MODIFICATION MODELS OF CONCEPTUAL COMBINATION

by

Ormond L. Wilkie, Jr.

B.A. Philosophy, California State University, Long Beach, 1984

M.A. Psychology, California State University, Long Beach, 1986

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Massachusetts Institute of Technology

February 1992

Signature of Author____________________
Department of Brain and Cognitive Sciences
January 17, 1992

Certified by____________________
Susan Carey
Professor, Brain and Cognitive Sciences

Accepted by____________________
Emilio Bizzi
Chairman, Department of Brain and Cognitive Sciences

MODIFICATION MODELS OF CONCEPTUAL COMBINATION

by

Ormond L. Wilkie, Jr.

Submitted to the Department of Brain and Cognitive Sciences
on January 17, 1992 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Cognitive Science

ABSTRACT

Smith & Osherson (1984) proposed a model to account for typicality effects associated with adjective-noun conjunctions. The Selective Modification model assumes a feature-based representation for noun concepts. Complex concepts (e.g. *red fruit*) are created from simple ones (e.g. *fruit*) through adjustment of feature-weights limited to properties corresponding to the adjectival modifier. The ability of the model to account for a range of phenomena was reported by Smith, Osherson, Rips & Keane (1988). Medin & Shoben (1988) recently reported evidence arguing that generation of a complex concept representation involves modification of the weights of properties in the concept prototype beside the ones corresponding to the adjectival modifier, an effect not predicted by the Smith et al. model. A model to account for this finding is described, which exploits interproperty correlations in generating complex concept representation. Two experiments compare the Selective Modification model and the correlational modification model to predict judged typicality of instances in complex concepts. The correlational modification model is shown to be superior.

Thesis Supervisor : Dr. Susan Carey
Title : Professor of Cognitive Science

Table of Contents

INTRODUCTION

Problem of conceptual combination

Complex concepts are expressed in language through combination of nouns with other nouns, adjectives or adverbs (e.g. *pond water*, *fresh air*, *very cold temperatures*). Some compounds have become idiomatic expressions in the language (e.g. *smart bomb*, *house coat*) with stable meanings that are not combinatorially derivable from those of the constituents. Here I am concerned with complex concepts that are non-idiomatic novel combinations of nouns and adjectives. In this paper, the term "simple concept" will be used to refer to concepts lexicalized by a single word (e.g. *tree*, *bicycle*, *radio*).

Like simple concepts, complex concepts have prototype structure; that is, they have prototypes (Tversky & Kahneman, 1982) and differ in the extent to which various objects are seen to represent the concept (Tversky & Kahneman, 1982, Smith, Osherson, Rips & Keane, 1988).

Simple and complex concepts differ, however, in one important respect. Simple concepts are thought to have stable representations that are stored in long term memory. In contrast, since the number of potential combinations of concepts is countless, the mental representations for these expressions or phrases must some how be produced on the spot. Our ease in comprehending complex concepts in conversation and reading is evidence of the speed and efficiency with which the production process must routinely be engaged.

In the classical view of concepts, suited to the requirements of logic rather than to those of a model of mental representation, a solution to the problem of complex concepts is framed in terms of operations on sets (Cohen & Murphy, 1984). Concepts are represented by a set of entities which may include all members of the category corresponding to the concept, and perhaps a prototype (Osherson & Smith, 1981). Complex concepts refer to those entities that are members of the intersection of all of the constituent concepts. Judgments regarding a complex concept, such as category membership and the truth of category inclusion statements involving complex concepts, are presumed to be based on functions over these sets (Osherson & Smith, 1981).

This characterization, while adequate for purposes of logic and semantic analysis, fails to provide a basis for a psychological model of conceptual representation (Osherson & Smith, 1981, Smith & Osherson, 1988, Cohen & Murphy, 1984). Alternative solutions are based on intensional models of conceptual representation (Cohen & Murphy, 1984). Rather than being represented by a set of elements, concepts are viewed as lists of (weighted) features that reflect the central tendency of the category (Smith, Osherson, Rips & Keane, 1988).

Given an intensional representation of concepts, the prototype structure of a concept can be explained by reference to a <u>prototype</u>, as a summary representation of the members of a category. Rosch & Mervis (1975) showed that judged typicality is positively related to feature overlap among members of the category, with the best exemplars being those members with the greatest overlap of features. Similarity, measured by feature overlap, is a strong candidate for the relation associating an instance and a category, with that instance's judged typicality in the corresponding concept. A single representation for the concept, as the average member of the category, may be used as an alternative expression of this instance-category relation. Given weighted feature lists to represent category instances and prototype, typicality of an instance in the concept can be derived by a similarity function over the representations of the prototype and the exemplar (Tversky, 1977).

Extending this account to complex concepts requires a process for producing representations for these concepts. The process, ideally, would be "closed" (Murphy, 1988); that is, it would take as input only the representations of the constituent concepts of the compound. Such a model is far more appealing on grounds of simplicity than one requiring other, "external" information, such as reference to other exemplars, context, or higher level knowledge.

