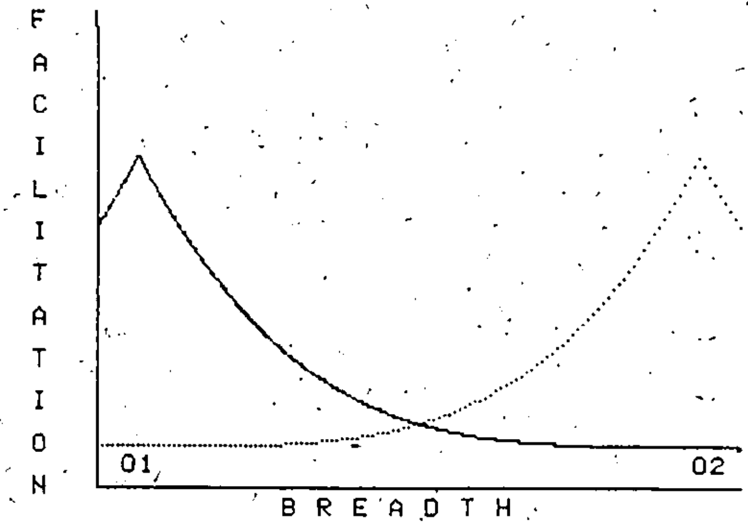


Fig. 15a. Hypothetical underlying generalization gradients of two items (01 and 02).

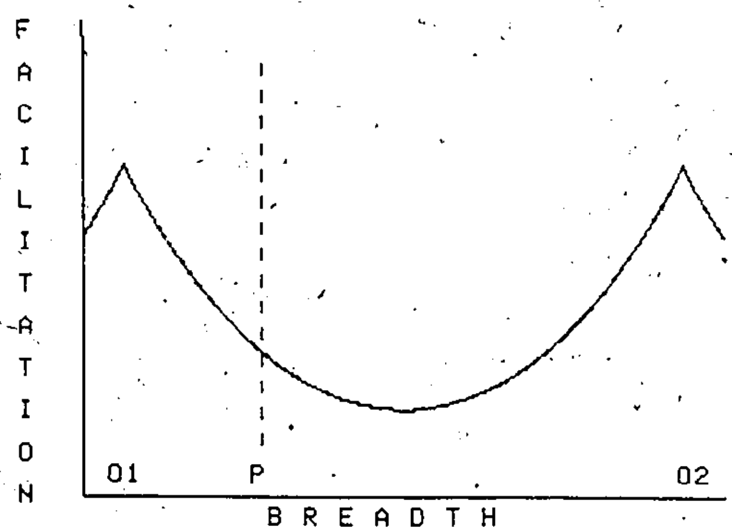


Fig. 15b. Resultant transfer surface. Probe (P) receives much the greater part of its facilitation from the nearer item (01).

generalization gradients around two old items. Fig. 15a shows the underlying distributions of two items that are very dissimilar; Fig. 15b demonstrates the resultant performance gradient, assuming that performance is determined by the algebraic sum of the generalization gradients (an assumption of long standing in the generalization literature; see, for example, Spence (1937)). Because the items were too far apart for their generalization gradients to reinforce significantly, the performance gradient has the same appearance as the underlying gradients. However, if the old items are less dissimilar, as in Fig. 15c, a probe item may gain benefit from both, either through simultaneous access of the two items, or through having two neighbours nearby, either of which can assist. The resultant performance gradient is seen in Fig. 15d: again it is an additive composite of the two underlying gradients. The probe item P is the same distance from the old item O1 in Fig. 15d as it was from O1 in Fig. 15b. However, its probability of correct identification has increased due to the nearness of the second old item.

Supposing something like this to be underlying the results of Experiment 7, it is clear why Ia items are better perceived than III items. While both types of probes are equally similar to the nearest old item, Ia items possess the advantage of being closer to the remaining old items, which also may assist in their processing, and gain probability of correct identification from that advantage. In fact, Ia items share an average of 2.6 features with old items in general, while III items share an average of only 1.8.

It becomes interesting to consider the advantage possessed by old items themselves, which results in their being the best-perceived of all types of transfer items in this preparation. They are of course identical to one old item (themselves) but are not as close to the remaining old items as are the Ia items.

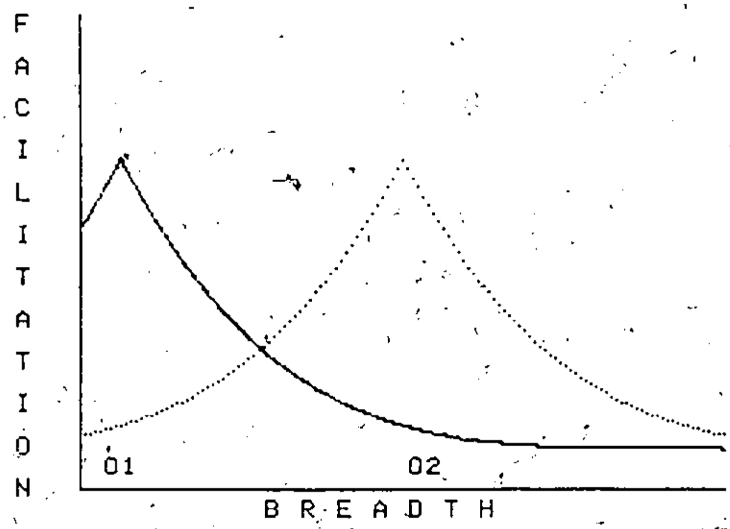


Fig. 15c. Underlying generalization gradients from Fig. 13a in closer proximity.

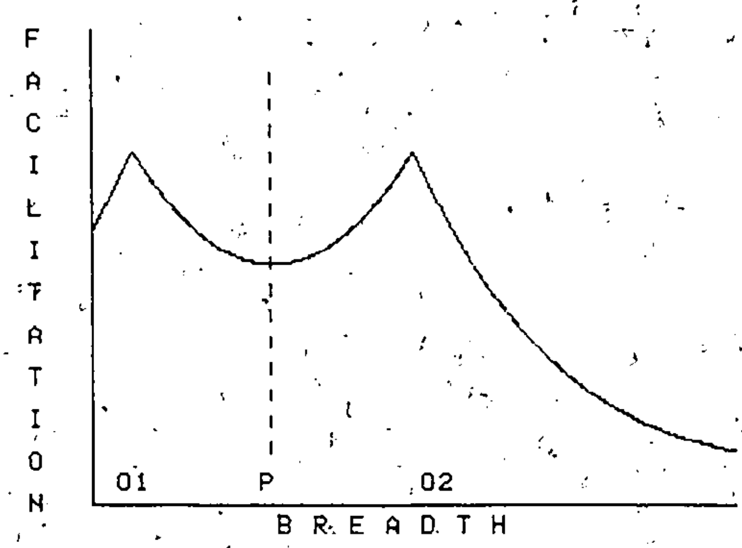


Fig. 15d. Resultant transfer surface. Probe enjoys greater facilitation although located at the same distance from item 01 as in Fig. 13b.

In fact, while Ia items share an average of 2.6 dimensional values with old items in general, IIa_O items share less, an average of 2.2 values with old items in general. If the supposition above of the reinforcing effect of multiple underlying generalization gradients is correct, should not the Ia items be better perceived than the IIa_O items? The answer to this puzzle may lie in the shape of the underlying generalization gradients. If these gradients are broad and flat, then one would expect considerable facilitation of probes falling between the old items, as in Fig. 13d. However, as shown in Figs. 15e and 15f, if these underlying gradients are relatively steep and narrow, a probe at the same distance from the two old items will experience relatively little facilitation of perception. The next section provides evidence that the underlying gradients in these experiments are indeed relatively steep and narrow.

ii) Recognition Data: Evidence on the Breadth of Generalization

a) Single-Resource Models

Experiment 8 was run for multiple purposes. It employed the dependent measure of recognition in place of perceptual identification, the measure used in all experiments thus far. Recognition is used to describe the judgement made by subjects in the sense of indicating that one feels one has encountered a specific item at some prior time. In employing recognition as a dependent measure this experiment broadens the generalizability of the conclusions of the study to a second function, one which has been considered important by abstraction theorists. It also provides evidence of a convergence between the perceptual and memorial

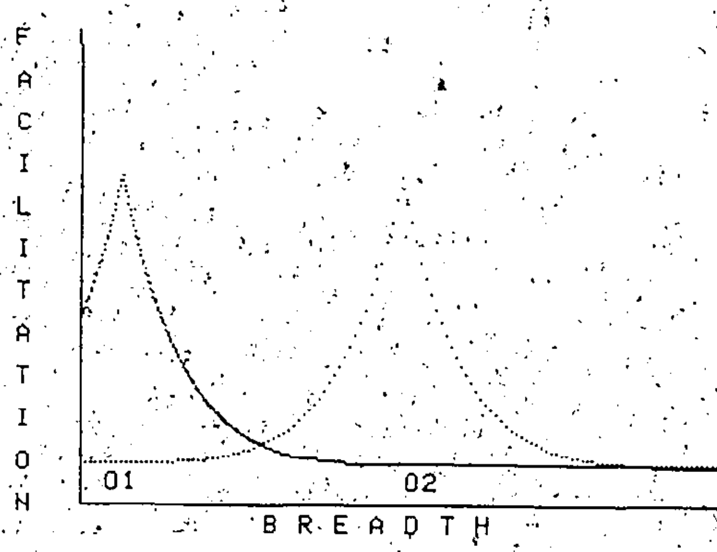


Fig. 15e. Items are at same distance as in Fig. 13c but have steeper generalization gradients.

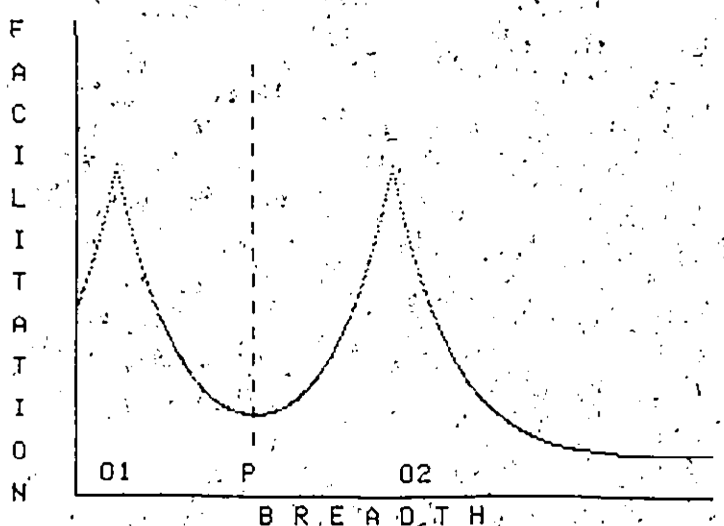


Fig. 15f. Resultant transfer surface. In this case the probe benefits little from its proximity to a second item.

measures which is predictable from contemporary memory research. Further, it provides more compelling evidence on the slope of underlying generalization gradients than was obtained in earlier experiments, and also evidence relevant to the weak hybrid model alluded to above.

The training phase was identical to that of the foregoing studies: 10 $\underline{IIa_O}$ items were exposed to be copied, three times each in random order. (The experimenter-defined categories still have cue validities of .8.) As in previous studies, subjects were not warned in advance of the transfer task that they would be required to perform any further task on the trained items. Unlike the perceptual identification experiments, no performance baseline was taken before training.

In the transfer phase, subjects were required to make recognition judgements about 10 stimuli of each of the Ia, IIb, IIc, and III types, as well as of the 10 $\underline{IIa_O}$ stimuli. Subjects were required to make a compound recognition-confidence judgement on a six-point scale, consisting of high-, medium- or low-confidence old, or high-, medium- or low-confidence new, where old-new refers to whether or not one has seen that specific item previously. This task was selected on the intuitive grounds that accuracy ratings alone might be insufficient to separate performance on the various stimulus types. It differs in unknown ways from other tasks which might have been selected, such as successive ratings of recognition and confidence.

Fig. 16a illustrates the results, but is difficult to interpret. The same data are re-plotted in Fig. 16b, in the form of a cumulative frequency plot, which (like an ROC plot) facilitates visual inspection of the patterns. The initial question to be asked of these plots is whether the curves for the various

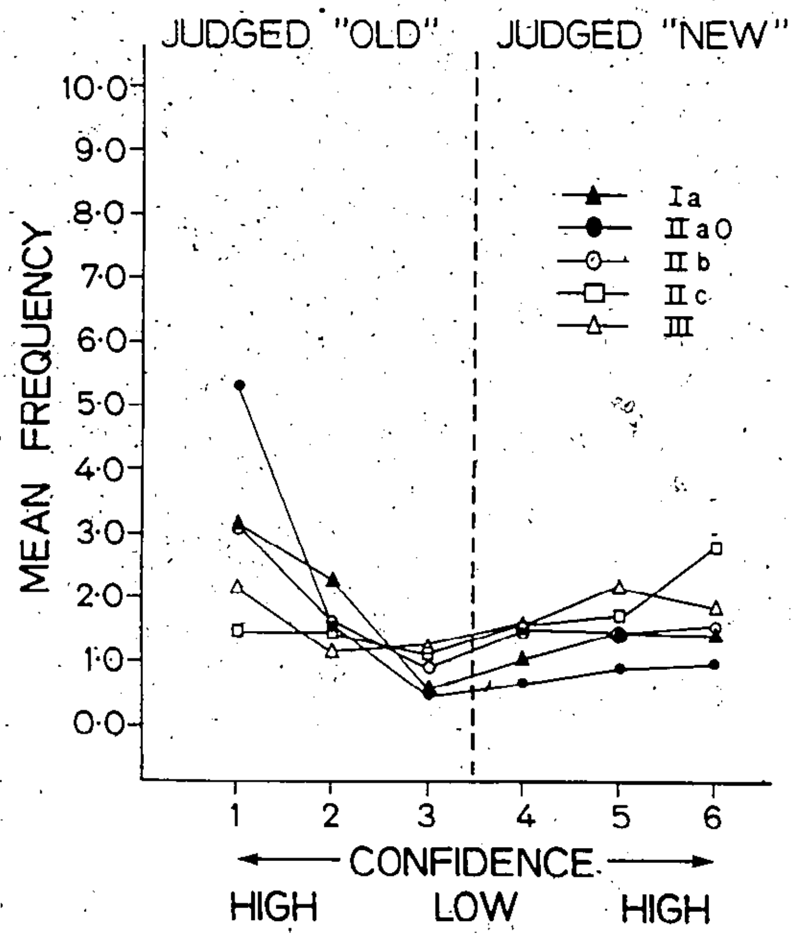


Fig. 16a. Mean frequencies of recognition-confidence judgements from Experiment 8.

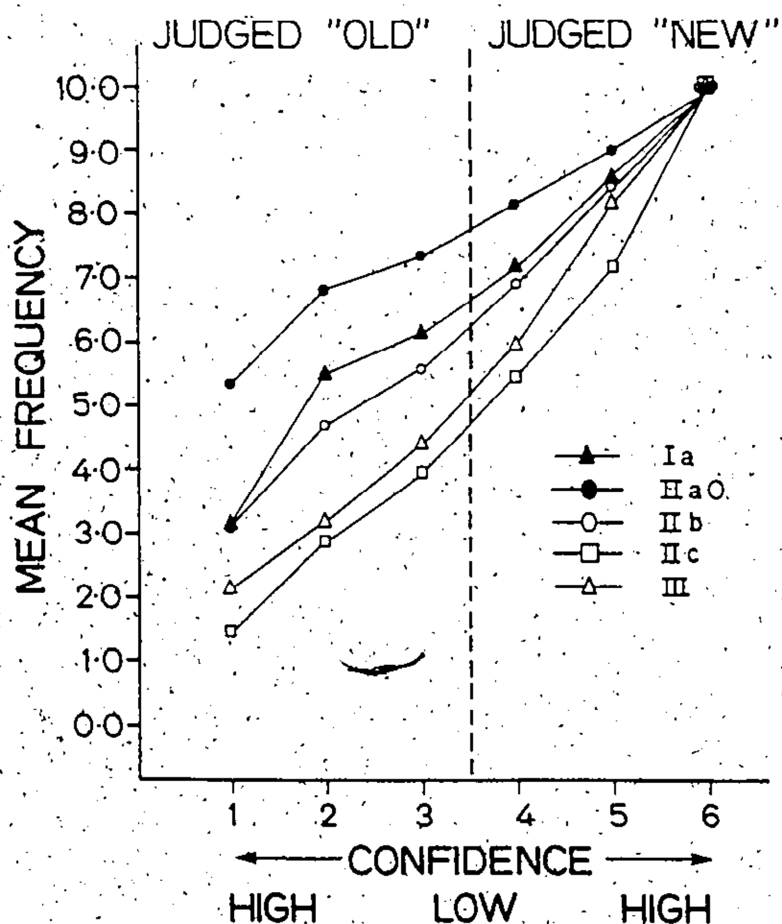


Fig. 16b. Cumulative mean frequencies of recognition-confidence judgements from Experiment 8.

stimulus types across confidence intervals really differ from one another: whether different stimulus types actually give rise to differing patterns of performance. The answer to this was determined through comparing adjacent curves, employing chi-square goodness-of-fit tests. (Because goodness-of-fit is assymetric, comparisons were made in both directions, and the more conservative value of chi-square was accepted for the comparison; and because four tests were made on the same data pool, the minimally accepted probability of alpha was (.05 x 1/4 =) .0125 .) This analysis indicated that the members of all pairs of adjacent curves, that is of the pairs IIa-Ia, Ia-IIb, IIb-III and III-IIc, differ from each other (p < .01 for each pairwise comparison, although the pair Ia-IIb differ only marginally). Different stimulus types apparently do sponsor different patterns of performance.

The next question is whether these different patterns of recognition performance could be accounted for in terms of the similarity of the item judged to items actually seen previously: whether the probability of judging an item to be old could be accounted for in terms of its similarity to the prototype, to old items in general or to particular old items. In an attempt to characterize the differences between performance patterns in a form which could answer such questions, the data were recast in three different ways. For the first of these, it will be recalled that each subject made 10 compound recognition-confidence judgements for each stimulus type. Fig. 17a plots the mean of these judgements for each stimulus type. The ordinate is the six-point scale upon which subjects made their judgements: it is read from high-confidence old at the top to high-confidence new at the bottom, with low confidence in the middle. The figure indicates that, although four-fifths of the transfer items are actually novel, only

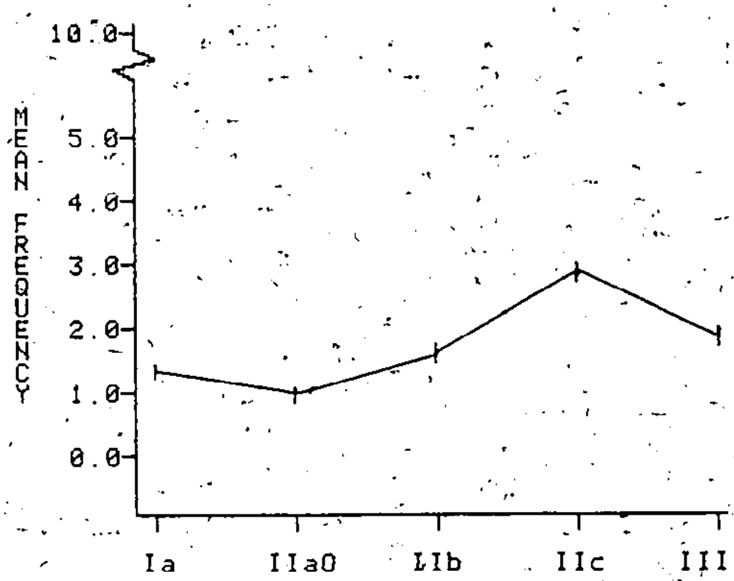


Fig. 17b. Mean frequency of judgements of high-confidence "New". Non-significantly differing pairs (p > .01) are IIaO - Ia, Ia - IIb and IIb - III.

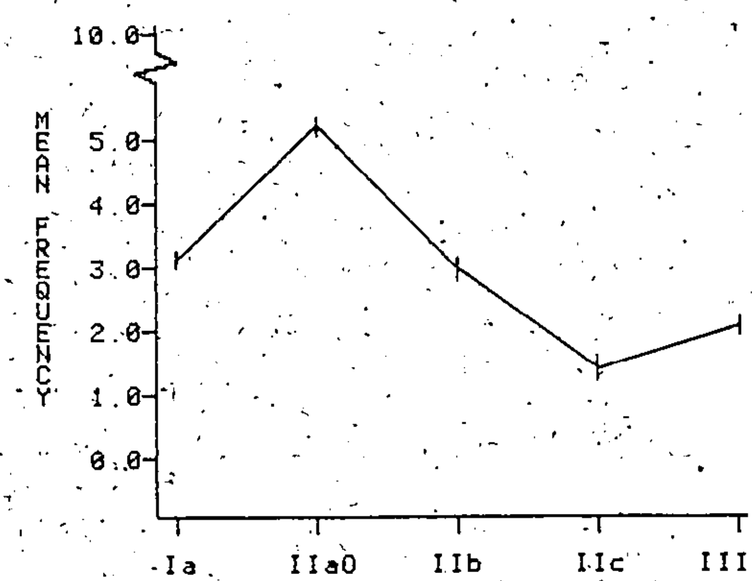


Fig. 17c. Mean frequency of judgements of high-confidence "Old". Non-significantly differing pairs (p > .01) are Ia - IIb and IIc - III.

the IIc and III items are judged on average to be novel, and those only with marginal confidence. Thus as a result of the training, many novel items appear familiar to subjects.

The second and third ways of illustrating the differences among the performance patterns are shown in Figs. 17b and 17c, which highlight information already presented in Fig. 16a. These figures pick out the differences among stimulus types in frequency of respectively high-confidence new and high-confidence old judgements. These two intervals were selected because they seem to carry most of the differences between performance patterns. It should be noted that there is no intention to portray these three figures as independent sources of information providing convergent evidence; rather, the three are presented as mutual checks, ensuring that whatever pattern of differences among the various performance curves is decided upon fairly represents the real differences between stimulus types detected in the distribution plotted in Fig. 16. Differences between adjacent points plotted in all three figures were tested via the Wilcoxon ranks test; non-significantly different pairs are indicated on the relevant figure.

These three figures clarify the pattern of differences among the five transfer stimulus types. All three figures yield the same ordering on the recognition-confidence scale: $\underline{IIa_O}$ items are judged to be most clearly old, followed by Ia and IIb items to-gether (with Ia perhaps judged old more confidently-frequently than IIb), then followed by III items, and finally by IIc items, which are judged to be most clearly novel. This ordering is a characterization of the differences in recognition performance among the five stimulus types; it is this ordering which must be accounted for through the similarity of transfer items to some memorial representation of the training items.

TRAINING	TRANSFER
P	P
Ia Ib	$\boxed{\text{Ia}}$ Ib
$\boxed{\text{IIa}}$ IIb IIc	$\boxed{\text{IIa}}$ $\boxed{\text{IIb}}$ $\boxed{\text{IIc}}$
III	$\boxed{\text{III}}$
IV	IV
V	V

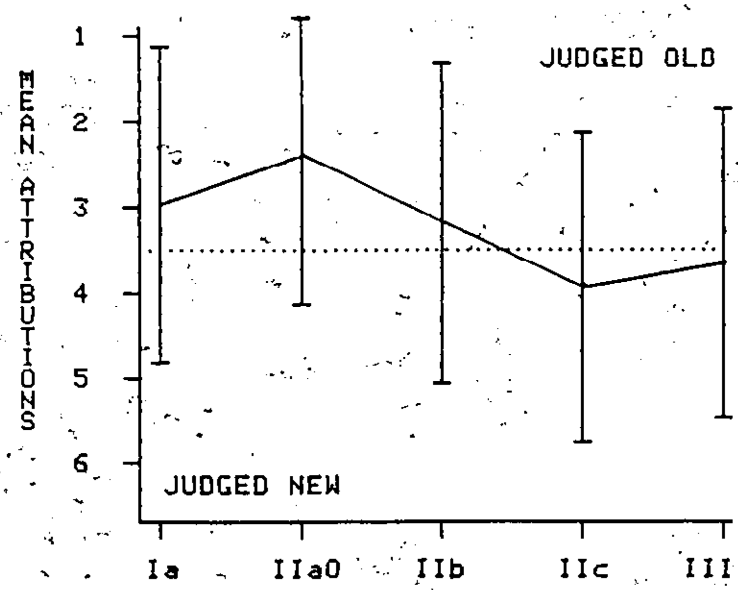


Fig. 17a. Mean composite recognition-confidence judgements from Experiment 8. All pairs differ significantly (p < .01) except Ia - IIb.

This ordering of transfer items is familiar from the last section. In that section, it was shown that the local performance surface surrounding an old item peaked at the old item, fell off symmetrically around the old in the around-the-prototype dimension ($\underline{IIa}_O$ > <u>IIb</u> > <u>IIc</u>), and assymetrically in the out-from-the-prototype dimension ($\underline{IIa}_O$ > <u>Ia</u> > <u>III</u>), falling off more steeply toward the periphery. However, it was impossible to judge accurately from that data the relative height of probability density at two points on the different dimensions, since they were obtained between-subjects: for example, the relative perceptibility of <u>Ia</u> items from Experiment 7 and <u>IIb</u> items from Experiment 6 could not be assessed. The ordering of recognition obtained in Experiment 8 provides a more complete look at what appears to be exactly the same performance transfer surface. Figs. 18a and 18b compare the perceptual identification results from Experiments 6 and 7 with the relevant recognition results from Experiment 8. These figures show a clear resemblance between performance in identification and recognition.