The representation of a complex concept may be the outcome of either a selective, or a compositional process. Selection models assume that complex concepts are represented by a set of exemplars, and that judged representativeness of a test exemplar is a function of its relatedness to this set. Since lions, tigers and jaguars may best represent the concept fierce feline for an observer, exemplar models claim that the judged typicality of an exemplar is dependent on its similarity to these animals. A selection mechanism, then, would be expected to describe a procedure for identifying a set of relevant exemplars for a given complex concept. Explicit proposals for such a set identification mechanism have yet to be presented. Set identification processes aside, selection models are closed processes to the extent that reference is not made to exemplars outside of the category in the formation of a complex concept representation.

A compositional process, in contrast to selective, is one in which a new representation is formed through combining the representations of the constituent concepts (Smith & Osherson, 1984) of a complex concept.

Composition might be thought to involve essentially equal contributions from constituent representations to the new complex concept representation. An analysis of noun-noun compounds provides evidence, however, that constituents of complex noun phrases don't contribute equally to the new representation. In analyzing the role of <u>mediating relations</u> in interpreting the meaning of noun-noun compounds such as *morning flight* , Cohen & Murphy (1984) suggest that there is an asymmetry in the roles of the concept constituents. For example, an understanding of *morning flight* requires knowing that "morning" refers to "a time of departure <u>for</u> a flight". *Morning* and *flight* are not simply blended; an understanding of the phrase requires the recognition that a morning flight is just a <u>type</u> of air flight. Through the mediation of the relation ("time of flight"), a particular value ("in the morning") replaces the default value for an attribute of the head noun. In this manner, the initial concept of the noun phrase modifies the final, head concept. Mediating relations appear to function similarly in adjective-noun combinations; understanding *red apple* requires knowing that the concepts are mediated by the "color" relation (Cohen & Murphy, 1984).

A model of conceptual composition, then, should account for the productivity of complex concepts and prototype structure. We seek a closed generative process which produces a new representation through modification of the simple concept representation by impact of the adjectival modifier.

Smith & Osherson (1984) presented a model that fits these requirements. For complex concepts composed of a noun and an adjectival modifier, the Selective Modification model explicitly describes mechanisms for transforming a weighted feature representation of the simple concept prototype into a representation of a complex concept prototype.

In what follows, I will show through arguments and experimental evidence limitations in the Selective Modification model as a model of conceptual combination. It will be argued that a model of conceptual combination is very likely non-compositional. Finally, a non-compositional modification model will be described and supported with evidence.

Structure of conceptual representations

In line with Smith & Osherson (1984), the structure of conceptual representations is here viewed as a "frame"(Minsky, 1975) composed of a set of attributes, each with one or more "slots" that correspond to particular values or features (see Figure 1). To illustrate, the attribute *number of wheels* for the concept *vehicle* may have slots for "2", "3", "4" or "18". The weight associated with particular features may be either binary or real, depending on the feature. For attributes such as *number of wheels*, in which members of a particular type of vehicle are not likely to vary (e.g. # wheels [car, 4]), the associated weight is 1 for the feature "4" and 0 for the rest. Weights assigned to features for attributes such as "color" are taken to represent the salience or intensity of the attribute for the concept. These weights in Figure 1 are arbitrary and are not required by the model to sum to any particular value.

Insert Figure 1 about here

In addition to the attribute-feature structure, a measure of the diagnosticity of each attribute in discriminating the concept from other concepts is associated with each attribute.

The Selective Modification Model

The Selective Modification model comprises three mechanisms which together generate a complex concept representation from that of a simple concept prototype and adjectival modifier: these are 1). feature selection, 2). weight migration[1], and 3). diagnosticity enhancement (see Figure 2). Feature selection assures that the modifier in the complex concept picks out the corresponding (target) attribute and feature in the simple concept representation[2]. Given selection of the target attribute, the weights of its features are redistributed. The cumulative weight of all features of the attribute is assigned to the target feature, and "0" gets assigned to all others. Finally, the diagnosticity weight of the selected attribute is boosted to reflect its increased prominence in the representativeness assessment.

Insert Figure 2 about here

A great virtue of the model is its simplicity. Since the modification mechanism operates on a single attribute, the Selective Modification model has no effect on other non-target attributes. The independence of attributes and the limited scope of the selection mechanism allows modification to proceed independently of other attributes. Since no information other than the simple con-

1. Smith, et al. (1988) refer to this mechanism as "vote shifting".
2. The selection mechanism selects a feature as well as one of its attributes. The feature and attribute selected will be referred to as the "target" feature and "target" attribute.

cept and modifier are required[3], this model is closed.