This resemblance is not a happy accident: it would be predicted from trace notions of perception and memory. Various memory theorists (e.g. Kolers, 1974; Mandler, 1980) have suggested that there are two forms or bases of recognition memory. Jacoby and Dallas (1981), finding independence between recognition and perceptual identification, argued that subjects could alternatively base recognition judgements on the fluency of perception of the target or on the reinstatement of the extra-item context, while perceptual identification must be primarily based upon fluency alone. In support of this idea, Jacoby and Witherspoon (1982) reduced the retrievability of study context through minimizing the meaningfulness of items (using pseudowords), and were able to demonstrate

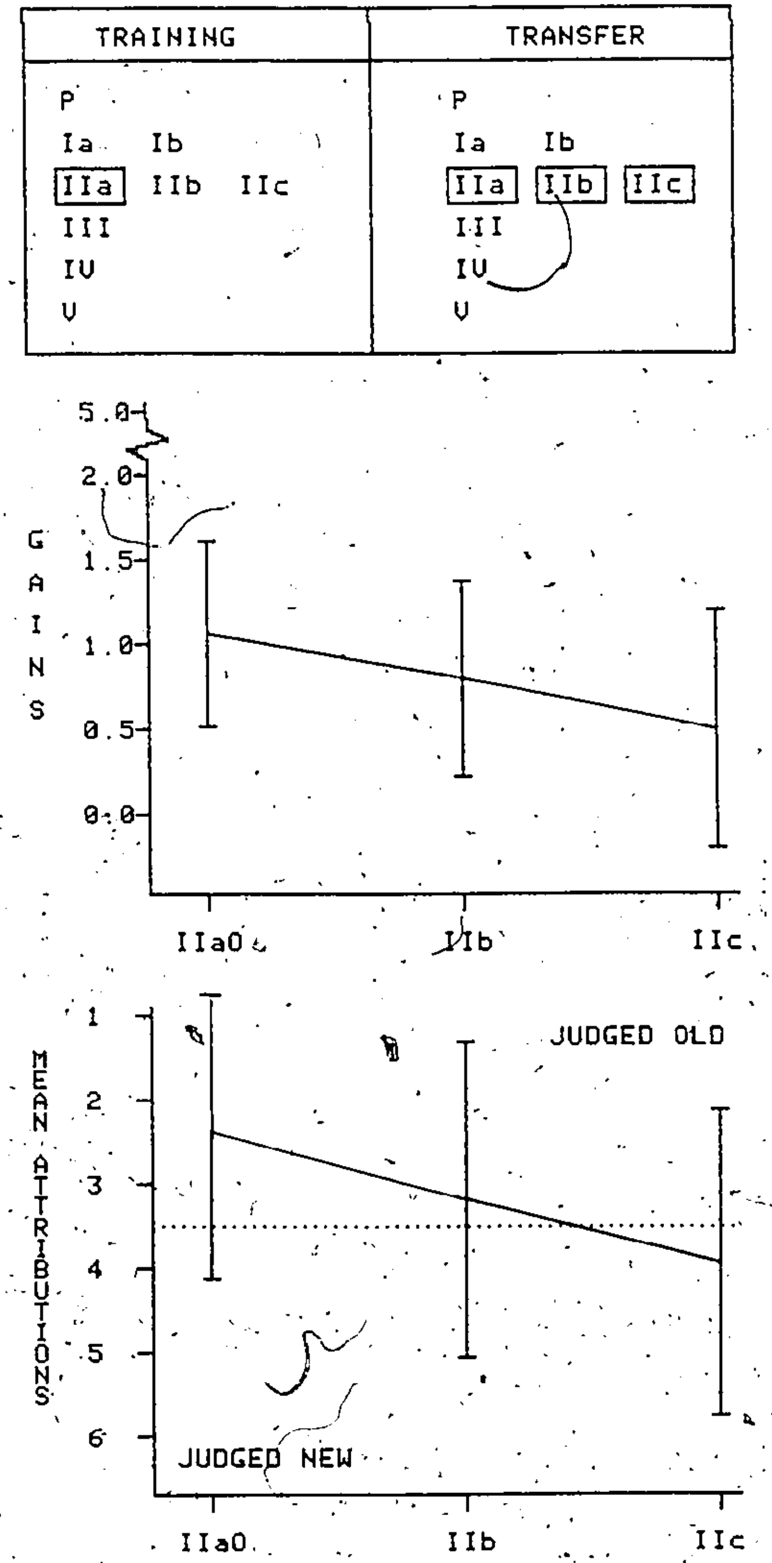


Fig. 18a. Comparison of perceptual identification scores from Experiment 6 (above) with recognition scores from Experiment 8 (below).

TRAINING	TRANSFER
P	P
Ia Ib	$\boxed{\text{Ia}}$ Ib
$\boxed{\text{IIa}}$ IIb IIc	$\boxed{\text{IIa}}$ IIb IIc
III	$\boxed{\text{III}}$
IV	IV
V	V

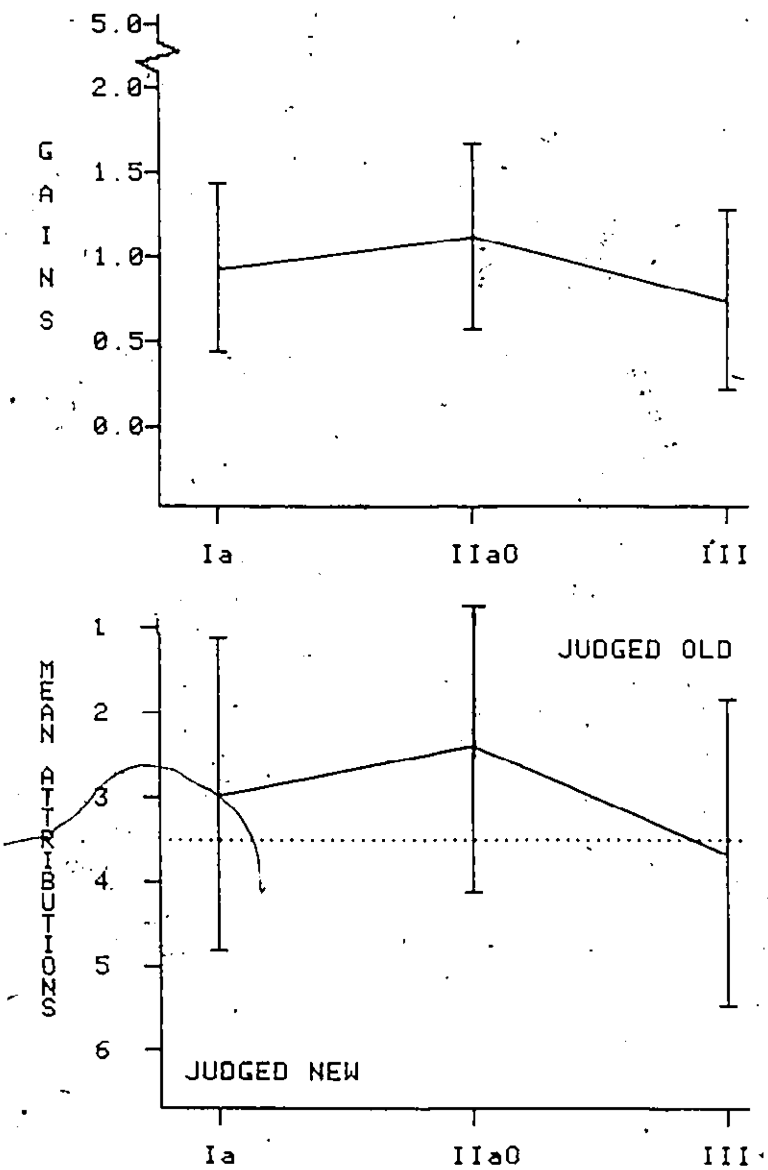


Fig. 18b. Comparison of perceptual identification scores of Experiment 7 (above) with recognition scores of Experiment 8 (below).

dependency under those conditions between recognition and perceptual identification. In the present experiments, not only were pseudowords used as stimuli, but also all differences in extra-item context were held to a minimum. Hence one would expect that recognition would be chiefly a function of the fluency of perception of the item, and would therefore be strongly related to perceptual identification performance on the same items. However, this insight does not provide leverage on the nature of the representation supporting the fluency.

Having discovered that the recognition transfer surface is arguably the same as the identification surface, and having noted a probable explanation, it is profitable to return to an examination of the slope of their putative underlying generalization gradients. In this discussion, the term "resource" will be used to refer to an item which has been previously processed, and which in consequence may be thought to have some effect upon the processing of subsequent items. This term was selected to avoid the structural connotations of the term "trace", but is not intended to convey anything other than the availability to the system of the experiences of prior processing.

A preliminary attempt was made to account for the saddle-shape of the transfer surfaces in terms of a set of generalization gradients around old items. Construction of such gradients requires a minimum of three pieces of information: a scale of the objective (experimenter-defined) similarity between a probe item and an old (resource) item in memory, a scale of the subjective (functional) similarity between probe and assumed resource (predictive of the amount of facilitation of performance rendered to the probe by utilization of the resource), and a parameter relating the two scales. For the preliminary attempt, and all subsequent versions, the objective scale selected was the proportional overlap in letters-in-position between resource and probe; that is, a scale of distance in terms of the experimenter-defined features. These scale units were selected so that the resulting generalization gradients could be interpreted in terms of something already familiar. As argued above, these units are approximately those the prototype perspective would prefer to employ, since they approximately maximize the covariation of sub-aspects of items; from the instance perspective, they are merely convenient units. The scale runs from a minimum of zero (no letters-in-position overlapping) to 1 (all five letters-in-position identical between resource and probe).

The scale of functional similarity selected consisted of convenient units. Such an arbitrary-sounding choice is justified by the fact that the goal of these constructions is to account for the ordinal relations among the degree of facilitation experienced by various stimulus types, not their absolute values. The values selected were the same values used on the objective distance scale, using endpoints zero and unity.

This leaves free only the parameter(s) describing the relationship between the two scales (a single parameter was fitted for all following analyses). The convention is adopted of calling this parameter $\underline{r}$, in specifying the functional similarity (S) between a probe and assumed resource as

$$S = \left(\left(\sum_{i=1}^{n} (d_i)\right) / n\right)^r$$

$$= O^r$$

where d_i is the value of the match ($d_i = 1$) or non-match ($d_i = 0$) between a resource and probe on the $\underline{i}$th feature, and each stimulus possesses n features, such that $\underline{O}$ is the proportional objective overlap between the resource and probe. The first attempt to fit this parameter used the value $\underline{r} = 1$, a value selected for reasons which will become evident directly. (The parameter $\underline{r}$ will be identified below with the Minkowsky distance metric.)

Fig. 19a illustrates the resultant plot of similarity vs. objective distance. An assumption of these analyses is that perception of an item is facilitated to the extent that that item is similar to a resource in memory, whether that resource is considered to be the prototype or an encoded old item. What this figure really illustrates is a transformation between the experimenter-defined distance between a probe-resource pair and the functional similarity between them, given the subjects' mode of processing the items. If gradients defined by $\underline{r} = 1$ account for the order of facilitation of various stimulus types, then a good description of subjects' performance is that they act as if they are computing the similarity between resource and probe in terms of the sum of experimenter-defined features overlapping between the two. In this case the facilitation value of a resource for a probe is just the proportion of letters-in-position shared between them.

This account specifies the relationship between one probe and one resource. It is a sufficient account to attempt to fit the parameter $\underline{r}$ for models which use a single known resource on each trial, for example the prototype and deterministic nearest-neighbour models. The assumptions of this account fit the prototype perspective very well. As pointed out above, the experimenter-defined features which constitute the objective distance scale are non-arbitrarily the best candidates for features under the assumptions of the prototype perspective, since in the current experimental domain they are the largest nearly invariant units.

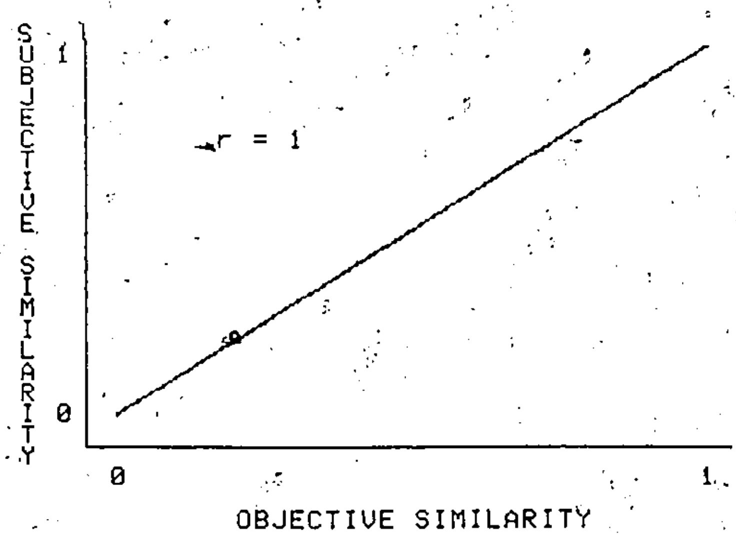


Fig 19a. Plot of objective vs. subjective similarity for r = 1.

The assumption that the scale of facilitation is identical to this scale of feature overlap, which is instantiated by the assumption that $r = 1$, is also a natural one for the prototype perspective. It is implied by the independence processing assumption which is a feature of nearly all prototype models. The assumption that the subject attends to and stores information about features independently is identical to assuming that facilitation of items is an additive function of the match – non-match values of the separate features: that is, that facilitation is a direct function of the proportion of overlap in experimenter-defined features between the prototype and a probe item.

For simple independent prototype models, the experimenter-defined similarity values for the five types of probe items can be computed as the sum of overlapping letters-in-position between the prototype pattern and the probe. For the average stimulus of each of the various types in Experiments 6 to 8, these values are Ia: 4, IIa$_0$: 3, IIb: 3, IIc: 3 and III: 2. (These values may of course be reduced to proportions by dividing each by five, for the five positions of each string.) Under the assumption that $r = 1$, the computed similarity values are the same as the objective distance values. This means that Ia items should be perceived most fluently, followed by IIa$_0$, IIb and IIc in a tie, followed by III items. However, as indicated above, the obtained order is IIa$_0$ best perceived, followed by Ia and IIb in a tie, followed by III, and finally IIc. Evidently the simple independent prototype model provides a very poor fit to the recognition data, as it was seen to do to the perceptual identification data in Experiments 4 and 5.

For complex independent prototypes, the calculation of distance from a probe to the prototype representation is more problematic. This may be why prototype researchers in general have ignored the representation of variability in assessing the predictions of prototype models. Taking a comparatively straight-forward case, the prototype might be thought to store modal and departure information, with no bias differentially weighting the access to one or the other. In this case, the objective similarity of a probe to the prototype may be computed as the number of matches with either modal or departure values presented in a particular position during training. Under this assumption, <u>every</u> element of <u>every</u> type of stimulus is found to be matched either by a modal or departure value stored in the prototype. Hence all transfer items are predicted to receive equal and complete facilitation, which is obviously not the case. Under a more complex set of assumptions, the prototype representation may be thought to be biased toward elements in proportion to their relative presentation frequency. Since in this preparation (Experiment 3) modal letters occur in each string position on 3/5 of trials, and a particular departure occurs on 1/5 of trials, it can be computed that the objective similarity values are <u>Ia</u>: 2.6, <u>IIa</u>$_O$: 2.2, <u>IIb</u>: 2.2, <u>IIc</u>: 2.2, and <u>III</u>: 1.8. (This set of values is identical to that obtained by determining the distance in overlaps of each transfer item to all old items. Thus this model is formally identical to an average distance model, which can be thought of as a multiple-resource, independent instance model. An example of such a model is discussed in the next section.) This model yields an order of transfer <u>Ia</u> > <u>IIa</u>$_O$, <u>IIb</u>, <u>IIc</u> > <u>III</u>, which is a very poor fit to the obtained order. Thus the complex independent prototype model also fails to account for the observed recognition data.

The deterministic nearest-neighbour model fares a little better. Under this notion, the experimenter-defined similarity values between target items and their nearest old item (a $\underline{IIa}_O$ item) are $\underline{Ia}$: 4, $\underline{IIa}_O$: 5, $\underline{IIb}$: 4, $\underline{IIc}$: 3 and $\underline{III}$: 4. Again, under the assumption of $\underline{r} = 1$, the computed facilitation values are the same as the experimenter-defined similarity values. This gives an order of facilitation $\underline{IIa}_O$ perceived most fluently, followed by $\underline{Ia}$, $\underline{IIb}$ and $\underline{III}$ in a tie, followed by $\underline{IIc}$. This is almost exactly the obtained order, except that Experiments 7 and 8 discovered that $\underline{III}$ items were more poorly perceived or less recognized than $\underline{Ia}$ or $\underline{IIb}$ items. This evidence is strong enough to discredit this model.

Since the independence assumption fails to provide predictive single-resource models, it may be abandoned, along with the parameter value $\underline{r} = 1$ which it implies. It's place may be taken by parameter values from zero to infinity. To understand the impact of these parameter values it is important to understand the essence of the objective similarity scale. Items in the present experimental domain are considered to consist of five elements. Each element can be considered to be a value taken from a pool of possible values for a particular position of the string, that is, from a nominal dimension. This is how the stimuli were conceptualized for their construction. It would appear necessary to employ a 5-dimensional matrix to index the similarity of two such items. However, as indicated in Appendix 1, their similarity can be coded by a single value, a count of the number of matches occuring over the five positions (although the information regarding the particular dimensions on which concordance occurs is lost). That is conceptually how the objective distance scale which in part defines the underlying generalization gradients is constructed. It is the value of a probe item on this scale that is raised to the $\underline{r}$th power to compute the similarity index; that is, similarity is conceptualized as the power of the sum of positional matches, not the sum of the powers of individual positional dimensional values (see Appendix 4).

We can now consider the meaning of raising this sum to various exponents. Table 8a compares values on the objective and subjective similarity scales for the parameter value $r = 1$. (This relationship was plotted in Fig. 19a.) The two sets of values are of course identical. Table 8b repeats this exercise for $r = 2$; the relationship is plotted in Fig. 19b. The plot demonstrates that subjective similarity of a probe and resource falls off relatively rapidly compared to the objective similarity. Compared to the case where $r = 1$, objective similarity (feature overlap) must be much greater to effect the same degree of facilitation. This is interpreted to mean that the processing system is attending to information at a higher level of integration than the individual letter-in-position: to some degree, inter-positional contingency information is being used as part of the basis for similarity comparison. It suggests that the target and probe are being compared at a relatively holistic level, relative to the letter-by-letter comparison entailed by $r = 1$. As can be imagined, increases in the value of r toward infinity have the result of steepening the slope of this gradient, until at an infinitely high value of r the facilitation index value for a probe and resource that are identical in letters-in-position is one, while a difference in any letter-position drives this index value to near zero (Fig. 19c, Table 8c). At this point the whole string can be thought of as the functional unit of similarity, the unit of processing. In practice, values of r greater than about five behave almost indiscriminably similarly to the infinite value. For this reason, if the attempt to account for the order of facilitation of the various stimulus types requires a value of r in the range of five upward, we have good cause to describe the basis of similarity comparison as holisitic.

Table 8

Comparison of Objective and Subjective Similarity under Various Metrics

Objective Similarity	Subjective Similarity			
	a) r = 1	b) r = 2	c) r = 10	d) r = .5
0	0	0	0	0
.1	.1	.01	.0000000	.32
.2	.2	.04	.0000001	.45
.3	.3	.09	.000006	.55
.4	.4	.16	.0001	.63
.5	.5	.25	.001	.71
.6	.6	.36	.006	.77
.7	.7	.49	.028	.84
.8	.8	.64	.107	.89
.9	.9	.81	.349	.95
1.0	1.0	1.00	1.000	1.00

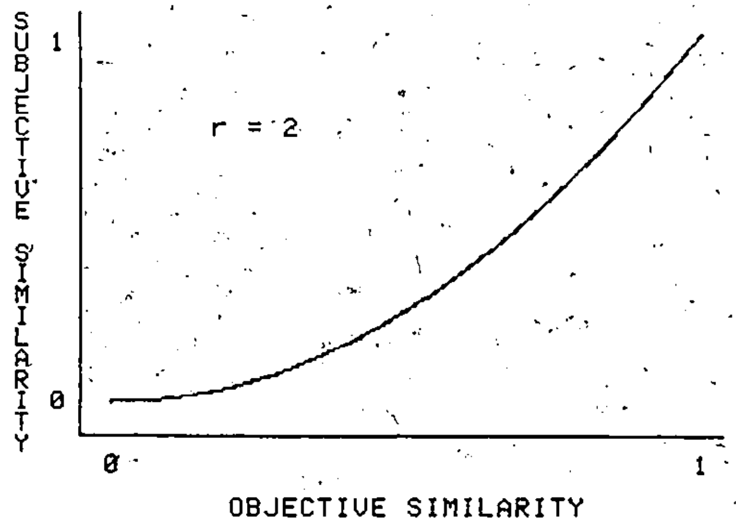


Fig. 19b. Plot of objective vs. subjective similarity for r = 2.

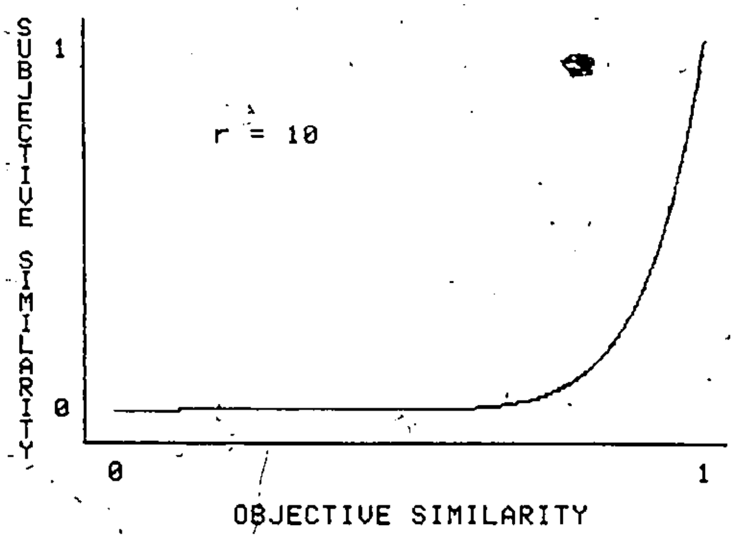


Fig. 19c. Plot of objective vs. subjective similarity for r = 10.

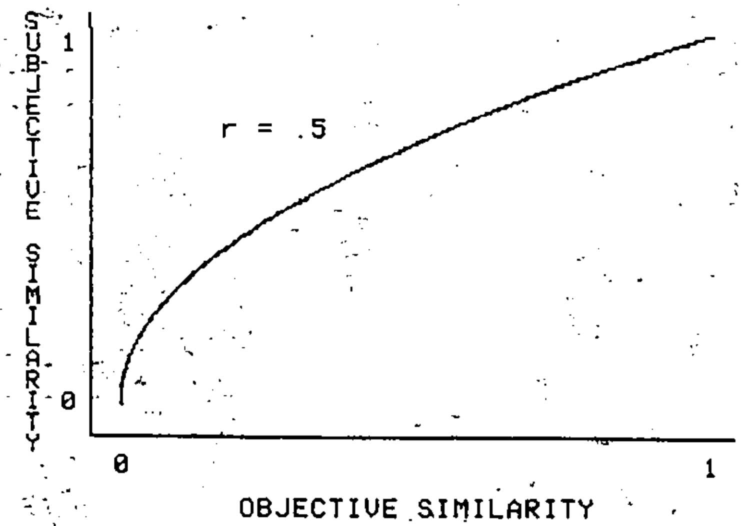


Fig. 19d. Plot of objective vs. subjective similarity for r = .5

In contrast, it is quite possible to consider values of $\underline{r}$ between zero and one (for an example of $\underline{r}$ = .5, see Table 8d and Fig. 19d). In this range of cases subjective similarity falls relatively slowly compared to objective similarity. These cases can be interpreted to indicate that the functional feature, the best-fitting unit of processing, is at a lower level of integration than the experimenter-defined feature, perhaps at the level of the letter segment. Similarity comparisons made on the basis of such levels could be described as relatively atomistic.

The plots traced out by differing values of the parameter $\underline{r}$ illustrate various slopes of the generalization gradient around a single resource in memory. At high levels of $\underline{r}$, the gradient is steep, such that facilitation occurs with great magnitude (or generalization occurs with high probability) only for items nearly identical to a resource in memory. As the value of $\underline{r}$ decreases toward zero, generalization to items of moderate objective similarity increases radically, such that for values of $\underline{r}$ near zero great or highly probable facilitation <u>fails</u> to occur only for items with no objective similarity to the resource.

The impact of these notions on models which employ a multitude of memorial resources on each trial, whether actually or potentially, is enormous, and will be discussed in the following section. However, they have less effect upon models which, like the simple prototype and deterministic nearest-neighbour models, access a single memorial representation on a given trial. Altering the parameter $\underline{r}$ from the unity value tested above, that is, dropping the independence assumption in favour of more atomistic or holistic bases of similarity comparison, does not alter the order of facilitation these models predict for the five stimulus types under study. That is because the effect of employing any value of r between zero and infinity is to alter only the ratio between each pair of stimuli for which the objective similarity values differ. However, the ordinal relations remain identical to those given above, and hence no manipulation of the value of this parameter will rescue these models. Thus a putative simple dependent prototype model would fare no better than did the simple independent model.

b) Multiple-Resource Models

Two kinds of models were mentioned above that employ multiple representations in memory on each trial. These are the probabilistic nearest neighbour and concert-of-instances notions (Brooks, 1978, personal communication). Both are instance notions: that is, both specify that concepts are represented memorially by a multitude of the categorical instances presented to the learner. They basically differ only in terms of whether the whole set of items in memory is actually accessed or only potentially accessed on a given trial. The concert notion suggests that all encoded episodes facilitate the perception of a probe, to the extent they are similar to the probe. In contrast, the probabilistic model suggests that only one memorial episode is accessed on a given trial, but will be accessed with probability dependent upon its similarity to the probe: the nearest neighbour is the likeliest to be accessed, but more distant neighbours have some probability of being accessed. Data averaged over a series of trials probably cannot not be used to separate these two notions, since the average facilitation over several probabilistic accesses would be expected to equal that afforded by a single, parallel access of many instances of differing similarity. However, these notions can be dealt with simultaneously through the notion of underlying generalization gradients introduced in the last section. They differ on whether these gradients represent probability distributions of access of a single resource out of many, or distributions of facilitation due to parallel access of multiple items: but the gradients predicted by the two models are mathematically similar, and make similar predictions about the relative facilitation of items.

The notions introduced in the last section about generalization gradients apply only to the slope of the gradient around a particular trace in memory. In order to apply those notions to multiple-resource models, one further piece of conceptual apparatus is needed: some means of combining the strength or probability inputs of multiple memorial representations, that is, some means of combining the separate gradients of particular resources. The following simple notion is submitted as an initial approximation. The facilitation of perception of a probe due to access of multiple representations is some function of the psychological similarity of the probe to all representations in memory. For current explanatory purposes, perceptual facilitation is taken to be an identity function of subjective similarity. (See Appendix IV for a further discussion of this function.) The psychological similarity is a function of the overlap of the probe with each of the representations in memory. Letting ST be the total functional similarity of a probe with all items in memory, S_i be the total functional similarity of a probe with a particular resource, and O_i be the overlap of a probe with a particular resource,

1) $ST = \operatorname{sum}_{i=1}^{n} (S_i)$, and

2) $S_i = O_i^{r_i}$,

where r_i is a parameter reflecting the degree of integration of encoding of the representation of the ith resource, and n is the number of representations accessed or potentially accessed to facilitate the probe, and

3) $O_i = (\sum_{j=1}^{m} (d_{ij})) / m$,

where d_{ij} is the match ($d_{ij} = 1$) or non-match ($d_{ij} = 0$) of the jth invariant element of a probe with the ith memorial representation, and m is the number of experimenter-defined elements in a probe.