Smith et al. (1988) tested the Selective Modification model in a series of experiments examining the model's ability to predict the typicality of exemplars in adjective-noun and adverb-adjective noun concepts. Only the former data will be considered here. The categories used were fruit and vegetables. Attributes, features, salience and diagnosticity weights were derived from subject-generated feature lists of exemplars of both categories. Lists of features for exemplars of both categories were tallied, and those mentioned frequently were compiled into a single attribute-feature (color: red, yellow, etc.) list. Representations were constructed for each of the fruit and vegetable exemplars. Property weights for the features of each exemplar were estimated by frequency of mention, and diagnosticity weights were given by a statistic which estimated the strength of each attribute in discriminating between the two categories. Prototype representations for *fruit* and *vegetable*, as average members of the category, were approximated by averaging the weights for each feature over the set of category exemplars. So, for example, the weight assigned for the feature "yellow" for *fruit* was the average of the weights of the set of fruit exemplars. A second group of subjects rated the typicality of the same instances in both simple (e.g. "vegetable", "fruit") and complex concepts (e.g. "long vegeta-

3. There is clearly room for debate on this issue given the broad grasp required of the "feature selection mechanism". It is assumed here that for the general case, the selection mechanism can be imagined to operate locally in identifying the feature and attribute encoded by the modifier. As noted by Medin & Shoben (1988), though, a system interpreting "golden rail" needs to determine whether a "color" or "material" property is to be modified.

ble", "round fruit"). Typicality ratings for each exemplar-complex concept pair was averaged across subjects.

Results from their first study, on adjective-noun conjunctions, provide strong evidence for the Selective Modification model. From representations of the simple concepts, complex concept representations were generated using the the mechanism of the Selective Modification model, corresponding to those complex concepts in which typicality of the exemplars was judged. For each exemplar-complex concept pair, a typicality judgment was predicted for the model by computing the similarity of the respective representations. Predicted typicality based on the model-generated prototype representations was found to be correlated with directly obtained typicality judgments of the exemplars in the superordinate concepts.

Multiple property change

Despite its demonstrated success, the Selective Modification model is unable to satisfy all of our intuitions about our knowledge with respect to complex concepts. Intuitively, it is often the case that the attributes viewed as important, salient or diagnostic about a complex concept are not limited to those that are picked out by the modifier. The effect of a modifier in a conjunction often seems to involve the adjustment of weights among a number of features not directly related to the modifier. *Mahogany desk*, for example, does not call to mind an image of a Steelcase office desk that happens to be made of mahogany wood. Rather, it calls to mind a large desk with a long wide top, many drawers, etc. The Selective Modification model, limited to modifying weights among closely related features, cannot account for change, due to modifier impact, in property weights for unrelated features.

Medin & Shoben (1988) support these intuitions with data from three experiments. Consider one example: subjects were found to judge "metal spoon" to be more representative of *small spoon* than of *large spoon*, whereas "wooden spoon" was judged more representative of *large spoon* than *small spoon*. If **size** and **material** are independent attributes, then a change in material feature weights should have no effect on the representativeness judgments for complex concepts involving size modifiers.

One explanation for this example is just that the representation of *wooden spoon* is not the result of a novel combination of independent concepts, but is rather a precompiled representation of a <u>kind</u> of spoon. A characteristic of this kind of spoon is that it tends to be large. Most of the examples tested, however, appear to be non-controversial cases of novel, non-idiomatic combinations (e.g. wild/cultivated flower, busy/empty street).

Medin & Shoben (1988) explain their results in terms of interproperty correlations. Among spoons, size is not independent of material. A change in the **material** attribute is associated with a change in the **size** attribute. Through modification of relevant feature weights of *spoon* guided by correlational information, the representation of *wooden spoon* generated turns out to be more similar to that of *large spoon* than to the representation for *spoon*.

Medin & Shoben believe that complex concept representations are the product of a process of prototype modification, but not of a closed operation. Since associations among properties show that properties are not independent, they conclude that they must therefore be linked. The explicit links between properties reflect causal relations that are codified in or are derived from causal theories. Modification of properties is guided by reference to these higher order data structures.

An alternative account is possible which does not require explicit links among properties or reference to theories. According to Rosch (1978), correlations among clusters of properties are a feature of common categories. Rather than deriving interproperty associations from causal schemas, it is possible that modification is guided by a mechanism that simply exploits knowledge of the correlational structure of the environment (Murphy & Medin, 1985).

Correlations among attributes of category exemplars have been investigated. Malt & Smith (1984) presented evidence of extensive interproperty correlation among the properties of exemplars of the categories "bird" and "furniture". Having beaks, as might be expected, was associated with having feathers. Further, Medin, Altom, Edelson & Freko (1982) presented evidence of the use of interproperty correlations in a classification task (although, see Murphy & Wisniewski, 1990). If interproperty correlations among objects in the environment play a role in the definition of categories and are used in classification, it would not be surprising if they were used for the construction of complex concepts.

Correlational modification could work in the following way: To generate a representation for *wooden spoon*, exemplars for the category *spoon* would be inspected for properties correlated with the feature "composed of wood". The weights for these related properties in the *spoon* prototype would be enhanced or diminished depending on the strength and direction of the correlation. Presumably, a strong correlation of "composed of wood" with "largeness" would result in an increase in property weight for the "large" feature for the new *wooden spoon* prototype. The size-material correlation would predict similar changes in the generation of the *large spoon* representation. The increase in similarity between the representations would account for the Medin & Shoben (1988) results.