Line 3) defines O_i as the proportional overlap between a probe and a given resource in experimenter-defined features, the invariant elements. The five values a probe takes on the five positions (nominal dimensions) are transformed into a single value on a single dimension of objective similarity to a given resource. Line 2) defines the functional similarity of the probe to a particular resource as a function of both its overlap in experimenter-defined elements with the probe and its level of integration. Lines 2) and 3) capture all the notions about the generalization gradient around each individual resource that were discussed in the last section. Line 1) indicates that the total functional similarity of a probe is the sum of the similarities of the probe to each resource it (actually or potentially) accesses.

The form of these functions may be made more clear by noting that in this scheme proximity is a function of similarity such that the proximity of a probe to all representations is

4) $$\text{proximity} = (\text{similarity})^{1/r} = (O_1^{\,r} + O_2^{\,r} + \ldots + O_n^{\,r})^{1/r}.$$

In words, total proximity of a probe to a set of resources equals the rth root of the total similarity. This equation defining proximity in terms of overlaps evidently takes the form of the Minkowsky distance formula (Torgerson, 1958). (Familiar instantiations of the formula are the Euclidean metric, or Pythagorean theorem, for which $r = 2$, and the city-block metric, for which $r = 1$). Taking the rth root of the total similarity to determine proximity simply returns the computation to the original scale values. (This is the purpose served by taking the square root of the sum of squares in using the Pythagorean theorem.) Thus the computation of similarity above rests on the transformation of a set of distances on experimenter-defined dimensions scaled in experimenter-defined units to a single value on a subjective dimension scaled in subjective units. Taking the rth root of this value returns the scale of measurement to experimenter-defined units. The similarity and proximity thus computed are monotonically related through the parameter r. Both are estimates of the amount or likelihood of facilitation of a probe; they are indices of the psychologically effective distance between the probe and memorial resources, given that the subject has processed the resources in a particular manner.

The value of the parameter r for the comparison reflects the degree of integration at which a resource was encoded, and defines the slope of its facilitation gradient. (The probe item has no parameter, since this model describes perception. At the beginning of the process, the probe is simply "out there": it is not integrated at any level, since it has not yet been processed at all.) At high values of r any difference between probe and resource makes all the difference: any change in the probe from identity with the resource causes the facilitation index to fall from unity to near-zero. At lower values of r, changes in the overlap of a probe-resource pair have a less dramatic effect on predicted facilitation. Essentially the value of r defines the level of the processing unit, the effective resource. Under r = 1 each element of each presented item is in effect a separate resource, contributing independently to the facilitation of the probe, whereas under higher metrics the elements are clumped into resource packages.

The effect of this formulation is to add the facilitation due to the separate memorial resources. This instantiates the parallel processing suggested by the concert of instances. It can be conceptualized as determining the simultaneous distance of a probe from multiple resources. In effect each resource is a dimension of an n-space in which we are attempting to determine the distance of a point (the probe) from the origin (which represents identity with all resources). The sum of overlaps between the probe and each resource is the value on that resource-dimension the probe takes in n-resource space. Conceptualized in this fashion, the parameter r represents the curvature of the space. This curvature represents the transformation of scale required to obtain distance estimates computed from the scale employed on the axes of the resource space which are commensurate with distances measured by some other device (a ruler or test) which assumes a different scale unit.

Another perspective on the significance of the parameter r is that it informs us of the degree or likelihood of use of each of the multiple available resources relative to each other. For r = 1, the city-block metric, alteration of the value of any entire resource (O_i) has an effect on the total outcome (ST) exactly equivalent to the value of the change in the single resource: a unit change in the similarity to one resource causes a unit change in the outcome. Thus this metric defines statistical independence, or separability in Garner's (1974) terms, of the resources. (Change in the similarity of probe to only one resource is likely to be difficult in clustered spaces, but this difficulty is irrelevant to the logic.) By contrast, for r = 2, the Euclidean metric, the resources are not independent. The actual value of outcome change depends on the magnitude of the original value changed by one unit. A unit change from .2 to .4 objective match causes less change in the total facilitation of the probe than a unit change from .8 to 1.0. In Garner's terms, this outcome defines integrality of the elements within each resource. For r = .5, a unit change causes more than a unit change in outcome: this is the obverse case of integrality (interaction of components). Another important case is where r approaches infinity, which is called the dominance metric by Garner. In this case, the only change which significantly affects the outcome is a change in the value of the resource which has the greatest value of all resources. In effect this metric predicts facilitation solely in terms of the nearest neighbour. A unit change in the objective similarity of the probe to the resource with highest overlap value results in various outcome changes, from near-zero if the original objective value is low, to near-unity if the original objective value is high.

This formulation gives us a descriptive model of concept formation with processing implications. It is a single-parameter model, the parameter being r, the degree of integration of encoding the old items. For purposes of this paper, this model will be called the "episode model", to reflect its emphasis on representation by means of encoding experiences of processing particular stimuli in a particular way. As indicated above, the extent of the processing unit which will determine the breadth of transfer (the functional encoded episode) depends on the level of integration at which old stimuli were processed, from relatively atomistic to relatively holistic experience of the items. The level of processing of old stimuli in turn is thought to depend on encoding variables such as strategy, task requirements and prior experience of other members of the domain. The task requirement of string copying is not manipulated in this paper, so that predictions of the episode model for variations of processing are not emphasized here. However, the ability of the model to account for performance in stimulus domains differing radically in distribution of density is tested in this section and in section V iv, and is discussed more fully in section V v.

Of course, various old items may have been encoded at various values of $\underline{r}$ during the course of the presentation of a set of stimuli. For present purposes, it is assumed that all old items have been encoded at a single level of integration, as a result of the encoding requirement (writing) being held constant. It is also possible that items are encoded in separate, integrated chunks (like syllables). While the model is in principle capable of dealing with such an eventuality, for present heuristic purposes the simplifying assumption was made that syllabification does not occur.

This model was fitted to the recognition data, attempting to discover a value of $\underline{r}$ which would predict the order of facilitation observed. Facilitation scores for the five types of transfer stimuli, derived from this model under the assumptions $\underline{r}$ = 1, 2, 2.5, 3, 4 and 10, are given in Table 9. As indicated above, the observed order of transfer derived from the recognition data (Experiment 8),

Table 9

Facilitation Predicted from Single Resources

under Various Metrics

Probe-Resource Overlap	Predicted Facilitation					
	r = 1	r = 2	r = 2.5	r = 3	r = 4	r = 10
5	1.00	1.00	1.00	1.00	1.00	1.00
4	.80	.64	.57	.51	.41	.11
3	.60	.36	.28	.22	.13	.01
2	.40	.16	.10	.06	.03	.00
1	.20	.04	.02	.01	.00	.00
0	.00	.00	.00	.00	.00	.00

and supported by the latter two perception experiments (Experiments 6 and 7), is $\underline{IIa_O} > \underline{Ia}, \underline{IIb} > \underline{III} > \underline{IIc}$. Table 10 gives the predicted facilitation score for a probe of each stimulus type. These scores were computed by adding the facilitation predicted for each member of a stimulus type by each of the ten training items (irrespective of the category of the training item), and then averaging these item-facilitation sums for each probe type, under the assumption that training items were $\underline{IIa_O}$. For example, any particular Ia item overlaps the set of within-category training items by 4, 3, 2, 2, and 2 letters, and other-category training items by 1, 0, 0, 0, and 0 letters. Table 9 indicates that for $\underline{r} = 1$ these overlaps predict separate facilitations of .3, .6, .4, .4 and .4 within-category, and .2, 0, 0, 0 and 0 from items in the other category. The sum of these individual facilitations is 2.8, which is entered in Table 10 as the facilitation from multiple resources for a Ia item under $\underline{r} = 1$.

Each column of Table 10 can be interpreted to make ordinal predictions concerning the relative magnitude of facilitation experienced by the five types of probe stimuli. For $\underline{r} = 1$ the predicted order of transfer from Table 10 is $\underline{Ia} > \underline{IIa_O}, \underline{IIb}, \underline{IIc} > \underline{III}$, which is evidently wrong. (This is is the multiple-resource independent instance model alluded to in the last section as being formally identical to a complex independent prototype model.) For $\underline{r} = 2$, the Euclidean metric, the predicted order is $\underline{Ia} = \underline{IIa_O} > \underline{IIb} > \underline{IIc} = \underline{III}$, which again is not the observed order. However, $\underline{r} = 2.5$ is a break point in predicted ordering: for this value of $\underline{r}$ the predicted order is $\underline{IIa_O} > \underline{Ia} > \underline{IIb} > \underline{III} > \underline{IIb}$. For the first time, old items are predicted to be best perceived (although only marginally so), and also for the first time, III items are predicted better perceived than IIc items (again, very marginally). However, Ia items are still

Table 10

Facilitation Predicted from Multiple Resources

Under Various Metrics

Probe Type	Average Predicted Facilitation					
	r = 1	r = 2	r = 2.5	r = 3	r = 4	r = 10
Prototype	3.00	1.80	1.40	1.10	.65	.00
Ia	2.80	1.52	1.17	.92	.63	.12
IIa	2.60	1.56	1.34	1.20	1.09	1.00
IIb	2.60	1.32	.94	.83	.57	.12
IIc	2.60	1.16	.82	.59	.32	.02
III	2.40	1.20	.91	.71	.50	.11

evidently predicted better perceived than IIb items, which does not conform to the observed ordering. For $r = 3$ and still more clearly for $r = 4$, the predicted ordering converges on the observed ordering. At $r = 4$ there is little difference between Ia and IIb items, and the direction of the difference is the same as the direction of the non-significant difference between these types found in the recognition data. IIa_O items are clearly predicted to be better perceived than Ia items, and III items better than IIc items. The conclusion to be drawn is that this multiple-resource model, employing similarity of the probe to all old items, and carrying an integration parameter for the similarity to old items, is capable of post-dicting the ordinal relations in the recognition and perception data.

One would not wish to be very precise about the actual value of r to be deduced from this analysis. However, it clearly must be 2.5 or greater. This lower bound is not based on the ratio properties of the similarity scale, for which the measures of facilitation in these experiments are a little crude, but upon its ordinal properties, which are quite robust: a metric of less than 2.5 would not yield the ordering obtained in Experiment 8. It is noteworthy that the metric must be greater than the Euclidean: this issue is discussed in Appendix 3.

The upper bound on this value is less certain, except that if it were near infinity, the model would predict nearly no facilitation for any of the four novel types of items. Even for $r = 10$, as shown in Table 10, the model predicts nearly no facilitation for novel items in comparison to old items, which is clearly wrong, and suggests that facilitation of novel stimulus types will be approximately equal, such that their transfer scores will be indiscriminable, which is also clearly wrong. In light of these considerations, the value $r = 3$ was taken to be a reasonable approximation.

The evidence given above for the validity of the model is very crude. It is based on ordinal predictions only, and the value of the parameter is derived from rather than predicted for the data. However, there was good a priori reason to suspect that a high level of r would be required. The parameter r reflects the level of integration at which old items are encoded, that is, how close to holistic was the encoding. The encoding requirement for subjects was writing. Compared to other encoding requirements attempted in the early experiments, writing intuitively appears to be a task which would be likely to promote relatively holistic encoding, whereas a letter-search task or spelling task appears intuitively likely to promote relatively letter-by-letter encoding. Secondly, three experiments described above (Experiments 4, 5 and 6) demonstrated inter-letter dependency in performance, under encoding conditions identical to those of the present experiment. That dependency suggests that the target letter was not encoded independently of the other letters of the string: in other words, it suggests relatively holistic encoding. A set of experiments to be discussed below (in part iv of this section) provides some predictive validation of the model.

iii) Conceptual Perspectives on the Necessity of Abstraction Explanations

This section argues that a conceptually basic demonstration of the necessity of prototypes, upon which much of the prototype tradition has grounded its assumptions, is at least as well accounted for by distributed-representation assumptions, and hence fails to support the necessity of explanations appealing to abstraction of summary information. That basic demonstration is the classic Posner and Keele (1968) study which found classification of prototypical patterns not seen in training to be as good as that of patterns actually presented in training, and better than that of other patterns not seen during training. This study has been widely cited as clearly indicating the emergence of a prototype. In fact, this popular conception of the conclusions of their study is somewhat misleading. Posner and Keele actually concluded that information about individual patterns must be retained and employed in performance (i.e. that purely summary prototype models are insufficient). This aspect of their study has generally been ignored in citation of this paper (e.g. Rosch, 1977; Homa, Sterling and Trepel, 1981). However, they also concluded that their results necessitated postulating some degree of abstraction, and that the weakest statement consistent with their data is that the prototype has a higher probability of recognition than other novel patterns; they left open the possibility that abstracted information has a greater role. They interpreted their basic finding to mean that the prototype is a unique pattern, such that it is better recognized than other novel patterns having the same average distance to stored exemplars. They concluded that some information about the central tendency of the presented exemplars must have been abstracted to account for this uniqueness, since the distance of a probe from the old items could not account for the differential recognition of two items of equal dissimilarity to the old items.

The most popular version of prototypes, the pure-summary prototypes, have been shown to be contradicted by the experiments above in all versions save the complex independent, which (as discussed below) does not store general information. The strong-form hybrid model, under which instance information has a limited role in the processing of novel instances, was also disconfirmed above.

Thus the weakest form of the abstraction position identified by Posner and Keele is also the only general-information prototype position still tenable. It claims that some general information must be abstracted and stored in order that the prototype may possess its unique characteristic of being better recognized than other novel patterns equally dissimilar to the old patterns.

The argument for the weak hybrid position rests upon an argument from necessity, that instance-only explanations cannot in principle account for the difference between performance on the prototype pattern and on other novel items. An account of this finding based on an instance perspective would destroy the basis of the claim. Further, the weak hybrid model, in order to avoid triviality, must claim a role for prototype information not only in the recognition of the prototype pattern, but at least to some degree in the processing of other novel items. That is, the primary evidence for the abstraction of general information may come from the uniqueness of performance on the prototype pattern, but the model must predict an impact of this abstracted information on the processing of novel items other than the prototype in order to avoid being a theory about the processing of a very special category member that rarely occurs.

The prototypical instance itself was never presented as a transfer item in any of the experiments reported in this paper. This was due to a methodological difficulty: while it is possible to obtain many repeated measures on a single subject regarding his ability to perceive most types of stimulus, as defined by the distance of the type from the nearest old and from the prototype, without repeating actual items, this is impossible with the prototype pattern itself, since there is but one prototype pattern per category. Inclusion of the prototype pattern as a transfer item would thus have led to a methodological awkwardness, with measurements of other transfer types to the prototype pattern being obtained in a ratio of 5:1. However, it is not really necessary to actually measure performance on this pattern to debate the necessity of abstraction of the prototype. First, the episode model described above is perfectly capable of predicting the apparently unique status of the prototype pattern observed by Posner and Keele. For example, Table 10 presented the predicted facilitation of the prototype patterns of the categories used in the recognition experiment (Experiment 8). It is evident from this table that for $\underline{r} = 1$ the prototype pattern is predicted to be the most fluently-processed of all types of stimuli; and this prediction is made from a purely instances basis. The prototype pattern achieves this eminence under instance assumptions because of its central location and consequent moderate similarity to <u>all</u> old stimuli. While the prototype pattern does not have the greatest similarity to <u>any</u> old item, it is most similar under a city-block assumption to old items in general. Under this assumption, many moderate similarities outweigh a few large ones. For the set of old items used to compute Table 10, the prototype pattern remains predicted to be the best-perceived pattern under the Euclidean metric and even under $\underline{r} = 2.5$. It is not until $\underline{r} = 3$ for these stimuli that the prototype pattern is even reduced to a level of predicted facilitation <u>equal</u> to the old items (the pattern of results found by Posner and Keele). However, for higher values of $\underline{r}$ the prototype is predicted to experience relatively less facilitation compared with items having greater overlap with a <u>particular</u> training item. Thus depending on the level of integration of encoding of training instances, the prototype pattern may be predicted by an instance model to be better, equally or worse perceived than training items. Thus the finding that the prototype pattern is as well perceived as the training stimuli does not necessitate an abstractive explanation.

A second theoretical point is that Posner and Keele's observation that the prototype pattern is better recognized than other novel patterns with "nearly the same average distance from the stored exemplars" (p. 362) is subject to re-interpretation. It is evident that "average distance" depends on the metric of distance employed. Posner and Keele assumed a city-block metric was appropriate, without any substantiation. They thus used the mean of the city-block distances from the schema pattern to each presented exemplar, and from each novel item to each exemplar, as their measure of similarity. Now, the mean distance to training items calculated in this fashion for the schema pattern is almost identical to the mean distance for the other novel instances to the training patterns, although the schema pattern is a goodly distance from each novel item (see Appendix 2). It thus appears impressive that the schema pattern is consistently better recognized than the novel patterns: it indeed suggests that there is a unique property belonging to the prototype that is above and beyond similarity to the instances.

However, as indicated in the prologue, their schema pattern was not constrained to be the average of the presented stimuli. Rather it is a prime form from which the presented exemplars are generated by means of random distortions at some level of probability. Their schema is therefore unlikely to be exactly central, and in fact is not (see Appendix 2). There is axiomatically a single minimum average city-block distance to the four old exemplars. An item possessing this average distance is axiomatically centrally located (although there may, of course, be more than one such central location (i.e. multiple modes) when the training items are located at differing distances to each other). Any items possessing close to this minimum average distance must also be close to central. In fact, an item cannot possibly be central, and also a similar average distance from the old items as another item, and yet be much more central than that other item (although such items may be far apart): any items sharing the first two characteristics are inevitably very similar in their degree of centrality. By this reasoning, Posner and Keele's novel items are nearly as central as their schema pattern, which refutes the contention that while similar in distance to the training items, the schema pattern is dissimilar to the novel patterns in being more central. In fact, as indicated in Appendix 2, none of the transfer patterns, including the schema pattern, is very close to the central tendency. The schema pattern is in fact more like one of the novel patterns in distance to the training items than like the central tendency. It is therefore surprising, in the light of contemporary prototype theory (which identifies the prototype strongly with the central tendency) to find that subjects reliably recognize the schema pattern more than the novel patterns; and this finding is paradoxically difficult for a contemporary prototype theory to explain. An abstracted central tendency would certainly not have this effect. It would lead to approximately equal recognition of the schema and novel patterns, since they would be approximately equally similar to the abstracted information.

However, instance notions can account for these results quite easily. Although the average distance from the schema pattern to old distortions is approximately the same as that of novel items to that of old distortions under city-block assumptions, the variance of the dissimilarity of the schema pattern to olds is higher. In fact, two of the four schema-to-old dissimilarities are smaller than any of the twelve tabled novel-to-old dissimilarities, one being only three-quarters of the <u>smallest</u> novel-to-old dissimilarity. Under these conditions it would require a very small increase in the metric over the city-block to predict a strong assymmetry in the facilitation of the two types. The Euclidean metric would certainly do so; and given that Posner and Keele required learning to a criterion of two completely errorless passes through a list of twelve training stimuli, it is not unlikely that their subjects encoded these stimuli to some degree of holism beyond that described by $r = 1$.

iv) Empirical Perspectives on the Necessity of Abstraction Explanations

This section contains four experiments that simultaneously satisfy a number of objectives. First, they test the predictive capability of the episode model. They all assume the appropriate metric to be $r = 3$, based on the reasoning that since the training for these experiments is identical to that in Experiment 8, also consisting of copying the training strings, the level of integration of encoding should remain at about the same level. (It will be recalled that the value $r = 3$ is a rough approximation: heuristically, the specific value of r is less important than that its value be greater than one, the independence value.) Secondly, while the model and value of the parameter were derived from recognition data, the experiments to which the predictions are now applied employ perceptual identification as their dependent measure, and so the success of the predictions rests upon the reality of the convergence of these two measures in the current domain. Thirdly, these experiments empirically assess the claims of the weak hybrid notion, the last prototype model to be examined that bears any real resemblance to the original conception of central representation which motivated prototype theory.

The first experiment of this set (Experiment 9) was predicted to yield exactly Posner and Keele's (1968) results, that items closer to the central tendency are as well performed on as training items; however, this prediction was made from the episode model. It employed the three sets of stimuli IIa, IIb and IIc as training items. They were presented once only each, thus achieving the same number of learning trials (30) that had been used in all foregoing experiments. The experimenter-defined categories in this preparation have cue validities of .8. The stimulus sets Ia, IIa_O and III were employed as transfer items, as in Experiment 7 (see Section V i). The episode model, with $r = 3$, was used to predict the rank order of transfer scores, employing all 30 training items as resources. The mean similarity scores computed in this fashion were Ia: .92, IIa_O: .87 and III: .58, giving a predicted order Ia > IIa_O > III, with only a marginal difference predicted between Ia and IIa_O. (The city-block metric yields similarity scores of 2.8, 2.6, and 2.35 for Ia, IIa_O and III. This is the same ordinal prediction as for $r = 3$.)

The episode model's prediction using $r = 3$, the value derived from the recognition study above, closely parallels Posner and Keele's (1968) observation that highly prototypical items are as well perceived as patterns which have actually been presented to subjects: in the present case, novel near-prototypical items are predicted to be very slightly better perceived than training stimuli. This result has been conceptually basic to the prototype tradition, and has been argued to necessitate a prototype explanation (e.g. Posner and Keele, 1968; Rosch, 1977; Homa, Sterling and Trepel, 1981). However, as the present case illustrates, the episode model is capable of making the same prediction, even with a value of the integration parameter specifying relatively holistic encoding, given the distribution of training stimuli in this experiment.

The results are shown in Fig. 20. The means of stimulus types are Ia: .81, IIa_O: .78 and III: .58. The means of Ia and IIa_O are not significantly different, while both of those means are greater than the mean of III ($p < .05$) in both cases. The pattern is remarkably similar to that predicted by the episode model for $r = 3$. However, since both $r = 1$ and $r = 3$ predict the same order of transfer, and the data is really a little too crude for interval or ratio comparison, this experiment is not a strong test of whether subjects' coding is best described as independent or dependent. Rather, the importance of this study is that the success of the episode model's prediction adds empirical force to the theoretical arguments of the last section indicating that appeal to abstraction is unnecessary to account for good performance on items that are novel but very prototypical.

The results of Experiment 9 are considered to be important in part because they demonstrate that the episode model does not demand that items processed previously will always be better perceived than novel items. As is argued below, performance on items is not a function of their exposure status, but rather of their subjective similarity to all items which have been processed. To make this point more strongly, Experiment 10 was run as a replication of Experiment 9, except that instead of perceptual identification it employed recognition-confidence as the dependent measure, as in Experiment 8. Despite the change of dependent variable, the transfer prediction for Experiment 10 is the same as for Experiment 9, since, as argued above, in cases where differential extra-item

TRAINING	TRANSFER
P	P
Ia Ib	Ia Ib
IIa IIb IIc	IIa IIb IIc
III	III
IV	IV
V	V

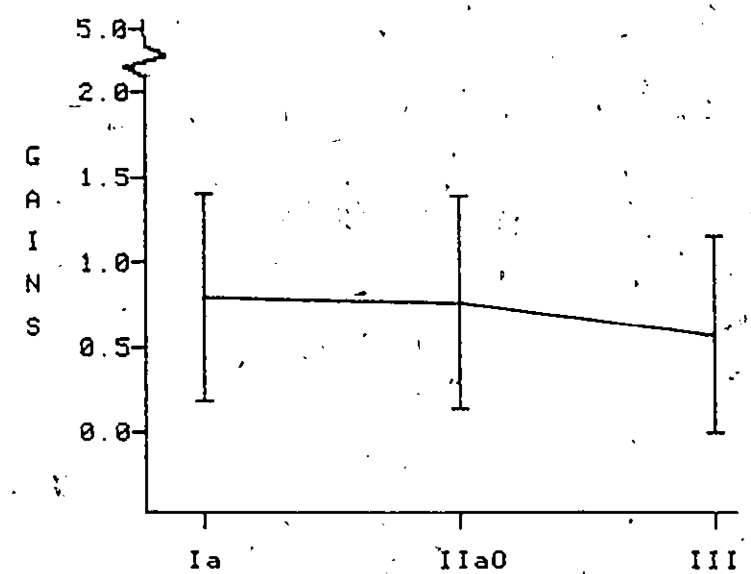


Fig. 20. Perceptual identification results for Experiment 9 (distributed training items).

context is lacking, recognition reflects simply the fluency of perception. For $\underline{r}$ = 3, therefore, the predicted order of transfer is $\underline{Ia}$ items judged marginally more likely Old than $\underline{IIa}_O$ items, and $\underline{IIa}_O$ items much more likely judged Old than $\underline{III}$ items.

The results are presented as they were for Experiment 8. Figs. 21a and b present mean frequencies and cumulative mean frequencies of recognition-confidence judgements. Chi-square goodness-of-fit tests on these patterns indicate that $\underline{Ia}$ and $\underline{IIa}_O$ are marginally different ($p < .02$), while the distribution of $\underline{III}$ items differs significantly from the other two ($p < .001$). These differences may be characterized through examination of Figs. 22a, b and c, which illustrate mean attributions, and mean frequency of high-confidence "New" and high-confidence "Old" judgements respectively. Mean attributions concerning $\underline{Ia}$ and $\underline{IIa}_O$ stimuli do not differ significantly (Fig. 22a), although the patterns of attributions differ slightly in that $\underline{Ia}$ items are more frequently judged both Old and New with high confidence than $\underline{IIa}_O$ items (i.e. with greater variability, which accounts for the difference indicated by the chi-square test: see Fig. 22b and c). Type $\underline{III}$ items are evidently treated as being more likely novel than either of the other types. As predicted, this pattern of results closely resembles that of Experiment 9. To-gether, these studies confirm the point that instance models need not demand that best performance is associated with previously-experienced items.

It is notable that this type of performance pattern (novel items performed on at the same level as old items) is achieved when the density of old items is relatively widely distributed in the domain. In the preparations used for Experiments 9 and 10, density of old items was not maximal at the central tendency (as it generally is in graded membership structures used to support the prototype

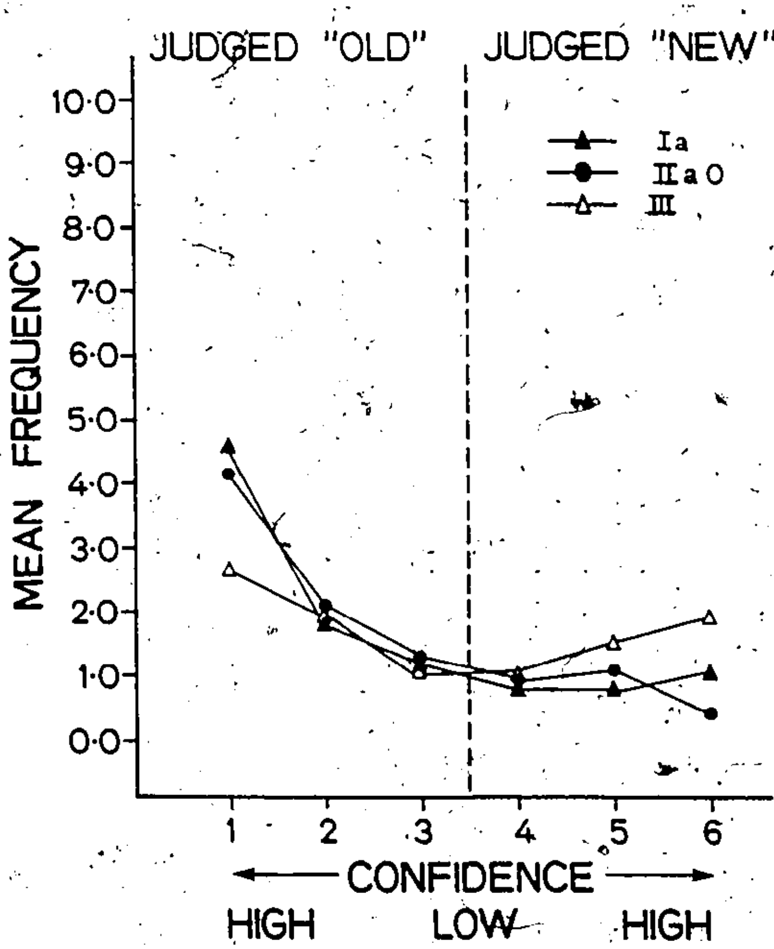


Fig. 21a. Mean frequencies of recognition-confidence judgements from Experiment 10.

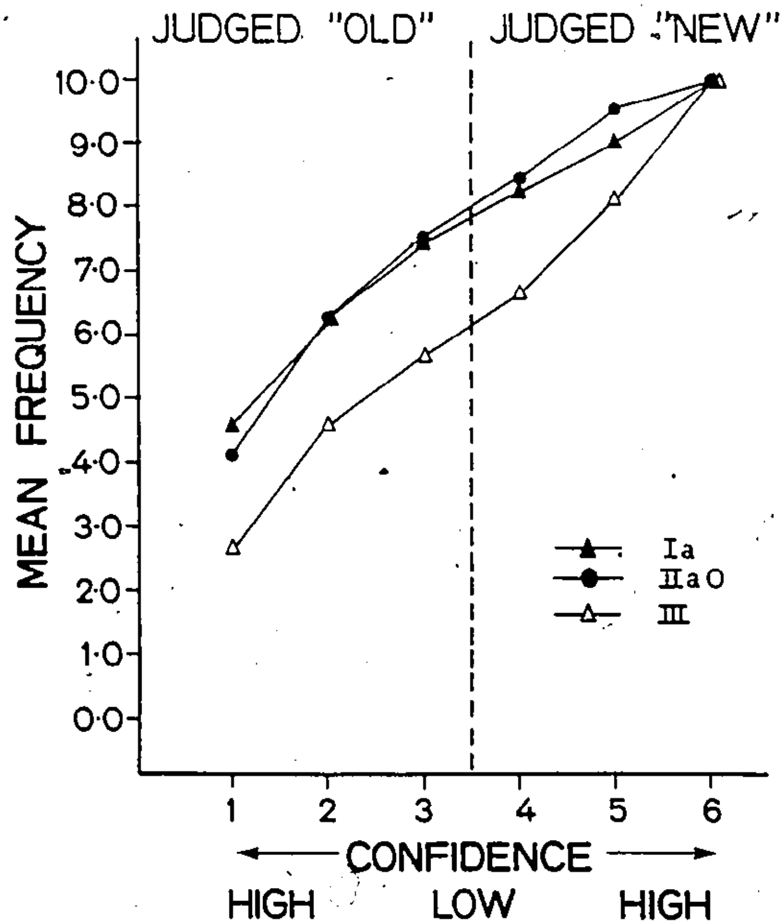


Fig. 21b. Cumulative mean frequencies of recognition-confidence judgements from Experiment 10.

TRAINING	TRANSFER
P	P
Ia Ib	$\boxed{\text{Ia}}$ Ib
$\boxed{\text{IIa}}$ $\boxed{\text{IIb}}$ $\boxed{\text{IIc}}$	$\boxed{\text{IIa}}$ IIb IIc
III	$\boxed{\text{III}}$
IV	IV
V	V

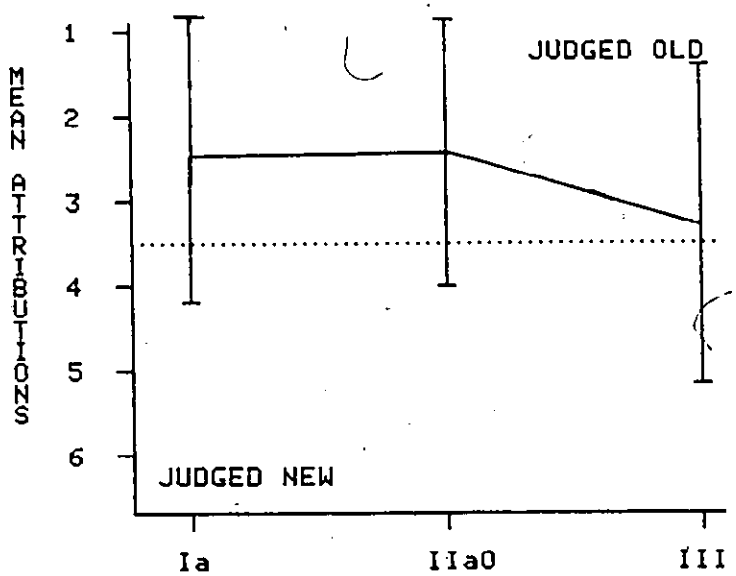


Fig. 22a. Mean composite recognition-confidence judgements for Experiment 10.

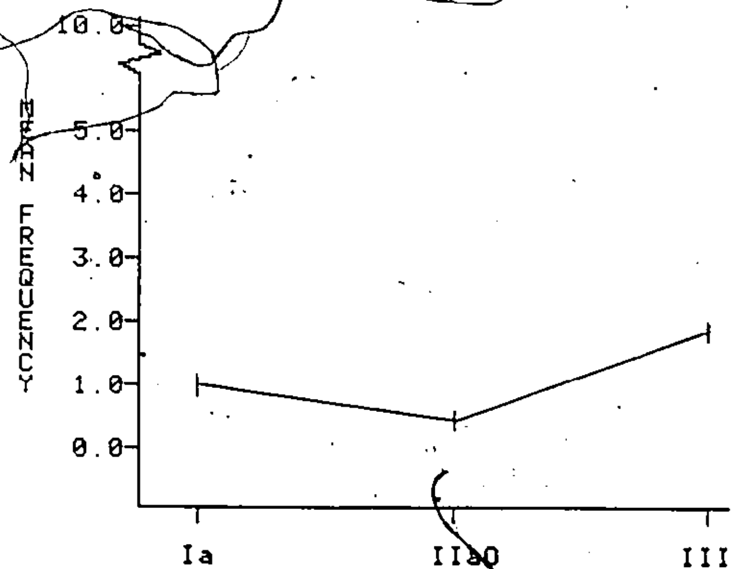


Fig. 22b. Mean frequency of judgements of high-confidence "New".

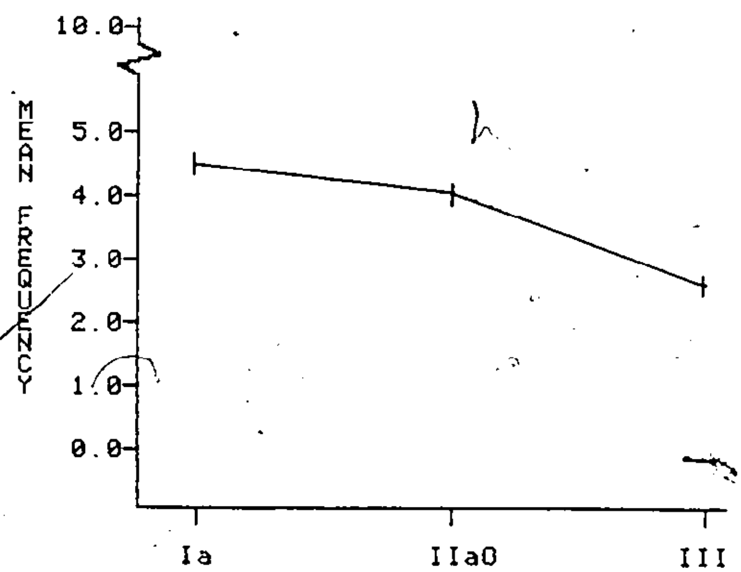


Fig. 22c. Mean frequency of judgements of high-confidence "Old".

position). Achieving this pattern of results appears to necessitate only having the training items so distributed that under an appropriate metric of similarity the items near the central tendency are more similar to the training items in general than are the training items themselves.

The contrast between these experiments and Experiment 7 (above) is particularly interesting, since they employed identical transfer stimuli and identical numbers of training stimuli, and basically differed only in terms of the density of the old items. In Experiment 7, only IIa items were used in training, repeated three times each, resulting in widely-separated concentrations of old items, whereas in the present preparation old items were more nearly uniformly distributed around the central tendency, although all were still at a distance of two deviations from the prototype. The difference in transfer patterns is remarkable ($IIa_o > Ia > III$ in Experiment 7, $Ia, IIa_o > III$ in Experiments 9 and 10), and entirely predictable under the episode model's assumptions. However, it appears to be very difficult for prototype theories to account for this difference, since the training items of these experiments have identical modes and numbers of deviations from the modes. Thus not only is an instance model sufficient to account for the "basic prototype demonstration", but it can also account for transfer patterns in spaces with different distributions of density of training items, where prototype explanations fail.

Relative to Experiment 7, Experiments 9 and 10 changed the distribution of density of old items from tight clusters of three repetitions of a string to loose clusters of similar strings. The distance between clusters was slightly reduced, while the distance between members of a cluster increased, resulting in a relatively uniform distribution of items at a distance of two deviations from the central tendency. The next study (Experiment 11) changed the distribution of density in a different way, by moving the tight clusters of Experiment 7 farther away from the central tendency, and consequently increasing the distance between clusters of old items while retaining the extreme density of members within a cluster. The result was a distribution of old items that was less dense overall than that of either Experiment 7 or 9, but was marked by concentrations of old stimuli, as in Experiment 7. It employed stimuli of type III as training items, repeated three times each, for a total of thirty training trials (as in all other experiments). The experimenter-defined categories had a resulting cue validity of .73, and represented the greatest cue validity for any way of splitting the space into categories, so that a summary abstract structure would still be thought to incorporate the central tendencies and variances of the definitional categories. The transfer set was identical to that used in Experiments 7 and 9, namely Ia, IIa and III. However, in contrast to Experiments 7 and 9, in this case the episode model, still assuming $r = 3$, predicts that the best-recognized type of stimulus will be those farthest from the central tendency: it predicts the order III_O > IIa > Ia. It can thus be seen that the episode model makes a wide variety of predictions depending on the distribution of density of the training items.

The results of Experiment 11 are shown in Fig. 23. The means of transfer types are Ia: .80, IIa: .79 and III_O: .96. These results indicate that items of type III are better perceived than IIa or Ia items ($p < .05$, one-tailed test), while IIa and Ia items are about equally well perceived. These results confirm the prediction of the model that the items farthest from the central tendency are best-recognized. The unpredicted lack of difference found between Ia

TRAINING	TRANSFER
P	P
Ia Ib	$\boxed{\text{Ia}}$ Ib
IIa IIb IIc	$\boxed{\text{IIa}}$ IIb IIc
$\boxed{\text{III}}$	$\boxed{\text{III}}$
IV	IV
V	V

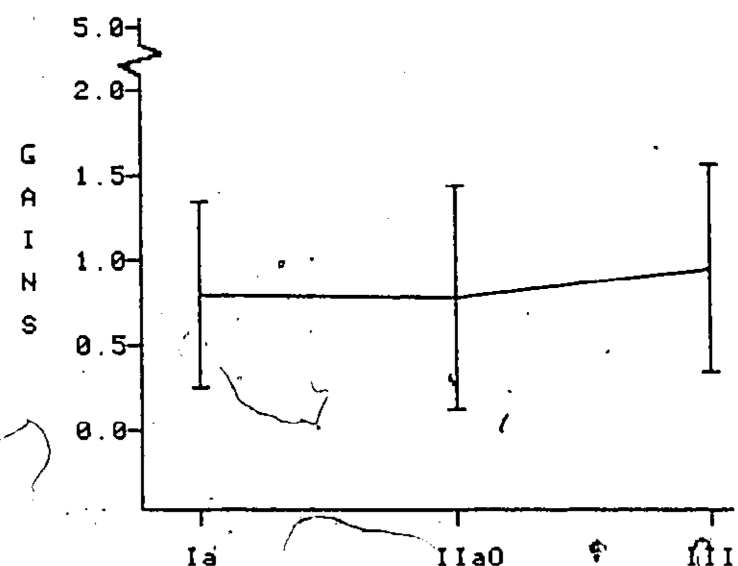


Fig. 23. Perceptual identification results for Experiment 11.

and IIa items is a minor concern: it would be predicted by a slightly lower value of r, closer to r = 2. It is at least still true that these results could not be accounted for under an assumption of independent processing of the letters, since the city-block metric predicts the order Ia > IIa > III_O, which is definitely wrong.

Taken to-gether, Experiments 7, 9 and 11 demonstrate that the episode model is capable of accounting for quite various results. Depending on the distribution of the training items, the model predicts that either old or novel items, and either items near or far from the the central tendency, will be best perceived. The model can predict results like those usually reported to support the prototype position, and also the opposite. Over the three experiments, it appears that "prototype results", the finding that novel items near the central tendency are as well perceived as old items, form a special case resting on a particular distribution of training items. Prototype theory predicts only the special case, while the episode model is more generally predictive.

Homa, Sterling and Trepel (1981) argued that exemplar-based generalization is effective only with small categories, such as those used by Brooks (1978) and Medin and Shaffer (1978), and that for larger categories, a synthesized prototype determines performance accuracy. However, the pattern of results of the last three experiments suggest an alternative explanation of Homa's findings with regard to set size. As set size increases, unless special constraints are imposed, the density of training items is likely to be greatest near the central tendency. It is also likely to be uniform around the central tendency, as in Experiment 9. (No such special constraints were imposed in the Homa et al study, which employed statistical rules to generate instances.) By contrast, small categories are likelier to have distributional characteristics more like those of Experiments 7 and 11, that is, density not maximized at the central tendency, and distributed in local concentrations. Thus set-size results like those encountered by Homa et al may be explained as due to the confounding of set size with change in distribution of density. The finding of "prototypical results" in large sets with centre-maximal density does not necessitate an appeal to the abstraction of general information.

Perhaps the most appropriate test of the relative impact of specific and general knowledge would be a pair of experiments in one of which similarity to the central tendency was held constant while similarity to instances was varied, and in the other of which the relationship was reversed. The first of these has already been described: Experiment 6 indicated a strong role in transfer performance for similarity of novel items to particular training items when similarity to the central tendency was held constant. However, in a clustered space the other experiment is impossible if one assumes that it is not similarity to the nearest, but similarity to all neighbours that is important. While variation in the distance of an item from the central tendency need not affect the distance to the nearest neighbour, as seen in Experiment 7, it does of necessity alter the similarity of that item to training items in general. The only exception is the special case in which, under a high metric, the old items are so far apart and consequently so far from the central tendency, or the level of integration of encoding is so high, that similarity of two items at different distances from the central tendency to all items but the nearest is effectively zero. Given the difficulty of achieving this special case, Experiment 12 was run instead to critically appraise the weak hybrid model.

Experiment 12 employed the IIa items in training, as most of the experiments above did, and the transfer set IIa_0 - IIc - III. This transfer set evidently pits similarity to instances against similarity to the central tendency: IIc items are two deviations from both the nearest item and also the central tendency, while III items are farther from the central tendency but closer to the nearest instance. More importantly, III items are more similar to the training items in general than are the IIc items under the assumption that $r > 2$, while IIc items are closer under $r < 2$ (see Table 10). The episode model with $r = 3$ predicts the ordering $IIa_0 > III > IIc$. By contrast a central tendency prototype model predicts IIa_0, $IIc > III$. Predictions for weaker prototype models are more difficult to make. The central tendency plus variance model also seems to predict $IIc > III$, since III items exceed the typical variance of training items, while IIc items match it. Following the spirit of the prototype tradition, the weak hybrid notion, which claims a role for both general and specific information, can be held to predict an ordering between those predicted by instances alone or general information alone: this model appears to predict that $IIc = III$. Although it is difficult to make strong claims for models with no formal processing assumptions, the author believes that no prototype model with an assumption of independence of processing between features would predict that $III > IIc$, which is the prediction of the episode model in this case.

The results are shown in the top of Fig. 24. They indicate that type IIa_0 items are best perceived ($p < .01$ in the comparison with both IIc and III), which again confirms the failure of non-hybrid prototype models. More importantly, it shows that type III items are better perceived than type IIc ($p < .05$), which is inconsistent with the predictions of any prototype model yet discussed. Moreover,

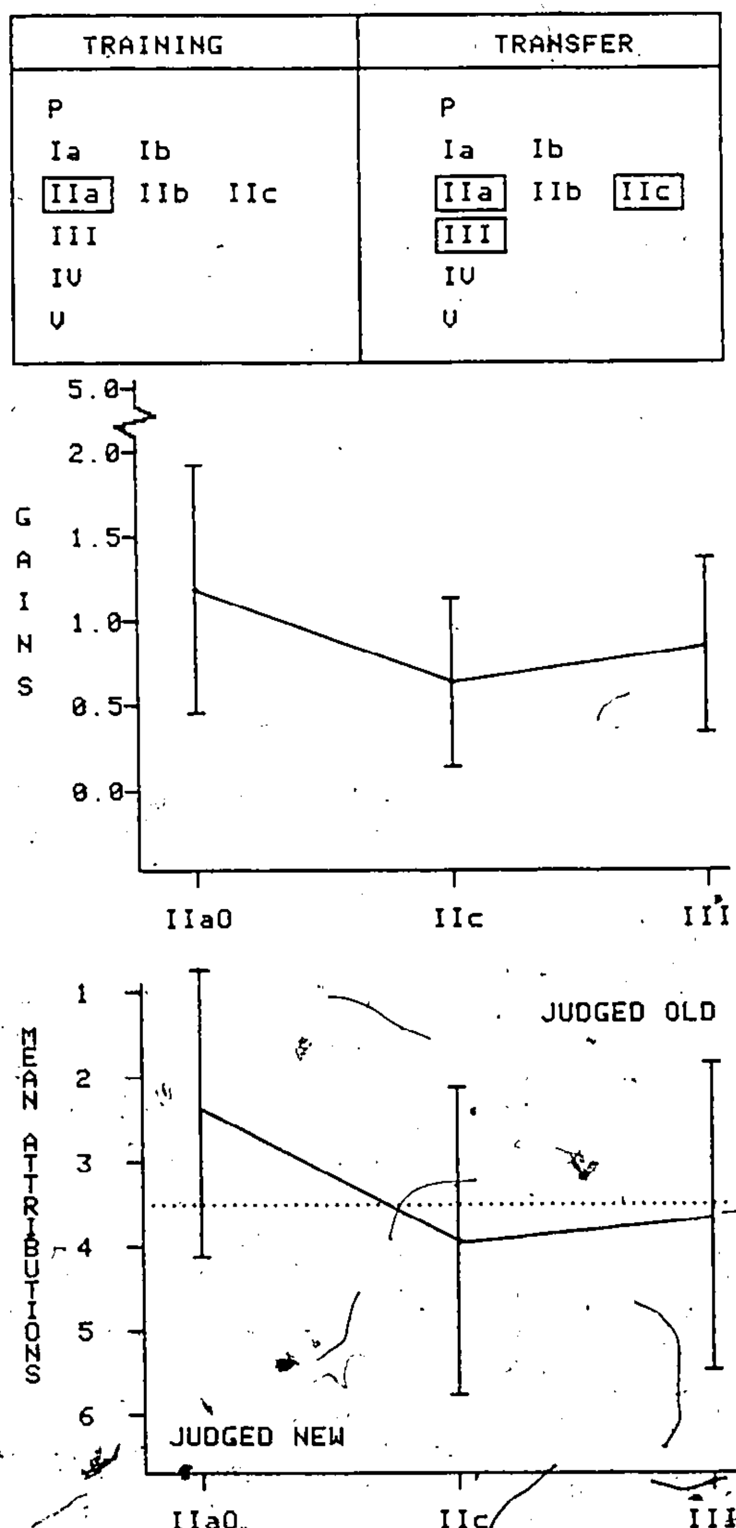


Fig. 24. Comparison of perceptual identification results of Experiment 12 (above) with recognition results of Experiment 8 (below).

there is convergent evidence that this is a reliable difference. It has already been shown in the memorial recognition experiment (Exp. 8) that with type IIa training items, the type III novel stimuli were judged more frequently and confidently to be old than the type IIc novel stimuli. Fig. 24 permits comparison of the patterns of transfer for these stimulus types in perceptual identification and recognition tasks.

The finding that novel items further from the central tendency are better perceived than novel items closer to the central tendency contradicts the fundamental observation which has supported the prototype tradition, the correlation of graded performance on probes with their distance from the central tendency. That fundamental observation appears to have rested, along with the success of prototype predictions, on the selective study of stimulus domains with centre-maximal density. The instance perspective has never had reason to insist on a correlation of performance measures with distance from the central tendency: that observation is not its motivating observation, as it is for the prototype perspective. The instance perspective can easily accomodate various distributions of density. In each of the distributions tested the episode model is able to predict performance well without appeal to general information.

Turning now to a slightly different point, the success of the episode model in Experiments 9 to 12 supports two assumptions regarding the similarity metric used to make the predictions. The first assumption was that the training task (copying the stimuli) encouraged relatively holistic encoding, such that a relatively high metric ($r > 1$) would be required to describe performance. This assumption appears to be justified by the failure of models in which $r = 1$ to account for the results of Experiments 11 and 12. This assumption was based on the second, that of commonality of process between the tasks of memorial and perceptual recognition for stimuli without differential context. This assumption appears also to be justified by the fact that the metric level derived from recognition performance (Experiment 8) effectively predicted performance in perceptual identification in Experiments 9 to 12. That commonality is thought to be the fluency of perception of the items.

V Dependent Abstraction Models

The finding of inter-feature dependence in the foregoing experiments eliminated the large classes of simple and complex independent models from consideration as explanations of the current data. The finding of differential perceptibility of novel items of differing similarity to particular training items but of equal prototypicality contradicted the strong hybrid models. Finally, the finding that novel items further from the central tendency and exceeding the standard variability of items, but closer to training items, are better perceived than those closer to the central tendency and of standard variability, but farther from training items, greatly reduced the utility and persuasiveness of the weak hybrid notion.

There remains the class of complex dependent models, which represent not only the expected value and variance of each dimension of features, but also the interdependence of features to some degree. Examples of models of this class are the higher-order feature-frequency models of Neumann (1974, 1977) and Hayes-Roth and Hayes-Roth (1977). Neumann's attribute-frequency model codes frequencies of pairs of neighbouring attributes when cues to the relevance of relationships among attributes exist (Neumann, 1974). However, since it fails to indicate what would or would not constitute a cue to the relevance of relationships, it is difficult to decide whether this model should be held to code the features of letter strings in the foregoing experiments in an independent or dependent fashion. In principle this model could be made to predict dependence in this condition, and modified to code dependencies at higher levels than the pair of neighbouring attributes. The property-set model (Hayes-Roth and Hayes-Roth, 1977) codes the frequencies of all possible n-tuples of a presented stimulus under <u>all</u> training conditions. In terms of the descriptive terminology introduced above, it in effect records information about the stimuli at all levels of integration of stimuli from $\underline{r} = 1$ to $\underline{r}$ = infinity.

Prototype models (models employing distance rather than strength) of the complex dependent class may be imagined, although they have not been promoted in the literature. The representation of the concept in memory would consist of the central tendency and variance of each dimension of features, coding modal and deviation elements for each dimension, and additionally inter-correlation information specifying the co-variation of specified feature values on different dimensions. The basis of facilitation of items would be the similarity between a probe item and the mass of stored information. This representation is identical in complexity to that employed by the strength models: storing information about the contingency of specified modal and deviation elements at all levels of multiples directly codes as much information as storing the frequencies of all the presented n-tuples. (Both types of representation are substantially more complex than the additive regression models proposed in the decision literature, discussed in Chapter 1, Section <u>II i</u>.) As indicated earlier, prototypes of this class are not identified with the "best member" of the category, in fact cannot be identified with any member of the category. If they code only the lowest level of dependency, for example only the contingencies of neighbouring pairs of elements, models of this class may still claim some of the original sense of prototypicality, in that the representations are fairly general or summary compared to the instance perspective. However, the single, central representation has been replaced by a multiple, partial representation, and the "concrete image" has been replaced by a more diffuse representation. At higher levels of dependency than the pair, such as are suggested by the foregoing experiments, the variance-covariance prototype representation becomes paramorphic of instance representation, since it encodes enough information to regenerate the instances and multiple local generalization gradients, each surrounding a particular item to which the subject has been exposed. However, it surpasses instance models in complexity of representation, since instance models suggest that people code the instances as a single trace, and do not separately encode all the various interrelationships of elements within and between instances. When forced to a high degree of feature inter-dependency, prototype representation loses all its originally-envisioned virtues of simplicity, singularity, economy and generality of representation: instead the representation becomes multiple, complex and prodigal of information precomputation and storage.

The complex dependent strength and distance models are capable in principle of explaining any transfer results, since they encode all possible information about the elements and relations of elements within and between items. For this reason, no empirical evidence will count against them. However, a number of arguments may be made concerning the preferability of instance accounts. First, the strength and distance models achieve their explanatory ability through incredibly complex and detailed representation of the stimuli; the episode model posits much more economical storage. (It is ironic that this virtue of economical representation can be claimed for instance models. Economy has been one of the virtues traditionally claimed for prototype accounts.) Second, the abstraction accounts place all the cognitive load on pre-computation. The system must update its detailed summary on each presentation of a new stimulus. Pre-computation is generally thought to possess the advantage of reducing the proportion of work done at the time a response is demanded, permitting faster responding. However, given the complexity of the summary to be updated, it appears that abstraction purchases on-line speed only at the expense of a great deal of pre-computational effort. By contrast, the instance perspective has the system expend no effort in pre-computation, the subject being thought simply to process a stimulus and retain a record of the processing. Nonetheless, it permits fast on-line processing through its assumption of parallel processing of memorial resources. Third, the major point made by instances, the impact of specific familiarity on processing, is necessarily implicit in the complex dependent abstraction accounts. In order to account for empirical findings they were forced to encode enough information, and at a sufficiently complex level, to regenerate each particular old item. However, their abstract means of representing this information adds nothing to their predictive power, but does add to their complexity of representation. Fourth, there is no empirical evidence that an abstraction account is necessary for concept formation. The episode model is sufficient to account for findings in the experiments of this paper, and also appears to be sufficient to account for the findings of the various prototype papers referenced. Again, this is not to say that people do not or cannot abstract prototypes. However, it is suggested that such abstraction will arise primarily when people are required to generate a description of an item that is highly typical of its category, and that this task is comparatively rare.