In the present studies, a correlational modification model of complex concepts will be explicitly described, and compared with the Selective Modification model in predicting judged typicality. The model that best predicts direct typicality judgments presumably does so on the basis of having generated a representation that more closely approximates the mental representation of the complex concept.

If the Selective Modification model demonstrates a stronger performance, it will be evidence that the simpler modification model represents a more fundamental process. Evidence of correlated properties in complex concepts would then need to be treated as a special case.

A superior performance by the correlational model would be evidence that modification of value of many properties incident to adjectival conjunction is the effect of a general mechanism, and not limited to a few selected cases. In so doing it would also provide evidence that the interproperty correlations found to exist among category members are exploited in reasoning.

EXPERIMENT 1: COMPARISON OF MODIFICATION MODELS

Experimental groups

The design of this experiment required the input of two groups of subjects. Group 1 provided typicality judgments and rated features for exemplars and complex concepts. Since Group 1 subjects evaluated a small number of exemplars, a second group (Group 2) estimated feature weights for a larger number of exemplars which were used as a basis for computing correlations. Details of data acquisition for Group 1 will be described here; details for Group 2 will be described prior to the evaluation of the correlational modification model.

Typicality judgments, and simple concept prototypes

In order to conduct a within-subject analysis, it was necessary to acquire feature weights for an adequate number of exemplars and complex concepts. This required that the categories selected be describable using a relatively small number of features. By comparison, Smith, et al. (1988) arrived at a list of 25 attributes for fruit and vegetable, with a little over 7 features per attribute. The categories canine and feline were selected as smaller, more homogeneous categories that might require fewer features.

To ensure that subjects would not get bored or exhausted, experimental sessions were limited to one hour. The experiment was therefore conducted in two parts, one for the feature ratings and the other for typicality ratings.

Method

Materials

A list of exemplars of each category (11 feline, 12 canine), including wild and domestic varieties, were assembled so as to contain members that were familiar to subjects yet ranged in typicality in the category (for feline: bobcat, calico cat, cougar, jaguar, leopard, lion, manx, ocelot, Persian cat, Siamese cat, tiger; for canine: bulldog, Chihuahua, cocker spaniel, collie, coyote, dachshund, fox, German shepherd, St. Bernard, Scottish terrier, wolf). To these lists were added a few non-members that were similar to members (e.g. ferret, fox for feline; otter, grizzly bear for canine). These instances were intended to set the boundary for membership.

Weighted property lists

The list of properties used was constructed initially by the experimenter and qualified through review by a set of subjects. This approach is intended to avoid biases that might be introduced through the feature-list compilation method (Smith & Medin, 1981, Murphy & Medin, 1985). Subjects, in listing properties, may include only highly diagnostic and highly salient features which are often perceptual or function-related. Subjects may also not spontaneously list non-perceptual features and those not associated with the manner of human interaction, nor properties that the majority of instances might possess and therefore would be considered "obvious" (e.g. "having acute sense of smell" for canine exemplars). Since the qualification of and weight assigned to features are based on frequency of mention across subjects, property lists generated by subjects, as well as the weights, may be strongly biased toward salience of features rather than, say, toward extent of possession. The resulting database may therefore not be sufficiently sensitive to less salient and non-perceptual properties.

The "reviewed list" approach attempts to reduce these effects by providing subjects with a comprehensive set of features to be rated, and by having them evaluate the <u>association strength</u> of each exemplar-feature pair. The "psychological strength of association" between an exemplar and a feature is assumed to combine intuitions of feature importance or diagnosticity, and extent of possession for an exemplar. The representations used differed from that of Smith et al. in the signification of the weight assigned to features. Since diagnosticity was now a "submerged" component of the association strength evaluated for the features, all attributes were assigned a unit diagnosticity weight. Increase of the diagnosticity weight of the target attribute (multiplying by the **d** parameter) in the simple concept representation allows implementation of the diagnosticity enhancement mechanism.

In investigating models of induction, Osherson, Stern, Wilkie, Stob & Smith (1991) showed that it was possible to predict similarity from weighted feature representations of mammals derived from direct property evaluation. One group of subjects rated a subset of 48 familiar mammals on 85 features, which were then normalized and averaged across subjects for each mammal, creating a database of exemplars and their feature weights. A second group of subjects rated the similarity of pairs of mammals. The obtained similarity ratings were correlated with predicted similarity for each pair generated using the mammal-property matrix and a simple similarity function over the 85 features. The average correlation (30 subjects) was found to be .64 (N=40 mammal pairs). Since the relation between the similarity of an exemplar and a category and judged typicality of the exemplar and the corresponding concept has been documented by numerous other measures (for review, see Smith & Medin, 1981) it is reasonable to expect that representations of instances generated through direct feature rating of association strength will serve as an adequate basis for computing predicted typicality of instances in complex concepts.

A preliminary list of features was constructed by the experimenter, drawing extensively on the list of properties used in the Osherson, et al. (1990) study. The list of fifty-six features was presented to groups of subjects with instructions to review them with respect to one of the categories. Subjects were asked to indicate features that are useful in distinguishing members of the category from members of other categories, or for distinguishing members within the category. They were also instructed to add such features that were not on the list. Approximately twenty subjects reviewed features for each category. No features were added and several were removed as a result of the review. A final list had 49 items. The features were resolvable into 28 attribute sets of related items (e.g. {timid, fierce},{territorial}). Table 1 shows the complete feature list.