Fifth, the instance perspective is based on the notion that the memorial resources supporting categorical performance consist of what the person actually experienced, whereas the abstraction perspective suggests that it is the essence of the experience that is added into memory. For purposes of discussion we can split the event experienced into an item (the focus of the experience) and a context (the situation in which the focal event occurs). The instance position follows contemporary memory theory in suggesting that later performance will be mediated not only by prior experience with the item itself but also by the context of the item, to the extent the person processed the context in dealing with the item. That is, the instance notion expects context dependency of facilitation in many cases. This context dependency implicates retrieval processes as well as encoding processes as determinants of performance, which provides the perspective with greater fexibility, permitting it to account for failure of perception, recognition or classification when test and learning contexts differ. By contrast, the abstraction notions have no place for context effects. They contain the fundamental value that people abstract the features of things that are objectively similar or which have the same outcomes. This value implies that people abstract across situations. For example, we are thought to be able to classify an object as a bird because the mass of our experiences with birds assists the classification, irrespective of the similarity of the context at test to the contexts in which we have previously seen birds.

This insensitivity of abstraction models to context effects is one aspect of their failure to incorporate what the learner is doing on learning occasions. Another aspect is that people are thought to abstract over items within a situation, as predicted for the experiments in this paper. This abstraction is thought to be quite automatic, consisting of averaging or accumulating counts of whatever features are salient to the person. It cannot take account of changes in the way the person processes items, depending on the kind of task in which he finds himself. By contrast, the instance perspective is flexible in taking into account the processing the person does. The episode model ascribes control of the level of dependency at which the features of items are processed to the conditions of the training task, and is thus able to predict that the effective memorial resources may be at a high level of featural interdependence, consist of the features independently of each other, or even of sub-aspects of the nominal features, depending on the operations performed on the stimuli in a particular task. Thus the instance perspective is sensitive to the importance of encoding variables which the abstraction perspective cannot easily incorporate.

VI Classification Studies

Thus far this paper has examined the tasks of perceptual identification and memorial recognition, employing the dependent variables of accuracy and confidence-accuracy. This section extends the scope of the episode model to classification, the task primarily used in studies of abstraction in concept formation. The studies in this section do not test the varieties of alternative models considered above; rather, they demonstrate that the effects obtained in perceptual identification and recognition studies are also present in categorization. This parallel performance is used to argue that classification is not the only or even the best test of conceptual knowledge, but rather simply one of several tasks in which basic mental processes which may support conceptual abilities may be evidenced.

Experiment 13 employed the same stimulus structure as Experiments 1 to 4, using Ia items in training, and $\underline{Ia}_0$, Ib and V items in transfer. The two experimenter-defined classes (with cue validities of .9) were assigned class labels by arbitrarily associating underlining with one class and overlining with the other. Thirty training trials were administered, on each of which subjects were presented with one of the ten training stimuli, and required to guess its class under unlimited time constraint. Directly following their guess, the stimulus was shown with its correct class label (either over- or underlined), and the subject was required to copy the stimulus and label. Following training, subjects were shown the thirty transfer stimuli, and required to classify them without feedback, using the compound confidence-accuracy scale described above in the recognition studies (Experiments 8 and 10).

The data were collapsed across categories (across over- vs. underline), since the effects of the arbitrary categories per se were not of interest, and since there were in any case no systematic differences between them. Accuracy and confidence were analyzed separately, although gathered through a compound decision, in order to facilitate interpretation. As discussed below, these two measures appear to tap into different processes or aspects of the knowledge base in some tasks, a phenomenon which would be difficult to observe if the data were left in unseparated form and analyzed as they were in the recognition studies. It is unlikely that this separation does violence to the data since, unlike the recognition task (which can be regarded as a task of classifying stimuli into the two categories of "Old" and "New"), the categorical task consists of two classes which are completely symmetrical in experience, so that it is unlikely that any within-subject bias in category judgements will affect judgements about confidence or vice versa. That is, while in the recognition task confidence and classification are probably not independent, having a common basis in the fluency of perception, such that high confidence of itself may lead the subject to believe an item is old, in the classification task it is difficult to see how high confidence per se could lead to a preferential belief that the item belonged to one category or the other. Similarly, a bias in judging the category label is unlikely to have any differential effect on the confidence with which various types of stimuli within a category are judged in the classification task.

The predictions of the episode model for this experiment are as they were in Experiment 4: that the $\underline{Ia}_O$ items will be most accurately and confidently classified, because they are closest to what the model supposes is the memorial representation of the concept, that is, traces of the training items. For the same reason, $\underline{Ib}$ items will be intermediate, and $\underline{V}$ items most inaccurately and unconfidently classified. This prediction is made under the assumption that $\underline{r} = 3$, the value of the parameter derived from the recognition study and dragged across the perceptual identification studies. It was thought that this parameter would apply equally in the present situation because of the commonality of stimulus copying between the experiments, which has seemed to be an effective factor in causing relatively holistic processing of the items.

The resulting accuracy is shown in Fig. 25a. Accuracy means for the three transfer types were $\underline{Ia}_O$: 8.20, $\underline{Ib}$: 8.33, and $\underline{V}$: 7.10, against a maximum of

TRAINING			TRANSFER		
P			P		
$\boxed{\text{Ia}}$	Ib		$\boxed{\text{Ia}}$	$\boxed{\text{Ib}}$	
IIa	IIb	IIc	IIa	IIb	IIc
III			III		
IV			IV		
V			$\boxed{\text{V}}$		

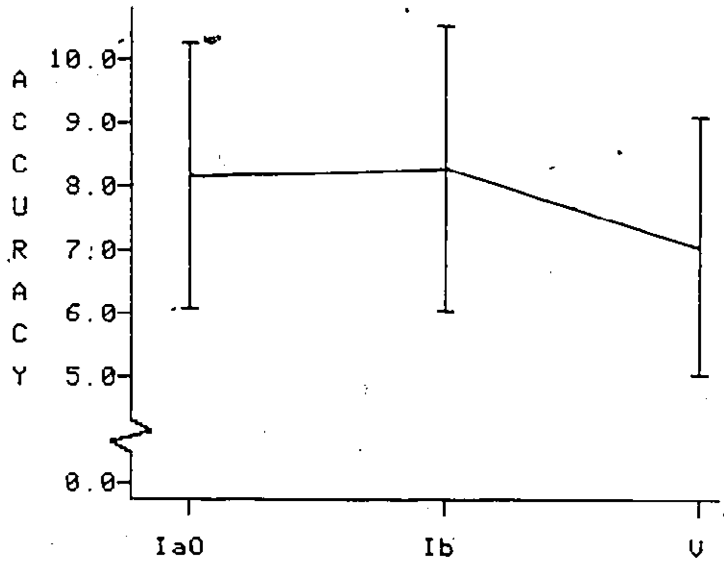


Fig. 25a. Accuracy of classification for Experiment 13.

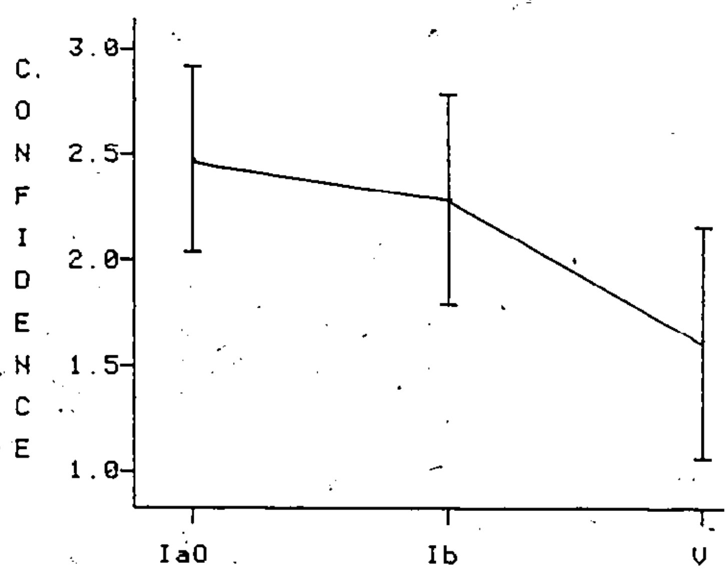


Fig. 25b. Confidence of classification for Experiment 13.

10 and a chance rate of 5 correct. The difference between $\underline{Ia}_O$ and $\underline{Ib}$ scores does not approach significance, but both are significantly greater than the mean for type $\underline{V}$ items ($p < .01$). Subjects apparently did learn something about the classification of items in this preparation: all three mean scores are significantly above chance ($p < .01$ in all cases; tested via normal approximation to binomial via Central Limit Theorem). However, this performance pattern would not be predicted by the episode model; rather these results look like those predicted by a central-tendency prototype model.

The confidences attending these judgements are shown in Fig. 25b on a 1-to-3 scale, where 3 is highest confidence, and 1 lowest. The confidence means are $\underline{Ia}_O$: 2.48, $\underline{Ib}$: 2.29, and $\underline{V}$: 1.62. All of these differences are significant ($p < .05$ for $\underline{Ia}_O - \underline{Ib}$, and $p < .01$ for $\underline{Ia}_O - \underline{V}$, and $\underline{Ib} - \underline{V}$). Unlike the accuracy results, these confidences are in line with the prediction of the episode model, and not with the predictions of prototype models.

To examine the generality of these results, another study (Experiment 14) employed the $\underline{IIa}$ stimuli in training (with resultant cue validites of .8 for each category), and the set $\underline{IIa}_O$, $\underline{IIc}$ and $\underline{V}$ in transfer. This is the same stimulus structure that was employed in Experiment 5, and the episode model makes the same prediction, that $\underline{IIa}_O$ stimuli will be most accurately and confidently classified, followed by $\underline{IIc}$ and then by $\underline{V}$ stimuli. Accuracy results are shown in Fig. 26a. The means for the transfer types are $\underline{IIa}_O$: 5.87, $\underline{IIc}$: 6.14 and $\underline{V}$: 5.5, again against a maximum of 10 and a chance rate of 5 correct. All three means are significantly above chance performance ($p < .01$, .01 and .05 respectively), although subjects appear to have learned very little; and none of the differences between the accuracy means is significant. Again these results do not accord with

TRAINING	TRANSFER
P	P
Ia Ib	Ia Ib
$\boxed{\text{IIa}}$ IIb IIc	$\boxed{\text{IIa}}$ IIb $\boxed{\text{IIc}}$
III	III
IV	IV
V	$\boxed{\text{V}}$

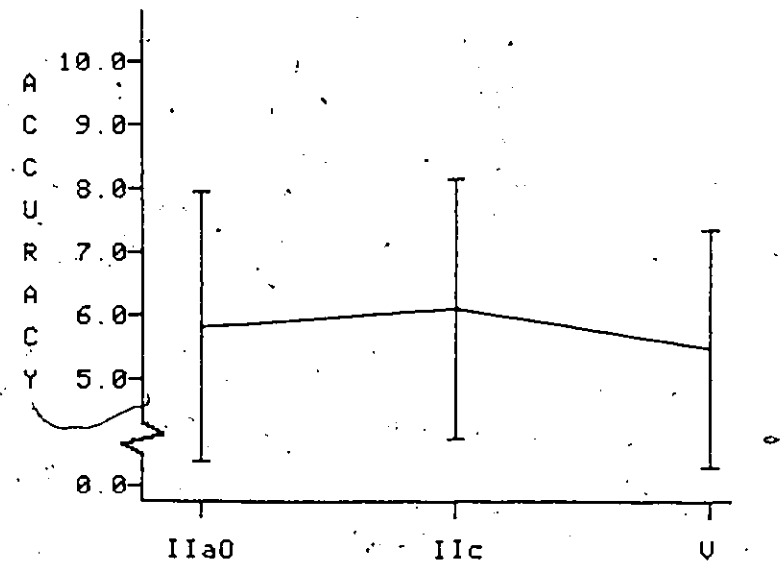


Fig. 26a. Accuracy of classification for Experiment 14.

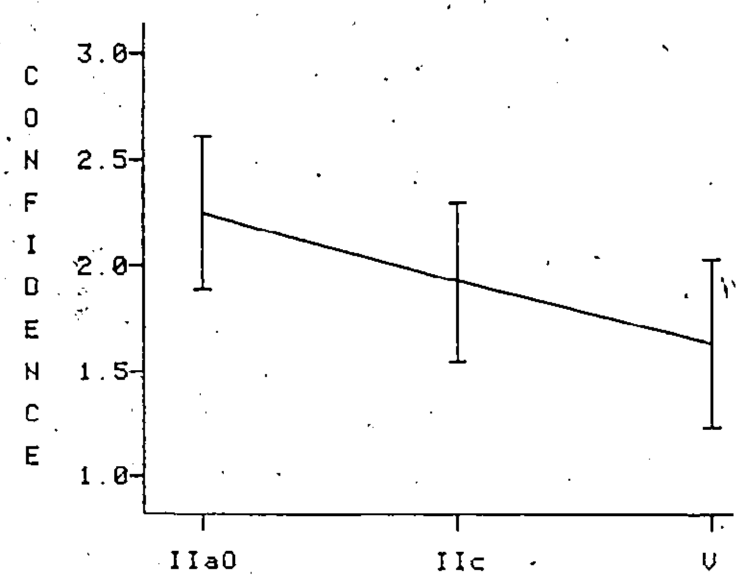


Fig. 26b. Confidence of classification for Experiment 14.

the episode model's prediction. However, the pattern is different in the confidence results, shown in Fig. 26b. The confidence means are $\underline{IIa_0}$: 2.26, IIc: 1.93 and V: 1.64 (all differences being significant, $p < .01$ in all cases): these results are predicted by the episode model. Thus in two studies the accuracy of categorization fails to follow the order predicted by the episode model, while the confidence of categorization follows precisely the predicted ordering.

Interpretation of these results is tortuous. Beginning with the confidence results of Experiment 14, it appears that performance is mediated by the similarity of transfer items to particular old items, as indicated by the episode model. However, the accuracy results of this experiment do not concur. It is as though, when confronted with an old item, subjects have no idea what category it belongs to, but feel that it's a very good member of whichever category. Prototype notions do not help solve this puzzle. They might account for the low accuracy on all transfer types as due to the difficulty of abstracting a prototype from a set of items each deviating from the central tendency by two features, but this argument cuts against the generality of prototype abstraction. Further, independent pure-abstraction models (as discussed above) would not predict the confidence results of Experiment 14, the finding that old stimuli are more confidently classified than novel but equally prototypical stimuli. The same difficulties are true of the interpretation of the two patterns of judgement in Experiment 13: the episode model would not appear to predict the accuracy findings, and prototype models would not appear to predict the confidence findings.

The most promising explanation of these findings is that subjects employed two different procedures to perform the two judgements. Being in a trial-and-error learning situation, subjects might attempt to analyze the commonalities of stimuli in each category. This strategy would be fairly easy and successful in Experiment 13, since in the training set each category-relevant feature occurs in 80% of the stimuli of a category, and is completely predictive of its category when it occurs. Thus a single extremely simple attributive rule like "Category I items have an F; Category II otherwise" would by itself lead to 90% accuracy on all transfer items. However, a more specific rule like "Category I items have an F in the first position; Category II otherwise" would lead to 90% accuracy on Ia_O and Ib items, but only 50% accuracy on V items, which were created by reversing the Ia_O items. In general, any set of one-feature rules will lead to equal accuracy on Ia_O and Ib items in transfer, while any position-specificity in the rules will lead to poorer performance on the V items. This appears to match the performance pattern in Experiment 13. Moreover, when questioned after the experiment, most subjects indicated that they had been attempting to isolate predictive cues, and many of their cue hypotheses included the position of the cue.

This strategy would work less well on the stimuli of Experiment 14. The training stimuli of this preparation each possess only three category-discriminative features, with the result that any relevant cue occurs only 60% of the time with a stimulus of its category. The simple attributive rule mentioned above would thus have an 80% success rate, lower than it had in the previous experiment. Perhaps more important, a particular relevant cue occurs in only 30% of training trials. If this is coupled with a tendency not to treat the absence of a cue as informative, the utility of particular cues is low, which likely leads to subjects changing hypotheses rapidly. The predictable result is that accuracy of classification of all three transfer types will suffer, leading perhaps to the transfer pattern shown for Experiment 14. Subjects in this preparation also indicated that they had attempted to isolate particular predictive cues.

It seems likely, then, that subjects in Experiments 13 and 14 were using a predominantly analytic strategy to learn to categorize the stimuli. This would presumably lead to a memorial representation of category knowledge in terms of a collection of cue-based heuristics. Such a knowledge base would predict the patterns of accuracy in transfer for the two experiments, but would not predict the patterns of confidence. Recalling from Experiment 5 that $\underline{\text{IIa}}_{\text{O}}$ and $\underline{\text{IIc}}$ items differ only in terms of the association of particular irrelevant cues with particular relevant cues, if the application of analytic heuristics is to account for the difference in confidence between these types, then the heuristics must consist of high-level compounds of relevant and irrelevant features. That is, to account for the confidence data, there would have to be as many heuristics as instances, and each heuristic would have to specify all the cues of a particular instance. However, if they did so, the accuracy of $\underline{\text{IIa}}_{\text{O}}$ items would exceed that of $\underline{\text{IIc}}$ items, which is untrue. In summary, an analytic knowledge base consisting of relatively low-level compounds (or more probably single cues) is the likely explanation of the accuracy data of these two experiments, but does not account for the confidence data.

That data may be accounted for by recalling that, as indicated above, subjects were not only required to guess the category of items in training, but also (following feedback on their guess) required to copy the item and its class label. The expectation regarding this requirement was that subjects would encode each item in a relatively holistic fashion, as argued in the previous experiments, and additionally encode its class label as part of the same trace. The expected result was that later classification of items would be done by analogy to traces in memory, resulting in most accurate and confident performance on old items. It appears that the first part of this supposition was incorrect, that the similarity of an item to items previously processed has little or no impact on classification accuracy: subjects appear to judge the class of an item purely in terms of their analytic heuristics. However, the second part of the supposition appears well justified. Of the three possible knowledge bases considered, prototypes, heuristics, and encoded episodes, only the encoding of episodes can account for the confidence patterns found.

Thus it appears that subjects attained a mixed knowledge base consisting of both low-level heuristics and also high-level traces of particular items, and used these two types of information differentially for the two aspects of the transfer task. In the terminology of Osherson and Smith (1981) and Miller and Johnson-Laird (1976) subjects employ different identification procedures for their judgements of classification and confidence. For judgements of confidence, they apparently employ the fluency of perception of the probe (argued above to be determined by the availability of specifically similar memorial resources). This information is not criterial to the confidence decision, as the degree to which the item fits known rules of category membership might be, but it is relatively fast and easy. (It is also inaccurate in this case, but see the following experiment.) The use of perceptual fluency to make judgements of confidence parallels its use in perceptual identification and recognition tasks, for which it is the <u>only</u> basis of decision available (in the absence of discriminating contexts). However, the classification training task offers subjects the opportunity to learn about the core of the concept, the rules constraining category membership. They appear to attempt to apply this core information in their categorical judgements, employing analytic heuristics as their identification procedure. This procedure is likely to be slow and difficult to apply relative to the use of perceptual fluency, but guarantees accuracy if the analysis of the core is successful.

The question of why encoded episodes had no effect on classification accuracy, contrary to prediction, now requires an answer. The explanation to be offered hinges on the notion of the integration of the class label with the trace of an item. The predictions of the episode model for accuracy in these two experiments depended on the assumption that the experience of copying each stimulus with its class label, either an underline or an overline, would result in a single, integral trace. However, if the class label is not integrated with the trace of the item, perhaps because subjects lack prior structure for the integration of over- and underlines with letter strings, then experiences of items and labels will be separately encoded. In this case a transfer item would benefit from familiarity only of the item identity. Greater familiarity of item identity might be supposed to lead to greater confidence, but could not support class identification.

The reasons why over- and underlines would be minimally integrated with letter strings will not be explored here in any depth. Pilot experiments suggested the same lack of integration when the class label consisted of the physical inflection of letter strings (strings arched up or arched down). This phenomenon appears to be related to findings that change of case of letters has at best a small impact on later perceptual identification (Jacoby, personal communication; Feustal, Shiffrin and Salasoo, in press). Such considerations bear more on the issue of how best to think of the trace of an episode than on the issue of whether particular episodes exert influence on later performance.

The next experiment (Experiment 15) attempted to assess the validity of the mixed-knowledge-base argument. While the stimulus structure in training and test was maintained unchanged from Experiment 14, the training task and category labels were altered. Whatever the basis for ease of integration of components of an experience, it was already apparent from the foregoing experiments that letters are easily integrated with other letters to form processing wholes. Experiment 15 capitalized on this, employing letter suffixes as the class characteristic. The training phase of this experiment consisted of an errorless learning task. As each training stimulus was presented, it was accompanied by a category label, either "NOUN" or "VERB". Subjects were instructed that when a stimulus was accompanied by "NOUN" they were to copy the stimulus, adding the suffix "ISM"; for "VERB" they were to add "ING".

The instructions of Experiment 14 (above) stressed that the subject's task was to develop the ability to classify: each training trial was also a test trial. Learning was by trial and error: the subject could only discover the correct classification of an item by first hazarding a guess. The category labels he was trying to associate with items were arbitrary, and are considered (in retrospect of Experiments 13 and 14) to be difficult to integrate with items. These appear to be excellent conditions in which to expect a hypothesis-testing strategy (albeit with dubious success in only thirty trials). By contrast, in Experiment 15, the ability to classify was not stressed. Rather the subject was required to perform a simple generation task which did not test his categorical knowledge. Learning, if it occurs at all in this preparation, need not proceed through correcting erroneous guesses, since the answer (the class label) is given along with the question. The category labels are associated with class characteristics for which the learner has prior structure (nouns in English <u>do</u> not uncommonly end in "ism", and verbs in "ing"), and the class characteristics are composed of the same type of elements as are the stimuli, namely letters. Thus the association of class labels and characteristics with stimuli is much less arbitrary than in the previous studies. It appears intuitively likely that the temptation to exercise an extensive hypothesis-testing strategy will not be strong in this preparation. In fact very few subjects in this experiment reported even having thought of testing cues for differential predictiveness.

Nonetheless subjects were predicted to evidence quite a lot of learning in this preparation, because they were required to copy the stimuli and class characteristics on each trial. The instance perspective suggests that because the subject has been put through these experiences, and because the class characteristic is probably fairly easy to integrate with the stimulus, test presentation of a stimulus will lead to accessing a similar memorial representation, complete with class characteristic, so that the subject will find one suffix more fluent than the other. He can use this differential fluency to perform the demanded classification (although he might, if asked, be hesitant to justify his judgement on this basis). It is this same fluency that the subject is thought to use to generate his confidence judgements. Thus the episode model predicts that both accuracy and confidence will co-vary with the similarity of transfer stimuli to the set of training stimuli. Specifically, it predicts the ordering $\underline{\mathrm{IIa}}_{\mathrm{O}} > \underline{\mathrm{IIc}} > \underline{\mathrm{V}}$ for both accuracy and confidence of classification.

The accuracy results are shown in Fig. 27a. The means of the transfer types are $\underline{\mathrm{IIa}}_{\mathrm{O}}$: 7.73, $\underline{\mathrm{IIc}}$: 6.83, and $\underline{\mathrm{V}}$: 5.03. Both $\underline{\mathrm{IIa}}_{\mathrm{O}}$ and $\underline{\mathrm{IIc}}$ types are significantly above the chance rate of performance ($p < .01$) while type $\underline{\mathrm{V}}$ is not. Most importantly, the differences among types are all significant ($p < .05$

TRAINING	TRANSFER
P	P
Ia Ib	Ia Ib
IIa IIb IIc	IIa IIb IIc
III	III
IV	IV
V	V

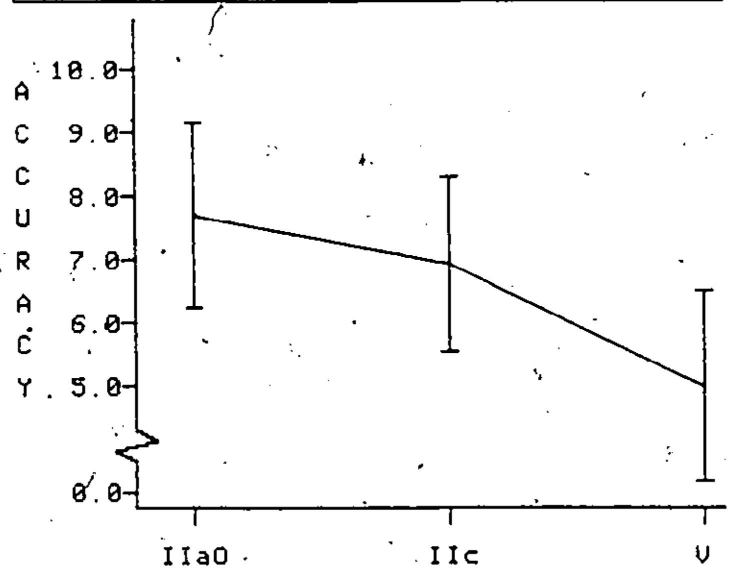


Fig. 27a. Accuracy of classification for Experiment 15.

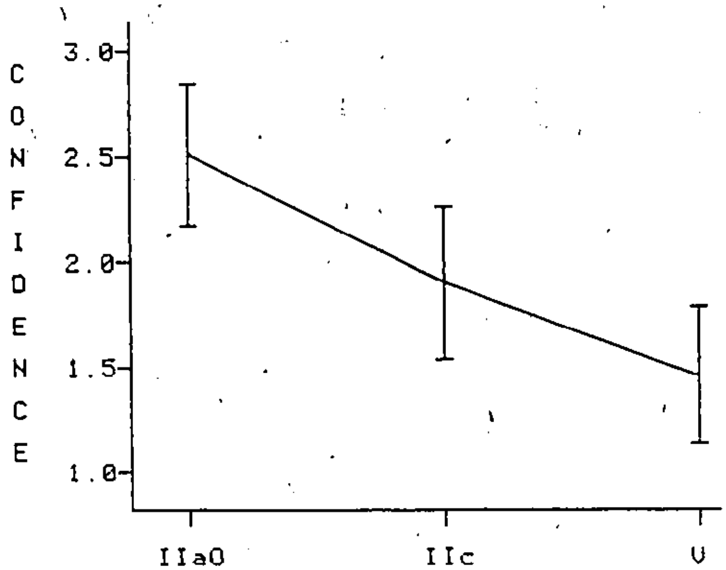


Fig. 27b. Confidence of classification for Experiment 15.

for $\underline{IIa}_O$ > $\underline{IIc}$, $p < .01$ for $\underline{IIa}_O$ and $\underline{IIc}$ > $\underline{V}$), and are in line with the episode model's predictions. The predictive success of the model is repeated in the confidence data, shown in Fig. 27b. The means of the stimulus types in confidence are $\underline{IIa}_O$: 2.52, $\underline{IIc}$: 1.9 and $\underline{V}$: 1.47 (all differences significant, $p < .01$). Neither a pure-form prototype model nor a low-order compound analytic model would predict the findings that $\underline{IIa}_O$ items are more accurately and/or confidently judged than $\underline{IIc}$ items.

Reflecting back on the assumptions made for these predictions, the training task in this experiment does appear to have avoided setting an analytic strategy, at least to the extent of permitting non-analytic processing to be evidenced. Secondly, the change in class label to one thought to be more easy to integrate with the stimuli also appears to have been effective. This was made particularly plain by a series of pilot studies (whose data is not reported) run attempting to find results like those in Experiment 15. Two of these studies employed an errorless training task, and were in fact identical to Experiment 15 save in using under- vs. overline or stimuli arching up vs. down as the category labels. The results of these experiments showed very little evidence of categorical knowledge, no difference in accuracy between the $\underline{IIa}_O$ and $\underline{IIc}$ items, but, as usual, the ordering $\underline{IIa}_O$ > $\underline{IIc}$ > $\underline{V}$ in confidence. Thus the manipulation of training task from trial-and error to errorless does not in itself seem sufficient to account for the difference.

The question remains whether trial-and error learning, as in Experiment 14, coupled with category labels easy to integrate with stimuli, as in Experiment 15, produces accurate performance. Pilot studies have been run employing parametric manipulation of trial-and-error vs. errorless learning, integral vs. non-integral category labels and string copying required vs. not required. Pilot data to date suggest that the trial-and-error learning task has the main effect of setting an analytic strategy which tends to overwhelm either the appearance or the actual use of holistically-coded instances in classifying items. However, this learning strategy is in general not particularly successful, even with integral category labels, unless subjects are additionally required to copy the strings coupled with their suffixes; and in this case there are marginal indications of an impact of specific familiarity on classification. The only condition in this series of studies which is solidly consistent with classification on the basis of encoded instances is the condition employed in Experiment 15 (although concurrent work by Brooks (personal communication) also demonstrates classification on this basis). Nevertheless, all conditions show the effect of similarity to particular old instances in the confidence of classification. Investigation is continuing on the conditions which differentially set analytic and non-analytic strategies, and the implications of the use of mixed knowledge bases for performance accuracy.

It must be mentioned that, for the sake of completeness, a study was run which was like Experiment 15 in all but the stimulus structure. Experiment 16 used the $\underline{Ia}_O$ - $\underline{Ib}$ - $\underline{V}$ stimuli, which made it parallel to Experiments 4 and 13 in structure. The prediction $\underline{Ia}_O > \underline{Ib} > \underline{V}$ was made for this study, but without great hopes, since the training stimuli within a category overlap to such an obvious extent that analysis of common components is very easy. Indeed subjects reported using such a strategy much more frequently in this study than in the previous one. The results are shown in Figs. 28a and b. The accuracy means are $\underline{Ia}_O$: 8.43, $\underline{Ib}$: 8.53 and $\underline{V}$: 5.7. All three means are significantly above chance performance (p <

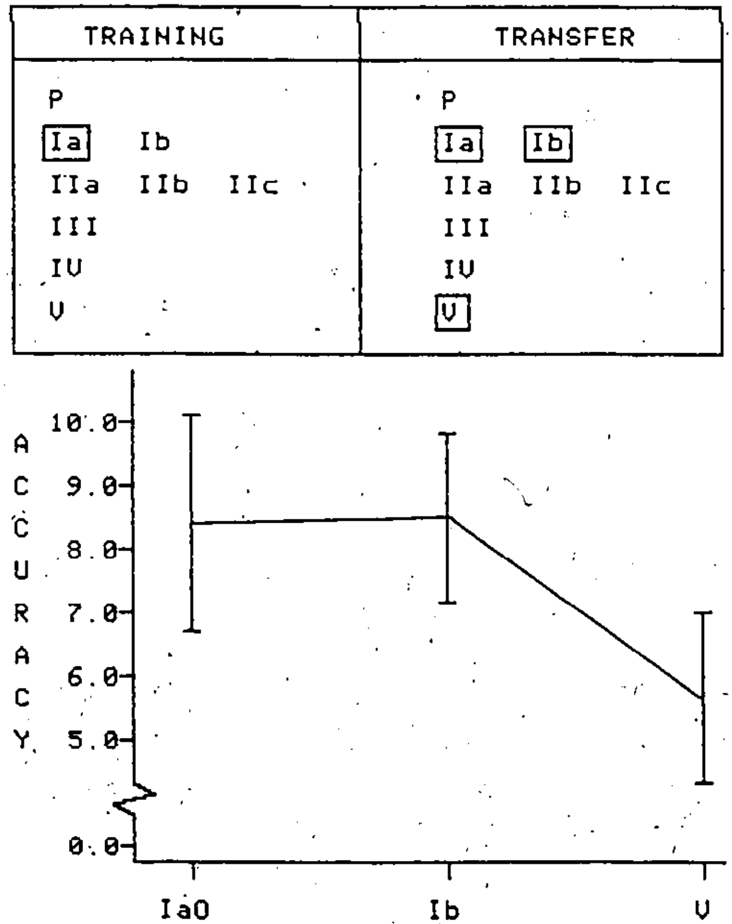


Fig. 28a. Accuracy of classification for Experiment 16.

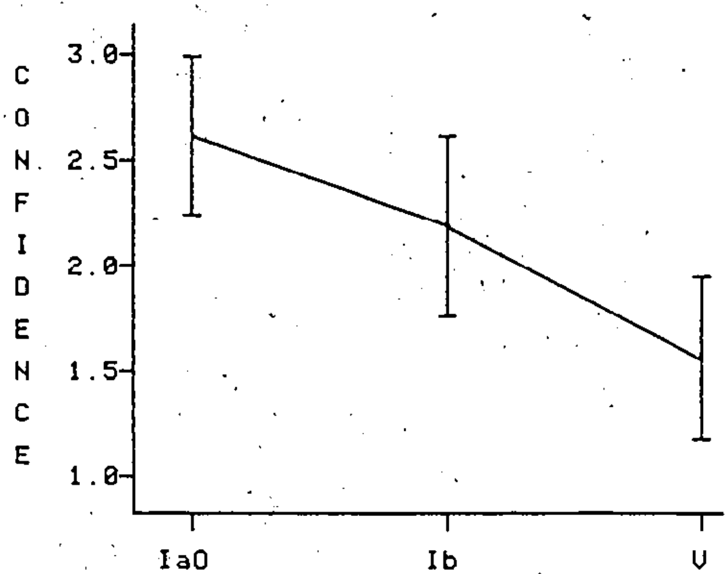


Fig. 28b. Confidence of classification for Experiment 16.

.01), and both $\underline{Ia}_{O}$ and $\underline{Ib}$ means are greater than the mean of type $\underline{V}$. However, the difference between $\underline{Ia}_{O}$ and $\underline{Ib}$ does not approach significant levels. The confidence means are $\underline{Ia}_{O}$: 2.62, $\underline{Ib}$: 2.19 and $\underline{V}$: 1.57, in accordance with the episode model's predictions (all means differ with $p < .01$). The accuracy results are complicated by an apparent ceiling effect for $\underline{Ia}_{O}$ and $\underline{Ib}$ items: almost a third of subjects achieved a perfect score on each type. Secondly, the fact that many subjects engaged in at least some analysis makes the results difficult to interpret. As mentioned above, the effect of the use of low-order compound heuristics in the transfer task is to eliminate performance differences between old and similar but novel items. This study serves more as a footnote on the readiness of subjects to engage in analysis in tasks which are clearly categorical than as a test of the hypothesis of interest. The confidence results, at least, still show the effects of specific prior familiarity.

Several general conclusions may be drawn from these results. First, prototypes do not appear to provide either a necessary or sufficient account of the data, while mixed analytic-plus-instances or pure-instances explanations do provide a sufficient account for particular training conditions. Secondly, Experiment 15 demonstrated that, under appropriate conditions, the specific familiarity effects obtained in perceptual identification and memorial recognition are paralleled in the accuracy of classification, while all the studies in this section demonstrated specific familiarity effects in confidence of classification parallel to those in perceptual identification and memorial recognition.

It is also apparent that classification is a particularly complex task from which to draw inferences about the conceptual knowledge base. Classification performance is evidently susceptible to the influence of a variety of factors to which little attention has been paid in the concept formation literature. First of all, depending on the task subjects are set, a variety of strategies may be employed in attaining the concept. Thus trial-and-error learning appears to bias subjects toward active analysis of stimuli, while errorless generation appears to tempt subjects much less to perform active analysis. Secondly, conclusions about the strategy employed and about the underlying cognitive structure may depend on the dependent measure of classification selected by the experimenter. In most of the experiments above, the confidence data presents a different picture of the underlying processes than does the accuracy data. The nature of categorical learning supporting classification also appears to be influenced by the ease with which category labels can be integrated with stimuli, which in turn appears to hinge upon the existence of prior structure for associating the two. The intentionality of category learning may also influence the development of a conceptual base. When category learning is incidental, as in the perceptual identification and memorial recognition studies, non-analytic strategies may be employed by default, while the intention to learn to classify may press toward active analysis. Strategic factors like these, and task requirements like spelling vs. copying, may have great effect upon determining the characteristics of the knowledge base available for later classification.

The intentionality of classification is also likely an important determinant of those aspects of the knowledge base accessed for performing classification. Thus if subjects are aware that their ability to classify is being tested, different aspects of the knowledge base may be accessed than if the classification task consists of incidental labelling (e.g. patting dogs and tickling cats under the task instruction "Play with these animals"). The requirement to justify class judgements is also likely to affect which aspects of the knowledge base are employed, by biasing subjects toward a "rational", verbalizable basis of classification such as analytic rules, while the demand to be generally right without justification of decision may be served by a non-analytic basis of classification. The general conclusion is that classification is an enormously complex task in terms of factors which could affect the development or later employment of a particular type of knowledge base. General statements about how people perform in classification tasks are at this point somewhat premature. The most useful approach to the classification task appears not to be demonstrations that people can in principle learn to classify in a particular fashion, but rather to conduct task analyses of particularly common or important occasions on which people are required to classify, and attempt to determine the basis of classification in those situations.

Another aspect of the studies in this section is that they begin to approach the problem of artificiality of the stimuli employed in these experiments. This problem is two-edged. Artificial materials were used in the first place because the specific familiarity of natural materials cannot be controlled or manipulated as precisely as that of artificial materials can be, since the learning history of the subject for natural materials is unknown. However, generalization of principles from artificial to natural materials is a risky undertaking, since one is never sure that the artificial materials have captured the essential characteristics of the natural materials. Experiment 15 introduced a first step toward combining these concerns. The demonstration of instance-guided performance in that study depended critically on the employment of a relationship between stimuli and categories that was non-arbitrary in the subject's history.

Specifically, it depended on the subject's prior structure for integrating suffixes with word stems, and for recognizing the syntactic class to which a suffix belongs. While only a bare beginning, these data point to an instance-based interpretation of a non-artificial phenomenon, namely people's ability to deal with lexical categories.

Experiment 15 also provided some clues about why many prototype experiments have required very many trials before subjects achieve accurate classification. (Subjects in an experiment by Homa, Sterling and Trepel (1981), for example, required an average of 17 repetitions of the entire training set to reach an errorless criterion.) Studies in prototype research have generally employed stimuli which are only arbitrarily associated with category names, for example the "A", "B" and "C" categories of dot patterns. The findings above suggest that these concepts would be difficult to learn in any fashion, the difficulty being the learning of the association between stimulus characteristics and category labels. Integral labels would likely greatly reduce the amount of exposure to stimuli required to perform accurately in these preparations. Further, selection of a different dependent variable, one which does not depend on producing a class label (as perceptual identification and memorial recognition do not), would probably demonstrate that significant amounts of conceptual learning occur with much less exposure to the stimulus set. The suggestion here is that learning the association of class labels with items is the block to the overt demonstration of a conceptual knowledge base. The argument that concept learning should not be identified with classification ability is enlarged upon below, in the Discussion.

There are two other possible explanations of why subjects require so many trials' experience with the stimuli in the prototype studies. Both explanations depend on the fact that these studies generally impose trial-and-error learning. The first explanation is that subjects attempt to employ an analytic strategy to attain the concepts, but fail to gain much from the attempt because the rules governing category membership are too complex. This explanation is of course suggested by the fact that trial-and-error learning was employed, but is not very tempting: there is little indication that subjects do attempt to analyze dot patterns. A more appealing explanation focusses on the fact that learning in trial-and-error conditions involves making errors. If these errors are encoded, and feedback is inadequate in erasing, correcting or competing with the erroneous trace, subjects will have some tendency to perseverate in an error once it is made. This of course is an instance explanation, since abstraction notions would suggest that information about particular episodes, whether right or wrong, should not affect performance on specifically similar later occasions. The possibility of such perseveration is important because of its relevance to pedagogical issues such as optimal learning, but is as yet unexplored.

CHAPTER 3

Discussion

I Summary of the Experimental Results

The intention of the experimental sections of this paper was to deny the necessity of abstraction explanations of concept formation while affirming the sufficiency of instance explanations. This intention has been fulfilled to a fair extent: a variety of cases were shown in which conventional prototype models could not account for performance, but the episode model could. In view of Section VI of Chapter 2, which demonstrated that prototype models could be modified to account for these results, the charge of insufficiency is not leveled at prototype models. More importantly, however, an instance model was shown to be capable of accounting for not only the performance patterns which have long motivated prototype theorizing, but also capable of predicting importantly different results. The sufficiency of an instance model to handle these data denies the necessity of an appeal to an abstraction explanation of conceptual performance, an argument which has been a mainstay of the abstraction position.

To recapitulate more specifically the argument pursued empirically in this paper, the early studies (4 and 5) demonstrated processing interdependency of the features of stimuli, eliminating all models which stipulate independent storage of features as explanations of performance under the task conditions. It was argued that most traditional prototype notions fall into this class. Experiment 6 (with all transfer items of equal prototypicality) demonstrated that similarity to old items predicts differential performance on novel items, which eliminated strong hybrid explanations of the data. Further, by demonstrating processing dependency between typical and rare aspects of stimuli, it eliminated traditional analog prototype representation as a contender, since such representation does not incorporate this dependency. Experiment 7 (using transfer items decreasing in prototypicality) showed that nearest-neighbour instance models also cannot account for subjects' performance, since items differing in prototypicality but of equal similarity to old items were found to be handled differently. However, this experiment, coupled with Experiment 6, began the task of tracing out the distribution of facilitation density through the stimulus space. This density obviously did not correspond to local independent domains of influence of particular old items, but was thought to be possibly due to the interactions of such local domains. Experiment 8 generalized the previous findings by demonstrating that the same facilitation density distribution was obtained using a recognition task. It thus provided a check on the reliability of the pattern of facilitation, as well as broadening the scope of the investigation to a second important dependent measure.

A model employing a parameter specifying the level of integration at which items are processed (the episode model) was developed and tested on this data. It was found to account for the obtained order of transfer well when the level of integration specified by the parameter was relatively holistic. This model was used to predict the results of four more studies (Experiments 9 - 12), and was found to be quite successful, in spite of the wide range of the patterns of results of these studies. Experiment 9 (using distributed training items) showed novel items nearer to the prototype to be as well handled as old items; Experiment 10 replicated these results, employing recognition as the dependent variable; Experiment 11 demonstrated a casê in which items furthest from the prototype are best handled; and Experiment 12 pitted similarity to the prototype against similarity to particular instances, and showed the latter to be the major determinant of performance. To-gether, these four studies were argued to eliminate even the weak hybrid notion (the notion that similarity to the prototype has an impact on performance independent of similarity to all the particular instances separately) as an explanation of the data. They also demonstrated the great flexibility of instance models in predicting widely various results.

Finally, a series of classification studies (Experiments 13 - 16) were conducted to demonstrate performance in classification tasks parallel to that in perceptual identification and memorial recognition, with the intention of demonstrating that the basis of concepts deduced from the latter tasks, the encoding of instances, is also the effective basis of concepts under at least some conditions of the classification task. The results of these studies were argued to give no support to the prototype position, but instead to suggest that instance knowledge may be frequently the basis of categorical confidence and, under appropriate training conditions, comprise the functional knowledge base accessed for the classification decision. Analytic strategies were also observed in these studies; however, it was suggested that in many cases concept formation and/or classification may proceed without the learner's awareness, in which case the analytic strategy is not expected to arise, leaving the encoding of instances as the predominant method of acquisition of a conceptual knowledge base, and analogy to encoded instances as the predominant basis of classification.

The empirical work of this paper demonstrated that traditional prototype models form an inadequate account of concept formation with materials structured to conform to the orienting assumptions of clustered distribution of events and objects in the world. The only abstractionist models which could account for the data assembled in this paper are those of the complex dependent class, including higher-order feature-frequency models and unconventional complex dependent prototype models. Section VI of the last chapter argued that the instances perspective possesses a number of virtues that the abstraction accounts do not, including economy of representation and pre-processing, and the ability to mesh smoothly with contemporary memory work in context dependency and encoding specificity. Further arguments concerning the preferability of an instance account are presented below, in part V of this section.

An important aspect of the studies of this paper is that they are based on multiple types of tasks. Concept formation has been mainly identified with the task of classification by prototype theorists, as noted above. This identification has perpetuated the split between concept formation and memory, accentuating task differences between the two traditions rather than emphasizing common processes in the mental functions underlying the two tasks. In contrast, the experiments undertaken in this paper, employing perceptual, memorial and classification tasks, form part of an attempt to seek basic mental processes which may subserve a variety of functions. Beyond the theoretical integration which may be gained by such an attempt, the methodology of the former tasks may be a very useful tool for the investigation of concept formation. As indicated in the last section, classification is a very complex task; while interesting in its own right as a task that people are called upon to execute, it is a difficult situation in which to investigate conceptual knowledge bases, in part because it is difficult to determine how the subject has processed the experimenter-defined features. By contrast, as illustrated by the experiments in this paper, perceptual identification permits precise, feature-by-feature analysis of productions, and hence permits testing of fairly fine-grained hypotheses.

II Status of the Episode Model

The episode model described in this paper has proven to be a good predictor in the domain in which it has been tested. However, this domain is rather limited. The model has only been tested in situations which were thought to be conducive to encoding items at a relatively high level of featural interdependency, and with artificial materials. An obvious next step is to test the model in a situation designed to promote relatively independent processing of features. Such encoding might be expected in an incidental learning task in which the letters of which stimulus strings are composed are widely spread across the display, and the processing task discourages continuity between one feature and the next, such that processing gestalts tend not to occur.

Another aspect of the model which has not been investigated is the completeness of encoding. The model is set up to take account of the level of integration of the features encoded, but not to take account of differing amounts of the stimulus encoded. The model could be modified to incorporate a parameter reflecting the relative completeness of encoding. This parameter would be most clearly identified with the variance of performance, since lesser completeness of encoding of resource items would cause greater variability of the similarity of a probe to the set of available resources. In effect, this parameter would reflect the salience of the various experimenter-defined features. Used in combination, the two parameters of integration and completeness of encoding could take account of syllabification or chunking of the stimuli.

However, the model is not intended to be worked up into a fully predictive and explanatory model of concept formation, with complete processing and representation mechanisms. The episode model's chief merit lies in its heuristic utility in drawing attention to the importance of considering the way in which particular items are processed, emphasizing the level of integration at which items are processed as a chief factor in their later utility as memorial resources. This aspect of the model gives it enormous flexibility, and unites several perspectives on how multiple resources may be utilized by the cognitive processing system, each instantiated by a level of the integration parameter. At the lowest level of the parameter considered here ($\underline{r} = 1$), the features of stimuli exercise independent influence on performance (a "co-operative" model). Essentially, each feature functions as though it is a separate trace. The impact of a resource item of any degree of feature overlap with a probe is strictly in proportion to the degree of overlap between the two, independent of the degree of overlap of other items. A second important level is defined by values of the parameter near the Euclidean metric (an "interactive" model). At this level, features are integrated to some degree within items. The various encoded representations of items work in concert to produce performance which is greater than any single representation would produce, although the impact on the outcome of traces less similar to the probe is relatively smaller (i.e. change in an item with greater overlap causes greater change in the outcome than change in an item with less overlap). At higher values of the parameter, this tendency becomes exaggerated, until the only item exercising significant influence on the processing of the probe is that which has greatest overlap with the probe (a "competition" model). The importance of the episode model is in relating these various alternatives within a common framework. Rather than supporting any one at the expense of the others, the episode model suggests that one's experiences of items may function as memorial resources at any of these levels, depending on how they were processed when presented, which in turn depends on the task demands and available memorial resources at the time of the experience.

The nature of the episodic trace and the process by which presentation of an item enlists priorly encoded episodes as resources for its perception, recognition or classification remain unresolved. The guiding heuristics used to generate the experiments in this paper were that the trace consists of a record of the operations undertaken to process a stimulus, that these traces are not organized in memory in any fashion, and that enlistment of a trace as a resource is accomplished by a process analogous to reverberation. It was thought that very early processing of a presented stimulus is mainly bottom-up. However, as the system engages in this processing, prior episodes of specifically similar processing are automatically recruited, and commence to guide further processing of the item in a top-down fashion. This engagement of processes performed earlier is not thought to result from an active search by a central executive for similar experiences (as, for example, suggested by Miyake and Norman (1978) or Williams (1978)). Rather it is thought that early operations call up further operations to the extent that they have been previously integrated in the experience of processing other items (a distributed, passive search process). This paper has only attempted to describe one characteristic of this process: that the probability of the enlistment of a resource (a set of previously-conducted operations) depends upon its similarity to the probe, and that this similarity depends on the level of integration of components at which the resource was originally processed. Another factor which is likely to determine the characteristics of this process is whether the memorial resource recruited to assist the processing of a novel item is accessed via active analogy or via a failure to discriminate. In the former case, the learner is actively exercising a similarity strategy, perhaps quite consciously. This may occasion quite a different process of recruiting a resource than the second case, in which the learner fails to realize that the probe is in fact novel, and treats it as a re-presentation of an item which is in fact only similar. The recognition data shown in Section V ii of the last chapter suggests that in the studies of this paper most subjects were in the latter mode, erroneously treating most novel items as though they had been processed previously. The implications of this factor, and the larger problem of how best to conceptualize the nature of the trace, whether as a bundle of features (Bower, 1967), or as a record of the operations conducted on stimuli (Craik and Lockhart, 1972) or in terms of fluent re-enactment of the operations (Kolers, 1973), remain open questions.

III Extension of the Model: Word Perception

The episode model was developed in the course of investigating the knowledge base underlying concepts, and has been applied in this paper to the perception, recognition and classification of members of non-arbitrarily structured categories. However, its implications are not limited to such structured domains, nor to stimuli which can be considered to be organized in categories. It is primarily a model of learning, of how particular processing episodes modify the organism's knowledge base and its consequent processing and performance. Word perception serves as an example of the extended implications of the model.

Much of the interest in the study of word perception has centred on effects which suggest that words possess holistic properties beyond the properties of their constituent letters. One such phenomenon is the word superiority effect, which consists of the observation that a letter presented in a word is better perceived than when presented in a nonsense anagram of the word or when presented alone (Reicher, 1969; Wheeler, 1970). The effect has also been observed in pronounceable non-words (e.g. Baron and Thurston, 1973; McLelland and Johnston, 1977), and extended to the demonstration of mutual dependency in the perception of letters in pronounceable non-words (e.g. Baron and Thurston, 1973). Other studies of word perception have attempted to determine the nature of the representations responsible for the apparent holistic properties of words demostrated by the word superiority effefct. As discussed above in Chapter 1, Section III, Murrell and Morton (1974) suggested that word perception is accounted for via abstracted morphemes (clusters of letters bound into a unit through the meaning they carry), while Feustal, Shiffrin and Salasoo (in press) found that similarity in letter clusters other than morphemes also facilitated perception of probes. These studies suggest that words consist of relatively holistic compounds of their constituent letters, and that perception of a word is facilitated by its similarity at a configural level to a word in memory. However, they do not provide an account of the breadth of transfer of perceptual facilitation from a word represented in memory to other words or pseudo-words, nor of how the overlap of a probe with multiple words might affect its perception.

McClelland and Rumelhart (1982) provided an account of systematic generalization from words in memory to probe items. Their "interactive activation model" posits detectors for visual features, letters and words. Excitation of a feature detector results in excitation of consistent letter nodes, which excite consistent word nodes and inhibit competing letter nodes. Activated word nodes inhibit competing word nodes, while feeding back excitation to each of their own constituent letters. Thus the model assumes conceptually driven processes to be interacting with data-driven processes in determining perception. This model has been shown to be capable of simulating a wide variety of the phenomena of word perception (McClelland and Rumelhart, 1982; Rumelhart and McClelland, 1982). The word-level nodes, consisting of highly integrated compounds of letters, appear to be a major factor in the success of the model: these nodes produce mutual dependence of perception of letters within a word (by mediating mutual activation of these letters) and produce better perception of a letter in a word than presented alone (through feedback activation). A second important feature of the model is that facilitation of probes falls off relatively rapidly compared to the letter overlap of the probes with word nodes, so that the facilitation of a probe depends primarily on items sharing with it three out of four letters. This non-linear relationship is due to the multiplying effect of mutual activation between the letter and word levels. Another important aspect is that perception of a presented item is a function of the number of words overlapping a probe: probes with many close neighbours receive more activation (excitation or inhibition) than those with fewer. This activation from multiple neighbours is parallel, in that the neighbours contribute independently, but is also synergistic, in that the activation contributed by one neighbour at the word level feeds back through the letter level to all word nodes.