Insert Table 1 about here

Subjects

13 MIT undergraduates were paid for their participation in this portion of the experiment (Group 1).

Procedure

All parts of the experiment were run on a Macintosh IIcx computer. The program, written in SuperCard, managed presentation of instructions, selection of category, ordering of stimulus items and data acquisition. Two sessions, one for typicality ratings and one for feature weight ratings, were each designed to last no longer than one hour. The typicality session was run first since the rating task took an entire hour.

Session 1.

One category was assigned by the program to each subject. Each subject was first presented with a list of mammals containing category members and several non-members. Of the thirteen Group 1 subjects, 6 completed ratings for canine and 7 for feline. Their task was to pick out items on the list that they were certain of being members of the category. This set of true members was next presented with instructions to select two members that were judged to be the most typical of the category. The remaining set of true members was presented with instructions to select two members judged least typical of the category. High and low typicality members were elicited to assure a range in judgment and property ratings. For each subject, all judgments of typicality involved these four exemplars with respect to either a simple or complex concept.

In selecting modifiers for each subject it was desirable to identify features that would have the greatest likelihood of being part of the representation for the simple concept. In a feature listing procedure such as was used by Smith et al., this function was served by the assumption that subjects tend to list features which are most salient and diagnostic; properties most likely to be represented were listed sooner and more often across subjects. Since the process by which the precompiled feature list was established did not involve any assessment of relative feature importance, it was required that this be directly assessed. To do this, each subject were presented with a list of 28 attributes and was asked to assign a number from 0-100 indicating the importance of each for distinguishing category members from members of other categories (e.g. *canines* from *bears* or *ungulates*). The feature members of the ten most highly rated attributes were put in a list. Since most attributes had more than one feature members (e.g. [diet: meat, vegetation, nuts/grain, fish], [habitat: coastal, desert, ocean, mountains, jungle]), this list often exceeded 10 items. 10 features randomly selected from this list were used as modifiers for that subject. Both the set of exemplars and modifiers, then, were selected by each subject.

Instructions for typicality judgments directed subjects to judge "how typical or how good an example" of the simple or complex concept each of their four selected high and low typicality exemplars were on a scale of 0-100, where larger values reflect greater typicality. Ratings could be entered either directly by use of the number keys, or by dragging a "button" along a scale marked with numerical endpoints and legends at appropriate ends labeled "good example" or "poor example". Each of the ten modifiers was adjoined to the superordinate concept label to form ten complex concepts (e.g. *large* feline, *domestic* feline, etc.). Typicality judgments were collected for all pairwise combinations of the four exemplars and ten complex concepts (e.g. "How good an example is *tiger* of *large feline* ?"), as well as the four exemplars of the simple concept (e.g."How good an example is *tiger* of *feline*?"). The 44 exemplar-concept pairs for each subject were randomly ordered, and presented one at a time.

Session 2.

"Cards", to be displayed on the computer, were designed listing all 49 features in several columns, with fields positioned beside each column for subjects to enter weights. To facilitate the task, attribute labels were placed above groups of broadly related features. A concept label was displayed at the top of the card, above a brief set of instructions. Subjects were told to assess the "strength of association" of the concept and each feature, indicating greater association strength by larger assigned value. Subjects were advised that since all instances were not likely to be associated with all features, it was neither required nor expected that a weight be entered for all 49 features for every exemplar or concept. No scale was imposed, but ratings were required to be positive integers.

Altogether, subjects evaluated 15 cards. One card was rated for the simple concept (canine or feline), one each for the four selected category members, and one for each complex concept formed from joining the 10 modifiers to the simple concept (e.g. *big canine*).

Subjects rated properties for the simple concept first. They were advised that they would not be permitted to revise these values after completing this card. Cards for each of the selected exemplars were rated next. The set of weighted properties associated with each exemplar will be referred to as a directly rated exemplar (DRE) vector.

After completing the exemplar ratings, subjects rated properties for 10 complex concepts. Weights entered for the simple concept were copied into corresponding fields for each of the ten complex concept cards. They were entered as both reference and default values in two columns labeled "Old" and "New". Values in the reference column could not be edited and subjects were instructed to review and edit values in the "New" columns. Subjects were free to move back and forth among the complex concept cards while rating, but were not able to change rated weights for the simple concept or exemplars. The set of weighted properties associated with each complex concept will be referred to as a directly rated complex concept (DRCC) vector.

Subjects were given a verbal overview of the property rating task, and were advised that they should be able to complete the task in one hour. To assist them in pacing themselves, each of the 10 complex concept cards had a number displayed in the corner reflecting its ordinal position in the stack, and current time was displayed in the upper right hand corner of the monitor.

Results

Analyses were all within subject, since there was substantial variation in which exemplars were viewed as most typical or atypical of the concept. Also, subjects exhibited variation in the properties selected as important, or diagnostic for the category.

Method validity

Recall that the concept representations used here differ in three ways from those used by Smith et al. They differ in the manner of selection of the properties to be rated, in the manner of assessing property weights, and in the assumptions regarding the quality of information represented (i.e. association strength as a measure of combined salience **and** diagnosticity). Before evaluating the modification models, it is necessary to establish that differences in the method of constructing the representations for exemplars and in the content of their representations are not critical to the investigation. It could be the case that the procedures for identifying or evaluating features result in the production of poor exemplar representations. For example, the reviewed list procedure used here may result in a bias toward diagnostic features that are not very important in typicality judgments. The representations generated by the reviewed list procedure might also be inadequate simply by not having a large enough number of features to capture enough of the important and distinguishing aspects of each concept. Where each concept was evaluated for 49 features in the present experiment, concepts for the Smith et al. study had weights assigned for over 100.

One way of establishing the validity of the exemplar representations is by showing that computed similarity between exemplar and complex concept representations, where all are obtained by the same methods, is related to patterns of judgment made by subjects of the typicality of the exemplars in the complex concepts. In other words, if the typicality of an object with respect to a concept is the similarity of the exemplar representation to that of the complex concept, and if the directly rated feature vectors are good approximations of their mental correlates, then the computed similarity of exemplar-complex concept pair representations should be correlated with the judged typicality of the exemplars with respect to the complex concept. This evaluation assumes that subjects evaluate features of complex concepts by the same processes of reasoning with which they evaluate those of exemplars.

To investigate the relation between computed similarity and obtained typicality judgments, weighted feature lists for each of the four directly rated exemplars (DRE) were compared with each of the 10 directly rated complex concept (DRCC) representations. The similarity, or predicted typicality of exemplar-complex concept pairs (weighted feature vectors) was then correlated with the judged typicality of exemplars in the complex concepts.

The similarity function used, a ratio model, has been used extensively in psychology (Gregson, 1975). The similarity of each instance-complex concept pair is given by the ratio of two sums. The extent to which the instances overlap (MIN) on each feature is summed across the set of features. Similarity is the ratio of this value to the summation across features of the maximum of each pair of values (MAX). The MIN term reflects the extent to which the instances share a feature; it reflects what is common to the instances. The MAX term is the sum of what is both common and distinctive for the pair.

$$\mathrm{SIM(A,B)} = \Sigma\ \mathrm{MIN}(A_i, B_i)/\Sigma\ \mathrm{MAX}(A_i, B_i) \tag{1}$$

The ratio similarity function returns values in the interval 0 to 1, and the similarity of an instance to itself is guaranteed to be 1, since there is no feature on which an instance and itself differ. Although the ratio function ensures symmetry and cannot therefore account for observed asymmetry in similarity judgment (Tversky, 1977), all judgments involved typicality and not similarity, and typicality judgments are asymmetric.

Predicted typicality is computed using the following equation, for an instance **I**, complex concept **CC** and attribute **i**.

$$\mathrm{TYP(I,CC)} = \Sigma\ \mathrm{MIN}(I_i, CC_i)/\Sigma\ \mathrm{MAX}(I_i, CC_i) \tag{2}$$

Tversky's Contrast Model of similarity (Tversky, 1977), a modified version of which was used in Smith, et al. (1988), was not used here because the function can return negative values, which might unnecessarily complicate the correlational modification model.

The set of 40 exemplar-complex concept (DRE/DRCC) similarity values was correlated with obtained typicality ratings for each subject. The average correlation[4] for the thirteen subjects was .62 (df = 38, *range.* = .40 - .83), $p \leq .001$. No difference was found for category, with averages for CANINE and FELINE

4. To correct for the skewness in the distribution of r, correlations for each subject are converted using the Fisher Z transformation prior to averaging, and the averaged Fisher Z converted back to a correlational value. This adjustment will be reflected in all averaged correlations reported in this paper.

subjects being .63 and .61, respectively.

In investigating complex concepts, Smith & Osherson (1984) found that subjects sometimes judge exemplars to be better examples of a complex concept than of the simple concept. It was found, for example, that apple is judged a better example of the concept *red fruit* than of *fruit*. The implication of this phenomenon, known as the conjunction effect, is that the exemplar is sometimes more similar to the complex concept than to the simple one. Another test of the directly rated exemplars, then, is their ability to match the pattern of occurrence of the conjunction effect that is found between direct judgments of exemplars in simple and complex concepts.

Smith & Osherson noticed that the conjunction effect tends to occur when the instance was found to be highly associated with the property value denoted by the adjectival modifier for a complex concept. Apples are strongly associated with being red in color. This type of instance was described as compatible with the complex concept. Since not all exemplars are expected to be compatible with every complex concept, in this sense, it is not expected that the conjunction effect would be observed to occur for every exemplar-complex concept pair.

A comparison of the judged typicality of exemplars with respect to both simple and complex concepts (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) revealed that on average, for the 40 judgments, a conjunction effect was observed in 11.3 (*SD* = 5.48) trials (28.3%) per subject.