The parallels between this model and the episode model are striking, despite their origins in different research areas. Both employ distributed representation to achieve their effects, rejecting summary, abstractive representation: the interactive activation model is an alternative to direct representation of orthographic regularities as a basis of word perception, while the episode model is an alternative to prototype representation of categories. Like the interactive activation model, the episode model employs parallel processing of multiple, highly integrated memorial resources (with relatively steep generalization gradients) to produce processing dependencies between the constituents of items and to account for the perception of novel items. The differences between the models, and between the preparations in which they are applied, are informative. The interactive activation model possesses inhibition parameters, which may be of great use in accounting for active discrimination (as opposed to the passive generalization upon which the work in this paper is based). Secondly, the interactive activation model is a recursive model, concretely instantiating the interactive effects of top-down and bottom-up processing assumed to underlie the episode model. However, the interactive activation model is not concerned with the acquisition of the knowledge base. In contrast, the episode model was constructed to account for the effects of various factors affecting a developing knowledge base, such as the distribution of density of the stimuli to which the system is exposed. Most particularly, the episode model reflects through its integration parameter a concern with the effect on the knowledge base of processing done to an item at encoding, which it asserts will determine the similarity of the encoded trace to further items, and hence the breadth of facilitation that the item will later support. Although the studies in this paper have not systematically manipulated the degree of integration of encoding, the model emphasizes processing variations due to changes in instruction, strategy or task as a major factor in determining the evolution of the knowledge base.

This concern with integration of processing of the episode model allows it to directly address the issues in word perception raised by the word superiority effect. As mentioned above, the major finding in this area is that words exhibit holistic properties, such that the conditional probability of perceiving ~~one letter~~ given that a second is perceived is greater than the unconditional probability. This was also the major finding in the early experiments of this paper, that gave impetus to the development of a model which can deal with non-linear similarity relationships. The demonstrations in this paper of the speed with which stimuli come to exhibit processing interdependence, and the differential level of that interdependence depending on similarity to other items already processed, can be thought of as a beginning to the examination of how strings of letters come to be bound up into perceptual wholes, or more simply how words come to be words. The model suggests that what is special about words that gives rise to the word superiority effect is the large number of very close, well-integrated neighbours that common words possess which can facilitate their perception. But it also suggests that orthographically regular non-words which are like many familiar words will share this benefit, even to the extent of being better perceived than some less familiar words; while irregular non-words will be at a disadvantage, particularly because their distance from memorial resources is exacerbated by the non-linearity of similarity to well-integrated familiar words.

IV Challenges to the Sufficiency of Instance Models

This section is devoted to commenting on a number of issues raised from the prototype perspective, particularly by Homa and his associates, which were not dealt with in the body of the paper. One such issue is that of the emergence of abstract cognitive structures over time. The experiments discussed above all employed an immediate test, and hence do not provide evidence on the issue of slowly-emergent general structure.

This issue was raised by Posner and Keele (1970), who tested subjects on the learning of dot patterns immediately following training or after a delay of a week. They discovered that training stimuli suffered significantly more classification errors after one week than in the immediate test, while the number of errors made on the schema patterns did not alter significantly between tests. They argued that these results are consistent with the abstraction of a prototype during learning, since if abstraction occurred at test, one would expect loss of information about particular training items to be accompanied by a loss of ability to classify the schema patterns. This argument has been reiterated by Homa, Cross, Cornell, Goldman and Shwartz (1973) and Homa and Vosburgh (1976). Robbins, Barresi et al (1978) and Homa, Sterling and Trepel (1981) encountered similar transfer patterns, but believed the evidence insufficient to establish when abstraction had occurred. The suggestiveness of this evidence for the abstraction of a central representation is undeniable. However, one aspect of their data was consistently ignored by these authors: that classification of the training stimuli was, after the longest delay, never worse than classification of the schema patterns, and usually better. Thus the argument made in these papers must be seen in perspective as being based on an apparent relative decay of information about the specific training stimuli, not on any superiority of the prototype pattern for delayed classification.

Hintzman and Ludlam (1980) provided an instance-based account of these data. They constructed a computer-simulation model (MINERVA) which used only the training instances as a data base. This model assumed only that classification occurred on a nearest-neighbour basis, and that properties of instances were lost in an all-or-none fashion. It is notable that the set of properties on which the probe was compared to its neighbours included the relationships among elements of stimuli; thus MINERVA incorporates the processing interdependence of elements that was found in the early studies of the present paper, and which became basic to the episode model. This model demonstrated differential forgetting of the prototype and instances, paralleling the findings in the delay studies cited above. Hintzman and Ludlam concluded that differential decay of performance on training items and prototype patterns is an insufficient criterion for deciding between instance and abstractionist explanations.

An issue related to the emergence of prototype structure with time is the claim that prototypes are seen to emerge only when large numbers of members of a category have been exposed (Homa et al, 1973, Homa and Vosburgh, 1976, Homa, 1978, Homa, Sterling and Trepel, 1981). The claim is most impressive when set size is manipulated between subjects (Omohundro, 1981), which avoids differential contrast effects between categories. These papers assessed the ability of instance models to account for this phenomenon, and found them wanting. However, most of these studies tested nearest-neighbour models; the exception is the paper by Homa, Sterling and Trepel, which additionally tested a fixed set assumption (under which the model always compares a probe to a fixed set of instances, however near to or far from the probe) and a complete set assumption (under which the model always compares the probe to all instances); and they assumed a city-block metric of similarity for all models tested. However, the episode model proposed in this paper assumes that the number of instances to which the probe is compared, and additionally the similarity of the probe to each instance to which it is compared, varies with the type of experience the subject has had with the training items. It is thus not constrained by the assumptions made by Homa et al. As indicated in Section V iv of the last chapter, the episode model does predict a shift from superior performance on items similar to particular training items to items similar to the central tendency as set size increases, so long as increasing the set size increases the density of the category space. It is irrelevant whether this increase in density is occasioned by populating the centre of the space, or by populating empty portions of the space between previous training items (so that the mean distance of the items to the central tendency remains unchanged). Either change results in probes near the central tendency tending to become more similar to training items in general than are the training items. (However, increasing set size by exact repetition of items (i.e. exact re-processing) would not lead to such a shift). The only assumption required to achieve this shift with increasing set size is that processing a probe invokes multiple memorial resources, which entails that the level of integration of item processing not be at an extremely high level, such that only nearly identical items have any effect on the processing of the probe. Increasing set size need not result in centre-maximal distribution of density for the episode model to predict the performance shift. However (as indicated in Section V iv of the last chapter) when stimuli are generated as random distortions from a central pattern, the likelihood of such a change in the distribution of density is high, and (as indicated in the Introduction) such a distribution of density leads all instance models to predict superior performance on near-prototypical items. Thus the set size issue, like the temporal decay issue, fails to discriminate between prototype and instance models.

A third issue raised by Homa, Sterling and Trepel (1981) is that studies supporting an instance view (e.g. Brooks, 1978 and Medin and Schaffer, 1978) used categories that were not ill-defined. They cite Neisser's (1967) statement that a category is ill-defined when it consists of non-obvious dimensions, and the variety among potential members is essentially infinite. They go on to indicate that ill-definition is demonstrably important because of the apparently prototype-based results they achieve when employing geometric forms consisting of dot patterns with the dots connected. It may be that ill-defined categories do most closely match the conditions of extra-laboratory concept formation, but this challenge may be discounted for two separate reasons. First, the stimuli used by Homa et al can hardly be considered to consist of non-obvious dimensions. They themselves, following Posner and Keele (1968), described their stimuli in terms of the locations of the stimulus elements on just two dimensions, a horizontal and a vertical axis, and they accounted for subjects' performance through the distance of various items in terms of these two dimensions. Perhaps Homa et al intended to convey that unlike some of the stimuli used by Brooks and by Medin and Schaffer their stimuli consisted of dimensions which were non-obviously perceptually separable. However, if they felt such to be the case, a Euclidean similarity metric would have been more appropriate than the city-block metric they employed (c.f. Garner, 1974), although it might have reduced the appearance of prototype-based performance. The stimuli employed in the experiments in this paper also consist of fairly obvious dimensions, but as was indicated by the analyses of the perceptual identification experiments, they were not treated by subjects as consisting of separable elements, given the type of training imposed. It is thus arguable that the present studies meet the first of the criteria set by Homa et al for appropriate stimuli, and yet still support an instance-based account of concept formation.

The second criterion is more easily met. Even given the restrictions placed on item membership in a category in the experiments of this paper, such that each of the five stimulus positions may take only one of five values, only two of which are relevant to category membership (discriminate between categories), and that no item may take a relevant value from more than one category, the number of potential members of a single category is very large. Taking just the set of items deviating from the prototype pattern by two elements (the type most frequently used as training items in the studies above), each category has 60 potential members; and the number of members for each additional degree of deviation increases dramatically.

Yet another issue is the training of items to criterion in concept formation studies. Most prototype studies require subjects to reach a learning criterion of one complete errorless pass through all the stimuli before being switched to transfer (e.g. Posner and Keele, 1968; Homa et al, 1973; Robbins et al, 1978; Homa et al, 1981; Omohundro, 1981). Homa et al (1981) criticized Medin and Schaffer (1978) for not employing an errorless criterion, indicating that it is questionable after such limited training whether the old patterns have been stored in memory, such that it is questionable whether generalization to old items could be properly evaluated. However, the criticism and the practice of errorless criterion reflect an implicit value that trace strength and abstraction are the prime principles of learning. The implicit idea is that repetition of an item leads to strengthening of the trace of the item, such that the final representation of the item is an abstraction over the separate occasions of exposure to the item. The instance perspective suggests in contrast that each presentation of an item is a separate encoding episode, and that a later presentation of the (nominally) same item leads to a separate trace. A single presentation may or may not have a measurable effect on later performance, depending on such factors as the distinctiveness of the trace given the fashion in which the items were processed; but later presentations will be processed differently than earlier presentations, at least in that later processing will likely recruit representations of earlier presentations to assist in encoding the item. Thus later traces likely differ from earlier ones in being richer, containing the experience of processing the item in terms of the earlier trace. (This issue of just what the later trace will be like is complicated: as Jacoby (1978) has pointed out, immediate repetition of the item may lead to the second presentation occasioning truncated encoding of the item, since the products generated by the previous trial are immediately available.)

The experiments in this paper generally employed three repetitions of each nominal item, except Experiment 9 (Section V iv of the last chapter) which employed only one presentation of each item, but whose results were interpreted to be consistent with episode-based performance. The three repetitions are not conceptualized to sum over trials to form a strong, single trace; rather they are thought to form three traces which, on the whole, will be relatively similar to each other compared to traces of other stimuli. That is, they are thought to form a very local cluster within the space. This would not have been true had subjects been required to perform a different encoding operation on later presentations. In this case the traces of a single nominal item would be thought to be dissimilar. And to the extent that processing an item while employing an earlier representation of that item as a resource differs from the earlier processing of that item, the traces of earlier and later presentations of items are thought to be dissimilar in the present case, although subjects were required to perform the same nominal processing on each presentation of each item.

Some pilot studies have been run to attempt to examine the effects of differing numbers of repetitions. Data are not reported here because the results are very complex, and the investigation is just beginning. However, they point to both set size and density of set as factors predicting performance, and suggest that the effect of repetition of items is not only to increase set size but also to affect local density, initially by increasing local density, but later decreasing it. This is at least consistent with the encoding variability notion implicit in the instance tradition: while a second presentation may be processed comparatively like the first, a fifth repetition, using what is by that time a mass of previous instances as resources for processing, is unlikely to be encoded in a fashion very similar to the first. At this time the available data cannot be used to evaluate the prototype position; rather it points to yet another difference in the perspective taken by abstraction and instance accounts, and another difference in the kinds of investigation they lead one toward.

V Heuristic Differences Between the Abstraction and Instance Perspectives

Thus far this discussion has revolved around various models and their ability to account for data. This is useful to the extent that it causes a re-evaluation of values basic to the various perspectives. However, its utility breaks down at the point at which models of each of the perspectives have been so modified that they all predict all relevant data. As indicated in the Introduction, it is frequently difficult to force abstraction and instance views to make differential predictions. Moreover, the section above on prototype challenges to the instance perspective illustrated that important issues such as durability of conceptual performance and set size effects are non-discriminating between the perspectives. The studies in this paper demonstrated a number of cases wherein conventional and some unconventional prototype models could not account for the data, but the stimulus domains used are probably fairly rare in one's experience. Moreover, it is evident that models from either perspective may be modified repeatedly to take account of challenging findings. Finally, as indicated in Section VI of the last chapter, no data could be thought of that would differentiate between high-level feature-frequency models and the episode model. The question is thus raised of whether there is any utility in the distinction of abstraction and instance models.

The prime importance of the distinction lies not in the ability of one form of model or the other to adapt itself to challenging data. Rather it lies in the heuristic value of the perspectives, the kinds of abilities of the organism the various sets of orienting assumptions of the perspectives lead one to investigate. Prototype and feature-frequency accounts have served a useful purpose in drawing attention to the evident ability of people to abstract the essence of a mass of data. This is an important ability, serving as it does the portability of decision principles across domains and the economy of social transmission of information. Nonetheless it is a basic contention of the instance perspective that it is easy to overestimate the prevalence of the use of this ability. The instance view points to the importance of context dependency of processing in conceptual tasks, suggesting that far from being transportable across domains, conceptual performance is relatively dependent on degree of reinstantiation of the original conditions of learning in many cases. Even within a stimulus domain, a change of processing context can affect the inclusion of objects in categories, as demonstrated by Labov (1973). No doubt from the prototype perspective the incidence or prominence of these effects are over-estimated. The prevalence of abstraction vs. specificity of encoding is an empirical matter. The concern expressed here is that a potentially important way in which the organism functions should not be overlooked by default.

How one performs on that large set of occasions in which one is called upon to perform toward a stimulus object which is not presented in good form is a second area in which the instance view encourages exploration (Brooks, personal communication). However, the prototype view has little to say about this issue. Thus, for example, while prototypes are well set up to explain the classification of a dog which appears to the observer in profile at a moderate distance, the theory is ill-prepared to explain good performance when cues are reduced (dog seen at a great distance), when an unusual view is presented (dog seen from below) or when cues are missing (dog partially occluded by an intervening object). These cases present neither essential information (as a profile would) nor typical information (neither most frequent nor average), the kinds of information which constitute the raw data of prototype-based decisions. However, such special cases are very frequent, and in fact appear to be the rule rather than the exception. From an adaptation standpoint, if one's general task in life is to deal with objects which rarely present themselves in the good form of which they are notionally capable, it appears plausible that one might employ the specific familiarity of particular experiences of the objects as a basis of decision rather than their similarity to an essential good form. At the very least, such a decision strategy would form a useful adjunct to the use of a central representation. Suggestive evidence for the dominance of familiarity over good form in dealing with common objects has been collected by Brooks and Whittlesea (Reference Note 2).

If one's task is to deal with a succession of special cases, then fast learning, consisting of durable effects of minimal presentations, becomes necessary. As has perhaps become clear from the above, a basic value of the prototype position is that the acquisition of concepts proceeds very slowly, requiring many trials of exposure to members of a category and perhaps some additional time to allow the prototype to gel. By contrast, the instance position is that conceptual learning occurs very quickly, proceeding apace with the encoding of category members. It is for this reason, among others, that the training phase of the studies in this paper contain few trials compared to prototype studies. The current studies have consistently shown that the ability to deal with members of classes arises rapidly, given appropriate encoding procedures and (for classification) non-arbitrary class labels. It is unlikely that theorists of the prototype persuasion would deny the existence of fast, instance-specific learning;

however, they are likely to place little importance on it, thinking of it as a short-term effect, and only a stage on the way to real concept learning, that is, the evolution of a summary representation. Such a basic value prevents one from looking for long-term effects of encoding of particular episodes. However, evidence of just such effects is presented by contemporary memory research. Jacoby and Dallas (1981) demonstrated superior perceptual recognition of briefly presented words after a delay of 24 hours, while Jacoby (1983) showed the same effect after a week, and Kolers (1976) found reading speed for text originally read a year earlier to be superior to that of novel text. None of these studies attempted to contrast training materials with prototypical material; indeed, the notion of "typical text" or "typical common words" seems to make little sense. However, the durability of what appears to be episode-specific facilitation of processing is impressive. The instances perspective is well prepared to integrate with such findings in memory research, while prototype notions would not lead one to look for such outcomes.

Another issue to which the instances perspective directs investigation is the context-specific setting of hypotheses. While subjects may indeed employ active analytic strategies to solve the basis of category membership and/or for later classification of items (as discussed in Section VII of the last chapter), the use of particular analytic strategies may be occasioned by the similarity of the current processing episode to prior ones. Thus while the actual processing engaged in may be best described as analytic, control over the setting of the processing may be episode-based. This issue has been discussed in depth by Brooks (Reference Note 1).

Another heuristic purpose served by the instances perspective is a more intensive examination of the objective structure of natural categories. While Rosch's (e.g. 1977) statement that natural categories are clustered is no doubt widely true, the striking effects of manipulation of the density of clustered spaces in the experiments of this paper suggest that it is worthwhile to examine more closely the distributional characteristics of categories of interest. This issue achieves special prominence in attempting to train experts in some field. For example, should the training of medical diagnosticians emphasize typical aspects of diseases or concentrate on exposing particular cases of the diseases? The answer to such questions lies in part in the distribution of the cases. If cases are distributed such that there are many typical and few atypical cases, or if utility considerations are weighted heavily in favour of being right about the typical case at the expense of the atypical, then prototype training is probably the solution. However, under one of many other distributions, ~~for example~~ uniform distributions or domains of relatively widely spaced cases, or where utility considerations inveigh against errors on rare cases, then a specifically-tailored training programme emphasizing instance learning is likely more appropriate.

VI The Nature of Concepts

Our thinking about concepts as psychological constructs has been changed dramatically by consideration of instance-based knowledge structures. For Hull (e.g. 1920), Bruner, Goodnow and Austin (e.g. 1956), and still more clearly for the modern abstractionists of the prototype school discussed above, concepts were thought to consist psychologically of abstracted commonalities of a category. The referent of the term "concept" was quite clear; the mental representation of a concept was thought to be very similar to a formal statement of the concept. It was as though learners were thought to have attained a concept when they had succeeded in developing internally a direct representation of the criteria by which the experimenter had constructed the categories, a miniature copy of the blueprint for class construction. These orienting assumptions about concepts produce confusion when applied to instance-based knowledge structures. In what sense does a person who has only encoded his experiences of multiple members of a class "have the concept"? He does not have any direct representation of the defining features of the class, nor even of the general structure of the class. He has no direct knowledge of the essential principles governing class membership.

The response from the instance perspective is that the concept is an emergent property of the knowledge base, not an immediate property. The learner doesn't "have the concept", he acts it. That is, his performance toward similar things is similar, so that an onlooker (including the actor introspecting) can detect patterns of performance. If similar things are in the same category, as Rosch (1977) suggests, the result is that the learner classifies similar things into one category, which can be interpreted by the onlooker as conceptual performance; but the abstraction of a commonality across the various responses is a property of the onlooker, not necessarily of the knowledge base used by the learner to produce the responses. The set of resources for dealing with items that are formally in one category can be thought of as formally structured (Rosch's point); the set of responses possesses a similar structure, for the onlooker; but the representation mediating between them does not necessarily possess this structure.

One aspect of the difference between the instance view of concepts and the abstractionist view is what sort of task will be considered to constitute an appropriate test of conceptual performance. As mentioned above, research in the abstractionist perspective has favoured the classification task as the most clearly relevant test of concept attainment. This appears to be due to the abstractionist identification of concepts with knowledge of the commonalities across members of a class which define class membership. In contrast, the emphasis from the instance perspective is on generalization as the definiendum of conceptual knowledge, whether in a classification task or any other task. In other words, from the instances perspective the central problem in concept formation which makes it distinct from other areas of memory research is to account for good performance in many types of task on material that is novel but related to material processed previously.

APPENDIX 1

This appendix is intended to clarify the role of correlation in giving categories good structure, and to reduce the confusion arising out of employing both quantitative and qualitative stimulus values in a similarity space.

Rosch (1977) claims that "the world does, in a sense, contain 'intrinsically separate things' " in part because "real-world attributes . . . do not occur independently of each other." (p. 39 - 40). While not denying either statement separately, the author wishes to point out that the lack of independence is not responsible for the separability of objects. As discussed in Chaper 1, Section I i, the multiplicity of clusters is determined by the distribution of the separate stimulus dimensions. However, the issue of correlation of dimensions is important for other aspects of structure. Correlated dimensions permit items coded as whole units (relatively holistic coding) rather than coded in terms of their elements separately to yield more effective categorization in two ways. First, less than all information about the constituents of an item is required in a correlated domain in order to make a categorization, since the rest of the information is predictable given part. Second, the level of integration of encoding and the level of correlation of the dimensions interact. Holistic encoding has the effect of increasing the dissimilarity between encoded representations (see Section V v of this paper). If the dimensions are correlated, the effect is to increase the similarity between representations in the same category and decrease similarity between representations in different categories. Contrarily, when dimensions are uncorrelated, independent encoding of features best supports categorization (yields maximal contrast between categories).

The foregoing applies to dimensions of features which differ quantitatively. For the case of features which differ qualitatively, the notion of cluster structure is not so direct of application. Clusters exist in some dimensional space; qualitative features do not offer any direct way of unfolding the dimensions of the space. However, we can create a space in which items possessing qualitative features may cluster, through creating a set of similarity dimensions. Each dimension of the similarity space is to consist of the number of features each object shares with a particular object; for a domain of ten items, the similarity space would consist of ten dimensions, on each of which all ten objects are located in terms of their overlap with one object. Alternative spaces of lesser dimensionality could also be created using fewer standards of comparison; for example, one dimension could be created for each of the supposed prototypes of the categories.

In this similarity space, clusters of items are created by exactly the same principle as in the quantitative case: by differential frequency of values of a dimension, and not by correlation of the dimensions. Thus in both quantitative and qualitative cases, multiple clusters hinge on at least a relative discontinuity between local areas. Such a discontinuity occurs only where dimensional frequencies are at least bimodal.

This picture is confused by the fact that the similarity dimensions computed from a qualitative-feature universe are nearly always correlated to some degree. This is due to the fact that such dimensions are computed from standards which usually differ to some degree. Because they differ, no item can be found which is identical to all the standards, which means that it is impossible to fill the similarity space uniformly. This lack of uniformity creates unpopulated areas in portions of the similarity space farthest from the origin, which generally results in an overall negative correlation among the dimensions. However, it is evident that unless the frequencies of the similarity dimensions are at least bimodal, no discontinuity occurs, and hence no multiplicity of clusters.

Adding further to the confusion, features belonging to different dimensions of stimuli generally bear contingency relations with each other across items. These relations are sometimes confusingly labelled "correlated features". In fact, the contingency need not be mutual, as in correlation. Moreover, contingent features do not necessarily entail correlated similarity dimensions. And, perhaps most important, items may indeed be clustered in similarity space while the features are non-contingent on each other; that is, contingency relations are unnecessary for cluster structure.

The confusion is somewhat clarified by differentiating between "dimensional logic" and "feature logic". Under dimensional logic we can feel free to talk about clusters in spaces defined by continuous or at least ordered dimensions. The discussion above of the relative roles of correlated dimensions vs. shape of frequency distribution associated with dimensions is appropriate to this logic.

Under the heading "feature logic" the immediate concern is diagnosticity, or cue validity. Here we are interested in the separability of groups of items not in terms of a discontinuity on one or more stimulus dimensions, but rather in terms of a discontinuity in the distribution of cue validities of items for categories. The feature-logic analogue of the continuous-dimension discontinuity consists of a relative lack of items having approximately equal cue validity for both categories: it is in this sense that the non-existence of pigs with wings defines a discontinuity between the categories of mammals and birds. As in the continuous-dimension case, a discontinuity exists only if the distribution of frequency of diagnosticities of items for a particular category is at least bimodal. (Note that there are no guarantees attached to the accuracy of either type of discontinuity used as a cut-point dividing items into categories: given fuzzy categories, either strategy may err.)

Contingent features are not necessary for high diagnosticity: non-contingent features may yield high cue validities. Nonetheless, where contingent relations exist between features within categories, they may be used to improve the diagnosticity of an item for its category, a fact employed in feature-frequency models which code feature compounds as well as inidvidual features (e.g. Hayes-Roth and Hayes-Roth, 1977). Moreover, contingent features also offer the second benefit of correlated dimensions in the continuous case: namely, reduction in the necessity to encode all elements of an item to effect classification.

APPENDIX 2

This note illustrates the computation of the central tendency of P and K's stored patterns, and demonstrates the difference between their schema pattern and the actual central tendency. The salient data, taken from Table 4 of Posner and Keele (1968), is as follows: the tabled novel patterns in which they were interested had mean city-block dissimilarities of 54, 58 and 55 from the four training items, while the schema pattern had a mean dissimilarity of 52 from those items. They argued that this dissimilarity was roughly the same as those of the novel patterns (of Level 5), their intention being to argue a prominence of the schema pattern beyond its similarity to the training items, that prominence being due to its centrality. They thus argue the schema pattern to be equally similar to the old distortions as are the new distortions, and claim that the better recognition of the schema pattern is because "the prototype pattern must share the most common properties with the set of patterns generated from it" (p. 362). However, this maximal sharing is a property of the mean of a set of items, as later prototype theorists became aware.

It is unfortunate that Posner and Keele tabled only the distances from two old distortions to the four stored patterns; this introduces some indeterminacy in the computation of the central tendency below. The relevant part of the reduced matrix they presented is shown in Table A1.

The computation of the central tendency begins with the realization that the four old distortions can be treated as four points in a three-space, in which distances between points are to be computed using the city-block metric. The

Table A1

Distances from Stored Exemplar Patterns

to Each Transfer Pattern

(taken from Posner and Keele, 1968)

Stored Pattern	Schema	Old Distortion		New Level 5s		
		2	4	C	D	E
1	36	73	66	49	48	51
2	43	0	77	65	71	59
3	65	88	65	54	51	56
4	65	77	0	51	62	55
mean	52	59	52	54	58	55

object then is to determine a set of co-ordinates in three-space which satisfy such information as is given on the distances between the points. Let us assume that pattern #2 has the co-ordinates (0, 0, 0) on three axes X, Y and Z. We can then locate pattern #4 along a single axis, X, allocating it co-ordinates (77, 0, 0). Next, we can establish axis Y such that X and Y are orthogonal, and the X - Y plane cuts through all three points #2, #4 and #1. This permits us to define the co-ordinates of #1 as (a, b, c) where abs(a) + abs(b) + abs(c) = 73 and abs(a - 77) + abs(b) + abs(c) = 66. The co-ordinates (42, -31, 0) satisfy this relationship. The computation of the co-ordinates of the last point are only partly determined. However, we can erect a third axis, Z, orthogonal to the first two, and define the co-ordinates of #3 as before such that the sum of absolute values of (a, b, c) equals 88 and those of (a-77, b, c) sum to 65. This permits us to ascertain the value of a as 50, and to state that the sum of the absolute values of b and c is 38; however, without specification of the value of the distance from #1 to #3 we can do no better than to assign these as 19 each. Rotation of the X - Y plane such that the co-ordinates of #4 become, for example, (60, 17, 0) allows us to determine, through repeating the above operation, that the sign of the Y co-ordinate of #3 is positive; the sign of its Z co-ordinate is unimportant, since no other point takes any value but zero on that dimension.