The ability of predicted typicality values based on directly rated feature lists to agree with patterns of occurrence of the conjunction effect found in the direct typicality judgments provides an additional means of checking the validity of the exemplar representations.

In order to compute a difference in simple and complex concept typicality for the directly rated vectors, a representation for the simple concept was required. In keeping with the sense of a prototype as an average representation, and in order to conform to the procedure of Smith, et al., a representation for the simple concept (canine, feline) was created by averaging the values for each property, across the set of exemplars (DRE's) for each subject. This average vector (AVGVEC) was the only <u>derived</u> representation in this analysis since feature list representations for both exemplars and complex concepts were directly evaluated by subjects.

For the analysis, two sets of values are required for each exemplar-complex concept pair. One set is a pair of values representing the judged typicality of the exemplar with respect to both the simple and complex concept. The second pair of values are the predicted typicality judgments of the exemplar with respect to simple and complex concept which are determined by computed similarity. For the pair {apple, red fruit}, then, judged and predicted values are needed for the comparisons : 1) typicality of an apple with respect to fruit, and 2). typicality of an apple with respect to red fruit. For each exemplar-complex concept pair, a conjunction effect can be said to occur when the value assigned to the second of these comparisons exceeds the value assigned to the first. For each pair, then, values from obtained typicality judgments may indicate the occurrence of a conjunction effect. Similarity, the model may predict its occurrence. Agreement of predicted and obtained occurrence can be assessed by a X^2 goodness of fit test.

Using a 2X2 table in which cells contain tabulations, across subjects, of occurrences (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) > 0) and nonoccurrences (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) <= 0) for computed and obtained typicality judgments, a X^2 goodness of fit test showed the similarity function and directly evaluated representations to agree, in predicted occurrence of the conjunction effect, with actual occurrence ($X^2 = 13.29$, $df = 1, p < .001$).

These result shows that the directly evaluated feature lists, as representations of the exemplars, succeed in capturing important aspects of the concepts. It also shows that an explicit weight for attribute diagnosticity in the concept representation, or the use of the Contrast Model of similarity is not critical to the relation associating representations and typicality.

Evaluation of the Selective Modification model

To evaluate the selective modification model, complex concept representations are generated from the simple concept representation by migration of weights among the features of the attribute whose set of features includes the modifier-encoded feature. These derived complex concept representations will be referred to as MODCC's to distinguish them from those directly evaluated by subjects (DRCC).

In feature weight migration, weights for all features in the target attribute are summed and assigned to the target feature (see Fig. 2, weight shift mechanism). All other features of the feature set are assigned a value of 0. Property weights for other non-related features are unchanged in the new prototype.

For attribute **i**, and feature **j**, where $n_{ij}(P)$ is the property weight for the prototype, the feature weight for the modified concept is defined as follows:

$$n^*_{ij}(P) = \Sigma\, a_j\, n_{ij}(P), \text{ if the modifier matches feature } \mathbf{j}$$

$$= 0, \text{ if the modifier is an feature of } \mathbf{i}, \text{ but is not } \mathbf{j}$$

$$= n_{ij}(P), \text{ otherwise}$$

For the new model-generated complex concept representation (MODCC), weights are unchanged except for the features of the target attribute. In addition to the increase in magnitude of the target feature due to weight migration, diagnosticity enhancement (Figure 2, mechanism 3) is captured in the following way. Attributes may be thought of as each having a constant indicating change of diagnosticity that is due to the modifier. In the similarity computation, the maximum and minimum for each feature of the attribute is multiplied by this constant, thereby modulating the influence of the attribute relative to other attributes. Diagnosticity enhancement of the target attribute is implemented by multiplying each attribute weight by $\mathbf{a}_j$ where the value is assigned as follows:

$$a_j = d, \text{ if the modifier matches feature } \mathbf{j}$$
$$= 1, \text{ otherwise}$$

The value of this parameter was estimated for each subject as the value for which the correlation of predicted and obtained typicality was maximized.

For each subject, 10 model-generated complex concepts (MODCC's) were produced from that subject's simple concept representation (AVGVEC) and the set of selected modifiers. For each exemplar-complex concept pair, the corresponding feature vectors (DRE and MODCC) were compared using the ratio similarity function.

$$TYP(P^*,I) = \sum MIN\,(a_j\, n_{ij}(P),\, n_{ij}(I)) / \sum\, MAX\,(\, a_j n_{ij}(P), n_{ij}(I)) \qquad (3)$$

This set of predicted typicality values for each subject was correlated with their judged typicality ratings. The average adjusted correlation for the 13 subjects was quite strong (r = .72, *range* = .45 - .87, df = 38, p < .001). However, an inspection of the **d** values for which correlations were found to be maximized shows that the model is defective.

For 10 of 13 subjects, correlations increased monotonically with the diagnosticity parameter, approaching a limit at values of **d** in excess of 100. Even for the remaining 3, the average value of **d** that maximized the correlation was 35.3, much larger than the parameter weights found by Smith et al. (8.4 for fruit, 4.2 for vegetable).