Having how sets of co-ordinates of all four points, we can determine that the co-ordinates of the mean point of this three-space are (42.25, -3.0, 4.75). This allows us to determine three sets of deviations from this mean, one for each dimension. The average values of these deviations are 21.25, 12.5 and 7.13 respectively for the three dimensions. Summing these values across the independent dimensions gives a value of 40.88, which is the closest approximation to the minimum average (city-block) similarity to the four training items possible given the available information.

This value of 40.88 appears to be remarkably smaller than the 52 given for the schema pattern: in fact the schema pattern is more like the novel (level 5) distortions (having city-block dissimilarities of 54, 58 and 55) with which it is contrasted than like the central tendency.

APPENDIX 3

Conversation with various interested parties has indicated resistance to the higher-order metrics, with a preference for discussing only the cases of $\underline{r}$ = 1 and $\underline{r}$ = 2. (This preference is also apparent in the prototype literature; see, for e.g., Reed, 1972 or Hayes-Roth and Hayes-Roth, 1977.) These conversations suggest that the objection is based on the belief that the Euclidean metric has a unique status among metrics, in that distances remain invariant under rotation of the axes only under the assumption that $\underline{r}$ = 2. If this belief is justified, then in a case where the dimensions of psychological salience to the subject are rotated from those assumed by the experimenter, the only metric which would provide the experimenter with the same set of distances as those psychologically effective would be the Euclidean. Under any other metric the experimenter would receive a false impression of the effective dissimilarities among stimuli. This conclusion has been extended to suggest that since distance is not maintained under rotation of the axes for metrics both greater and less than $\underline{r}$ = 2, these cases are in some sense parallel, being both cases in which judgements rapidly approach uni-dimensionality , and that therefore no consideration of metrics greater than $\underline{r}$ = 2 is necessary.

Taking these objections in reverse order, even if distances were not maintained under rotation for metrics other than the Euclidean, the operations suggested by metrics greater and less than $\underline{r}$ = 2 are quite different. For $\underline{r}$ = 1 the outcome is the sum of input values, while for $\underline{r}$ = infinity the outcome is the largest input value. It is evident that the outcomes of these operations may vary independently, for example if the largest input value is held constant while the others are varied. Secondly, there is nothing "uni-dimensional" about either of these processes: outcomes of both reflect processing of the entire set of inputs. Thirdly, it is not true that for all values of $\underline{r} > 2$ that the outcome is the largest input value, nor indiscriminably close to it: for example, $(2^3 + 3^3)^{1/3} = 3.27$. The largest input value may dominate the outcome, but does not become identical to the outcome until $\underline{r}$ = infinity (although there may be a practical difficulty in detecting the difference for moderately large values of $\underline{r}$). In principle, then, metrics above and below the Euclidean are quite different; and their difference in practice is shown by the very different patterns of relative transfer predicted in this paper for metrics of 1, 2 and 3.

Turning now to the fundamental objection, the author contends that it is false that distance is not maintained under rotation of the axes for metrics other than the Euclidean. The point of contention appears to have arisen from consideration of cases like the following. Let us assume two points, A and B, in a two-space. Further let us assume a pair of othogonal axes, X and Y, in terms of which A and B are located at (x_1, y_1) and (x_2, y_2). The Euclidean distance between A and B in terms of the axes is given by the Pythagorean theorem, such that the distance $D_{AB} = ((x_1 - x_2)^2 + (y_1 - y_2)^2)^{1/2}$. Since it is undisputed that horizontal or vertical translation of the origin (translation in the direction of either axis) does not affect distance under any metric, the above expression may be simplified by translating the origin until A lies on one axis and B on the other: this leads to $A = (x_1, 0)$ and $B = (0, y_2)$, such that $D_{AB} = (x_1^2 + y_2^2)^{1/2}$. To discover the effect on distance of rotation of the axes, we will make use of the following strategy: following each rotation of the axes about the origin, we will justify the axes vertically and horizontally such that the points A and B again rest each on one axis. The result of rotating the axes 360 degrees in small steps will be that the origin of the axes will be found to trace out the circumference of a circle passing through A and B, such that A and B are maximally distant. The axes X and Y trace out a set of pairs of chords from A and B, each pair meeting at right angles at one of the various loci on the circumference at which the origin has been located. The line length AB can be computed from any pair of these chords, using the Pythagorean theorem, and lengths computed from different pairs of chords (different roations of the axes) can be compared to see if the distance AB is maintained under rotation. However, that is scarcely worth the effort, since the length AB must be a diameter of the circle, and it is clear that whatever pair of orthogonal chords are used, their hypotenuse, the diameter, must remain constant; thus it is evident that length remains invariant under rotation of axes under the Euclidean (Pythagorean) metric.

However, let us use these relationships to determine two pairs of chord lengths, one in which the origin of the axes is at A, and another in which the origin is midway between A and B (but of course still on the circumference). If we assign the diameter a length of one unit, then in the first case the chord lengths are 1 and 0 units, while in the second case they are each $.5^{1/2}$ (i.e. $(.5 \times 1)^{1/2}$. If we use the metric $\underline{r} = 1$, the distance AB = x + y, which for the two pairs of chords gives us AB = 1 and AB = 1.41; apparently $\underline{r} = 1$ fails to maintain constancy of distance under rotation. For $\underline{r} = 3$, the distance AB = $(x^3 + y^3)^{1/3}$, which yields AB = 1 and AB = .63, suggesting that $\underline{r} = 3$ also fails to maintain distance under rotation.

However, the argument is circular. The chord lengths, the values taken by A and B on the X and Y axes, were computed under the assumption that r = 2: they were computed under the assumption that $y = (D_{AB}^2 - x^2)^{1/2}$, where D_{AB} was known, for selected values of x. This computation instantiates the Pythagorean theorem, with its assumption of a Euclidean metric. It should thus be no surprise to find that other metrics fail to yield constant outcomes from pairs of inputs selected under this metric. The argument only illustrates that distance remains constant under r = 2 if the rotation is done in Euclidean space. The argument assumes that $(x^2 + y^2)^{1/2} = k$, a constant, and then computes values of x and y to fit. If instead we rotate in a non-Euclidean space, for example using $x + y = k$, we will find that the distance AB is constant under rotation in city-block space, while the Euclidean assumption fails to maintain distance under rotation. Similarly, if $(x^3 + y^3)^{1/3} = k$, which defines a higher-metric space, we will find that distance is not maintained under rotation under either city-block or Euclidean assumptions, but is under the assumption that r = 3. In general, distance is maintained under rotation under any metric which satisfies the relationship $(x^r + y^r)^{1/r} = k$. This gives us assurance that the Euclidean metric is not unique under rotation: if the psychologically salient dimensions are rotated away from the experimenter's assumed dimensions, the experimenter will not be misled about the psychologically effective distances between stimuli.

Stated in the terms above, the constancy of distance under any metric assumption which was used to derive the x - y pairs is fairly obvious. However, it is not obvious in many situations from which we draw intuitions, perhaps because we are very accustomed to dealing with Euclidean space. For example, if one plots the distance from the origin to A against the distance from the origin to B for various rotations, in the case where the distances were derived under the assumption that r

= 2 the plot forms a curve, a quarter-circle. It is immediately evident that the distance from the origin to all points on this circle is constant. Repeating the process for x - y pairs derived under the assumption that $\underline{r} = 1$ yields a straight line plot of the distance from A to B. Habit (or something) encourages us to look on the diagonal distance from the origin to this line as the AB distance, and it is obviously non-constant. But the application of diagonal measurement is the application of the Pythagorean theorem, which assumes that $\underline{r} = 2$; if instead we measure horizontally and then vertically and add the measurements, we find that the distance is held constant across the range.

There remain a number of problems involved in the use of Minkowsky metrics for modelling subjects' behaviour, including the determination of the number of dimensions used by the subject and the difficulty of obtaining sufficiently fine resolution of measurement to detect differences between moderately high metrics. However, indeterminacy of distances under metrics other than the Euclidean is not a problem.

APPENDIX 4

Medin and Schaffer (1978) described an exemplar-based model of categorization named the "context model". This appendix compares the context model with the episode model described in this paper. The context model has two cardinal features: its representation assumptions and its computation of similarity (Smith and Medin, 1981). The model states that some of the elements or dimensions of the stimulus may not be represented in the memorial trace resulting from processing the stimulus. This reduction of the stimulus would be thought to occur under selective attention to particular stimulus aspects at the expense of others. This notion permits the model to incorporate limited abstraction, although the abstraction is over certain stimulus dimensions rather than of a central tendency. This abstraction has several consequences. First, the memorial representation of a stimulus is not a literal copy of the nominal stimulus, as is perhaps suggested by an extremely naive instance view; rather the representation reflects what the learner was concerned with at the time of encoding. Second, as indicated by Smith and Medin, if the learner focusses primarily on those aspects which occur frequently among category members, the result is a representation of the category consisting of comparatively full and detailed representations of typical members of the category and relatively incomplete representations of atypical members. This is one way in which the context model accounts for the common finding of typical category members being dealt with more effectively than atypical members.

The second cardinal feature of the context model is its computation of similarity. The model assumes that the similarity of two items (e.g. a probe and an exemplar in memory) is the product, not the sum, of the differences between the items. Specifically, the model posits a similarity parameter, alpha, for each dimension on which a difference can occur between the items. This parameter takes the value 1 when there is no difference on that dimension, and is decreased toward zero by a difference in the psychophysical values taken by the items on that dimension. However, the value of alpha is also affected by the salience of the dimension, such that a difference on a more salient dimension yields a lower value of alpha than the same difference on a less salient one. The salience factor instantiates the strategic abstraction referred to above: unprocessed dimensions have an alpha value of 1, whether the items differ on that dimension or not, whereas items differing on a dimension treated as a necessary condition yield an alpha value of zero.

Computation of the similarity between items as the product of the differences, rather than as the sum of the differences as is generally done in the computation of similarity to a prototype, has been stressed as the most vital aspect of the model. The multiplicative vs. additive issue has been used to argue that prototype models cannot account for a variety of findings in concept formation (e.g. Medin, Altom, Edelson and Freko, 1982; Medin and Smith, 1981; Medin and Schwananflugel, 1981). The basis of this argument is that prototype models posit processing independence of the elements which make up a stimulus. This is conceptually similar to the tactic taken in this paper of demonstrating processing dependence between the definitional elements of stimuli, and hence arguing that prototype notions as usually described cannot account for the data. However, the context model and the episode model approach this issue in rather different ways. Multiplication of elements is frequently taken to indicate an interaction amongst them, which may have been why Medin and Schaffer elected to instantiate their interest in feature dependency through building multiplication into their model. In contrast, the episode model instantiates interdependent processing through the distance metric. This section attempts to argue that the distance metric is a more fundamental measure of interdependence.

To begin this examination, let us note that the goal of both models is to predict facilitation of a probe given that the system has experienced a set of items. It is assumed that facilitation is a function of the similarity of the probe to the old items. Taking a simple case in which the system has experienced only a single item prior to the probe, and the stimuli consist only of two dimensions, 1 and 2, the context model assumes that the similarity between the probe and item, S, is the product of their similarities on the two separate dimensions, S_1 and S_2, such that

1) $S = (S_1)\,(S_2)$.

Noreen (Reference Note 4) has pointed out that these similarities are functions of the distances x and y between stimuli on dimensions 1 and 2, such that $S_1 = f(x)$ and $S_2 = f(y)$, while the total similarity is a function of the total distance, z, such that $S = f(z)$. Noreen's problem was to discover the function f such that $f(z) = f(x)\,f(y)$; that is, to discover a function mapping similarity onto distance such that the total similarity is a multiplicative function of the unidimensional similarities, as specified by the context model. Now, the distance z in terms of the distances x and y is given by the equation $z = (x^r + y^r)^{1/r}$, the general distance formula, and takes a variety of values depending on the value of r. Thus Noreen's problem was to discover the function f such that $f(z) = f(x)\,f(y)$ and $z = (x^r + y^r)^{1/r}$ for r varying between one and infinity. His solution is that

2) $f(z) = e^{k((x^r + y^r)^{1/r})^r} = e^{kx^r} \cdot e^{ky^r} = e^{k\,(\text{distance})^r}$.

In words, similarity is an exponential function of the rth power of the distance multiplied by a constant ($k = -1$ is convenient).

This solution provides a link between facilitation and distance through the multiplication of similarities as specified by the context model. However, two further points remain to be resolved: first, it is evident that the computation of similarity rests upon knowledge of the appropriate distance metric (i.e. that the metric issue is fundamental to prediction of facilitation), and second, the success of the derivation of the function f does not indicate that a multiplicative relationship between total and unidimensional similaritites is appropriate. For example, if instead we assume that $g(z) = g(x) + g(y)$, and $z = (x^r + y^r)^{1/r}$ for the same range of r, we find that

3) $g(z) = ((x^r + y^r)^{1/r})^{kr} = ((x^r + y^r)^{1/r})^{kr} = (\text{distance})^{kr}$.

In words, similarity is a function of the rth power of the distance multiplied by a constant. The difference between the multiplicative similarity formula and the additive one is purely a scaling difference, as can be seen by comparing lines 2) and 3). The distributions of similarity with respect to distance for various values of r have precisely the same shapes and relationships to each other. Indeed, the multiplicative version is essentially an exponential transform of the additive version: ignoring the constant, whose function is simply to reverse the direction of the scales,

4) $f(z) = e^{g(z)}$.

The question of which of these versions is more appropriate is simply a utilitarian question of which provides the most direct fit with empirically derived data. To put this argument in other terms, the context model is of the form

5) $$S = \prod_{i=1}^{n} (\text{alpha}_i)$$

where n is the number of dimensions on which stimuli are compared. Letting d_i be the distance between the two stimuli on dimension i, the model implies that

6) $$\text{alpha} = e^{d_i^r},$$

although this was never formalized by its authors. But this is equivalent to stating that

7) $$S = \sum_{i=1}^{n} (\ln(\text{alpha}_i)) = \sum_{i=1}^{n} (\ln(e^{d_i^r})) = \sum_{i=1}^{n} (d_i^r),$$

which is an additive model. More importantly, if it is thought instead that simply

8) $$\text{alpha} = d_i^r$$

instead of an exponential function of the distance, then

9) $$S = \sum_{i=1}^{n} (\text{alpha}_i) = \sum_{i=1}^{n} (d_i^r),$$

once again.

To reinforce the argument that given the distance formula $z = (x^r + y^r)^{1/r}$ that the only decision is between multiplying an exponential function of the unidimensional distances or adding the untransformed values, it may be noted that neither

10) $$S = \prod_{i=1}^{n} (d_i^r), \quad \text{nor}$$

$$11) \quad S = \sum_{i=1}^{n} (e^{d_i^r})$$

could be valid forms, since both imply that $z = ((x^r + y^r)^{1/r})^r = (x^r + y^r)$. Thus the decision of whether to prefer an multiplicative or additive form depends on the decision of whether alpha is best conceptualized as an exponential transform of the distance or directly in terms of the distance. This decision in turn rests on an assumption of the appropriate metric for measuring the distance, which is not a formal property of the context model.

In contrast to the context model, the episode model formally incorporates the distance metric. Under the assumptions of this model, similarity between two items can be stated as

$$12) \quad S = (\sum_{i=1}^{n} d_i)^r.$$

This expression is obviously different from that in 9) above. The reason is that the episode model treats items in memory as the dimensions of the memory space, rather than elements of items as the context model does. The context model focusses on how each feature of a single stimulus functionally differs from each feature of a second stimulus, in terms of the psychophysical difference and the salience of the feature. It is less concerned with the interdependence of processing of all the features of the stimulus. This is surprising, given that the intent of the model is to instantiate the interaction of cues, but follows logically from locating the major parameter of the model at the level of the individual cue. The episode model locates its parameter at the item level, so that it does typify the degree of processing interdependence of cues. This difference between the models can be observed in their statements of the total distance, DT, and total similarity, ST, of a probe to a set of items previously processed. The context model, focussing on elements, implies that for j items in the set,

13) $$DT = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} d_{ij}^{\,r} \right)^{1/r}, \quad \text{and}$$

14) $$ST = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} d_{ij}^{\,r} \right),$$

while the episode model states that

15) $$DT = \left(\sum_{j=1}^{m} \left(\sum_{i=1}^{n} d_{ij} \right)^{r} \right)^{1/r} \quad \text{and}$$

16) $$ST = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} d_{ij} \right)^{r}.$$

While not instantiating the concern with interactive cues, the focus of the context model on the salience of particular features is important. The impact of the two models could be combined into a single model which describes both level of integration of processing and also relative salience of particular cues, such as

17) $$ST = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} w_{ij} d_{ij} \right)^{r},$$

where w_{ij} is a parameter representing the salience weight of each feature.

Despite the logical problems described above, the context model has had fair success in accounting for patterns of performance in categorization. Moreover, Medin, Dewey and Murphy (Reference Note 3) succeeded in demonstrating a better fit of a multiplicative model than an additive, and vice versa, depending on instructional conditions. At first sight, this success appears to render specious the above arguments concerning the indifference of multiplicative vs. additive combination of the values. However, these results really belong to a completely separate issue. The discussion above was based on multiplicative vs. additive combination of similarity parameters, which require some assumption about the relationship of similarity and distance. In the studies by Medin and his associates, no assumptions were made regarding the relationship of alpha to distance. In fact, the values of $alpha_i$ employed were derived as regression weights to maximize the fit to the data of the two statements

18) $$S = \sum_{i=1}^{n} (alpha_i)$$

19) $$S = \prod_{i=1}^{n} (alpha_i),$$

These alpha values are not parameters, but simple weights. Supposing that the functional underlying metric were $\underline{r} \neq 1$, the fit to the data using weights derived under the first restriction would be perfect. Supposing r = 2, weights derived under the second restriction would provide a good, although not perfect, fit (as can be verified by simply computing $z = (x^r + y^r)^{1/r}$ and $z = (x \cdot y)$ for various values of x and y). As r tends from 2 toward infinity, the fit of both would become worse, although the weights derived under the multiplicative restriction would always be a trifle better. Thus the issue of multiplicative vs. additive combination does have some relevance to the issue of interdependence of feature processing, although the fit to these two restrictions is only a crude indicant of the level of processing interdependence. (The episode model directly represents the level of processing interdependence through the parameter r, and is thus much more flexible, able to make definite predictions when $1 < r < 2$, or when $r > 2$.) However, as indicated above, the multiplication - addition issue reduces to a scaling decision when included in model statements specifying parameter values rather than in regression equations deriving weights.

REFERENCE NOTES

1. Brooks, L. R. On the insufficiency of analysis. Ms. in preparation.

2. Brooks, L. R. and Whittlesea, B. W. A. Unpublished data.

3. Medin, D. L., Dewey, G. I. and Murphy, T. D. Relationships between item and category learning: Evidence that abstraction is not automatic. Ms. submitted for publication, 1982.

4. Noreen, D. A note on the relation between similarity in the context model and distance in the Minkowski metric space. Unpublished ms.

5. Psotka, J. Colloquium delivered at McMaster University, 1978.

6. Vokey, J. R. and Brooks, L. R. Taming the clever unconscious. Ms. submitted for publication, 1982.

BIBLIOGRAPHY

Baron, J. and Thurston, I. An analysis of the word superiority effect. Cognitive Psychology, 1973, 4, 207-288.

Bartlett, F. C. Remembering. Cambridge University Press, 1932.

Beach, L. R. Cue probabilism and inference behaviour. Psychological Monographs, 1964, 78, Whole No. 582. (a)

Beach, L. R. Recognition, assimilation and identification of objects. Psychological Monographs, 1964, 78, Whole No. 583. (b)

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation (Vol. 1). New York: Academic Press, 1967.

Bransford, J. D. and Franks, J. J. Abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Brooks, L. R. Non-analytic concept formation and memory for instances. In E. Rosch and B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N. J.: Erlbaum, 1978.

Bruner, J. S., Goodnow, J. J. and Austin, G. A. A study of thinking. New York: Wiley, 1956.

Craik, F. I. M. and Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behaviour, 1972, 11, 671-684.

Craik, F. I. M. and Tul[illegible] Depth of processing and the retention of words in episodic memory. [illegible]urnal of Experimental Psychology: General, 1975, 104, 268-204.

Dawes, R. M. The robust beauty of improper linear models in decision making. American Psychologist, 1979, 34, 571-582.

Elio, R. and Anderson, J. R. The effects of category generalizations and instance similarity on schema abstraction. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 397-417.

Elstein, A. and Bordage, G. The psychology of clinical reasoning: Current research and approaches. In G. Stein, F. Cohen and N. Adler (Eds.), Health psychology: A handbook. San Francisco: Jossey-Bass, 1979.

Feustal, T. C., Shiffrin, R. M. and Salasoo, A. Episodic and lexical contributions to the repretition effect in word identification. In press.

Franks, J. J. and Bransford, J. D. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65-74.

Friedman, A. Framing pictures: The role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General, 1979, 108, 316-355.

Garner, W. R. The processing of information and structure. Potomac, Maryland: Erlbaum, 1974.

Glushko, R. J. The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 674-691.

Goldberg, L. R. Man vs. model of man: A rationale, plus some evidence for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.

Hayes-Roth, B. and Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behaviour, 1977, 16, 321-338.

Hintzman, D. L. and Ludlam, G. Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. Memory and Cognition, 1980, 8, 378-382.

Homa, D., Cross, J., Cornell, D., Goldman, D. and Schwartz, S. Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. Journal of Experimental Psychology, 1973, 101, 116-122.

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

Homa, D. and Vosburgh, R. Category breadth and the abstraction of prototypical information. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 322-330.

Hull, C. L. Quantitative aspects of the evolution of concepts. Psychological Monographs, 28, Whole No. 123.

Jacoby, L. L. On interpreting the effects of repetition: Solving a problem versus remembering the solution. Journal of Verbal Learning and Verbal Behaviour, 1978, 17, 649-667.

Jacoby, L. L. Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory and Cognition, in press(a).

Jacoby, L. L. Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning and Verbal Behaviour, In press(b).

Jacoby, L. L. and Craik, F. I. M. Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak and F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, N. J.: Erlbaum, 1979.

Jacoby, L. L and Dallas, M. On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 1981, 3, 306-340.

Jacoby, L. L. and Witherspoon, D. Remembering without awareness. Canadian Journal of Psychology, 1982, 36, 300-324.

Kolers, P. A. Remembering operations. Memory and Cognition, 1973, 12, 347-355.

Kolers, P. A. Two kinds of recognition. Canadian Journal of Psychology, 1974, 28, 51-61.

Kolers, P. A. Reading a year later. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 554-565.

Labov, W. The boundaries of words and their meanings. In C. J. N. Bailey and R. W. Shuy (Eds.), New ways of analyzing variation in English. Washington, D. C.: Georgetown University Press, 1973.

Lindsay, P. H. and Norman, D. A. Human information processing: An introduction to psychology. New York: Academic Press, 1977.

Mandler, G. Recognizing: The judgement of previous occurrence. Psychological Review, 1980, 87, 252-271.

McClelland, J. L. and Johnston, J. The role of familiar units in perception of words and nonwords. Perception and Psychophysics, 1977, 22, 249-261.

McClelland, J. L. and Rumelhart, D. E. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 1981, 88, 375-407.

Medin, D. L., Altom, M. W., Edelson, S. M. and Freko, D. Correlated symptoms and simulated medical classification. Journal of Experimental Psychology: Learning, Memory and Cognition, 1982, 8, 37-50.

Medin, D. L. and Schaffer, M. M. Context theory of classification learning. Psychological Review, 1978, 85, 207-238.

Medin, D. L. and Schwanenflugel, P. J. Linear separability in classification learning. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 355-368.

Mervis, C. B., Catlin, J. and Rosch, E. Relationships among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomic Society, 1976, 7, 283-284.

Miller, G. A. and Isard, S. Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behaviour, 963, 2, 217-228.

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

Miyake, N. and Norman, D. A. To ask a question, you must know enough to know what is not known. Occasional paper No. 2. La Jolla, Calif.: Centre for Human Information Processing, University of California at San Diego, 1978.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Murrell, G. A. and Morton, J. Word recognitionn and morphemic structure. Journal of Experimental Psychology, 1974, 102, 962-968.

Neisser, U. Cognitive psychology. New York: Appleton - Century - Crofts, 1967.

Neumann, P. G. An attribute frequency model for the abstraction of prototypes. Memory and Cognition, 1974, 2, 241-248.

Neumann, P. G. Visual prototype formation with discontinuous representation of dimensions of variability. Memory and Cognition, 1977, 5, 187-197.

Omohundro, J. Recognition vs. classification of ill-defined category exemplars. Memory and Cognition, 1981, 9, 324-331.

Osherson, D. N. and Smith, E. E. On the adequacy of prototype theory as a theory of concepts. Cognition, 1981, 9, 35-38.

Posner, M. I. and Keele, S. W. On the genesis of abstract ideas. Journal of Experimental Psychology, 1968, 77, 353-363.

Posner, M. I. and Keele, S. W. Retention of abstract ideas. Journal of Experimental Psychology, 1970, 83, 304-308.

Reed, S. K. Pattern recognition and categorization. Cognitive Psychology, 1972, 3, 382-407.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 274-280.

Reitman, J. S. and Bower, G. H. Storage and later recognition of exemplars of concepts. Cognitive Psychology, 1973, 4, 194-206.

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

Robbins, D., Barresi, J., Compton, P., Furst, A., Russo, M. and Smith, M. A. The genesis and use of exemplar vs. prototype knowledge in abstract category learning. Memory and Cognition, 1978, 6, 473-480.

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

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

Rosch, E. Human categorization. In N. Warren (Ed.), Advances in cross-cultural psychology. London: Academic Press, 1977.

Rosch, E. Principles of categorization. In E. Rosch and B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N. J.: Erlbaum, 1978.

Rosch, E. and Mervis, C. B. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Rosch, E., Simpson, C., and Miller, R. S. Structural bases of typicality. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 491-502.

Rumelhart, D. E. and McClelland, J. L. An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. Psychological Review, 1982, 89, 60-94.

Smith, E. E. and Medin, D. L. Categories and concepts. Cambridge: Harvard University Press, 1981.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. Psychological Review, 1937, 44, 430-444.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Tulving, E. Relation between encoding specificity and levels of processing. In L. S. Cermak and F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, N. J.: Erlbaum, 1979.

Tulving, E. and Thompson, D. M. Encoding specificity amd retrieval processes in episodic memory. Psychological Review, 1973, 80, 352-373.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Wheeler, D. D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

Williams, M. D. The processes of retrieval from very long-term memory. Technical report No. 75. La Jolla, Calif.: Centre for Human Information Processing, University of California at San Diego, 1978.