The behavior of the diagnosticity parameter indicates that the only values that seem to matter for the model in predicting typicality are the feature weights of the target attribute for the instance and prototype. While the fact that it was not possible to find finite values for the diagnosticity weight for most subjects is a serious problem for the Selective Modification model, it is clear that even if the model had stabilized on a very large (e.g. 100) diagnosticity weight, it could not be correct. Any model, in fact, that does not incorporate some significant reference to the strength of relation between the instance and the superordinate concept cannot be right. To see this, compare the typicality of *daffodil* and *lemon* with respect to the concept *yellow flower*. Lemons are just as yellow as daffodils, yet most observers would find them less representative of *yellow flower*. The difference in perceived typicality, of course, is due to the fact that daffodils, and not lemons, are members of the category. But a model with diagnosticity weights exceeding 100 on "color" predicts that typicality for the two is the same.

A model that takes minimal or no account of instance typicality in the superordinate category makes sense under conditions in which the set of exemplars being considered are all pretty equally representative members of the category. Where category membership varies little, judgment seems to reflect greater sensitivity to factors that do show variation, such as property weight of *color* attributes in the present example. To extend the flower example, if the set of objects considered as examples of *yellow flower* included just members of the category *flower*, such as {rose, daffodil, tulip, daisy, marigold, lily}, then strength of association to the category might not figure as strongly.

The *flower* vs. *lemon* example demonstrates the importance of two factors in typicality judgment involving complex concepts: the value of particular exemplars on the modifier-encoded property, and the association of the set of exemplars to the simple concept. The flower example also suggests an explanation of how these factors interact with a given set of reference exemplars in the Selective Modification model.

An inspection of the Selective Modification model shows that it is responsive to both the weight of the exemplar on the modifier-encoded attribute as well as its typicality with respect to the simple concept. The weight shift and diagnosticity enhancement mechanisms focus on the target-relevant attribute. Similarity of exemplars to complex concept representation for non-target attributes provides, in effect, a measure of typicality vis a vis the simple concept. This is because the complex concept representation generated by the Selective Modification model, is, apart from attributes of the target property, identical to the simple concept representation, and the target attributes represent a single property dimension.

In the Selective Modification model, the size of the diagnosticity weight balances the relative contributions of 1). the target attribute of an exemplar, and 2). its typicality with respect to the simple concept in predicting the exemplar's typicality with respect to the complex concept. A small value of the diagnosticity weight reflects relatively greater influence of overall similarity (typicality of exemplars relative to the simple concept) and a lesser influence of target attribute of exemplars, where a large value for the diagnosticity weight indicates the reverse.

Smith, et al. assume that the diagnosticity weight is independent of the set of exemplars considered; that the relative contribution of the two factors remains fairly constant across conditions. For a given domain, then, the relative importance of the factors may differ, but the actual assessed typicality for an exemplar in a complex concept will <u>not</u> depend on the particular set of exemplars considered. The particular set of exemplars considered may impact the <u>range</u> of judgments elicited, but is not expected to change the value of the judgments themselves.

The performance of the Selective Modification model for the present study does not make sense under the "stable diagnosticity weight" assumption, since the parameter weights found imply that the typicality of exemplars with respect to the simple concept plays no role in judgment. The model's performance can sensibly be interpreted, however, under a different assumption. An alternative account is one that assumes a relationship to hold between the relative contributions of exemplar target attribute and overall similarity with respect to the simple concept, and the particular set of exemplars being considered. The value of the diagnosticity weight, then, would depend on the nature of the particular set of exemplars considered.

Note that this new proposal does not address a psychological process, but just one way in which the results of the Selective Modification model for the Smith et al. experiments and the present one can be consistantly interpreted. It may be the case that our judgment of representativeness is dependent on the composition of the set of exemplars before us, but the experiments do not present us with evidence of this and the account of the Selective Modification model to be described assumes that it is not. The current proposal, then, is intended to explain the performance of the Selective Modification model as an issue separate from the endogenous process underlying representativeness judgment.

The essence of the proposal that diagnosticity weight is a task-dependent variable that may be understood in statistical terms. To obtain a correlation among a set of pairs of data points requires variance in the values of the variables represented. Results using the directly rated feature vectors show that adequate range in values is expressed among obtained typicality judgments. The correlation found between obtained typicality and typicality predicted via the Selective Modification model suggests that there is variance in the predicted typicality values. Variance in predicted values must be derived from variance in either or both of the two factors, exemplar typicality in the simple concept or feature weights of the target attribute. The diagnosticity weight, in the similarity computation, acts as a fulcrum which balances the influence of these factors. Large values signify greater influence of the target attribute, and small values signify greater influence of exemplar typicality. In case variance for either of these factors is limited, then the model will predict the obtained scores best when it is most sensitive to the factor with the greater variance.

Now, notice that the amount of variance for either factor <u>must</u> be a function of the set of exemplars considered, and that the range expressed for these two factors is all that the model has to work with. Analysis of the following examples suggests how the diagnosticity variable works in the model. For the *yellow flower* example, given the set {sunflower, banana, lemon}, there is a range in the typicality of the exemplars, but none for the values associated with the feature *yellow*. The model will be most sensitive when it places less weight on exemplar typicality than on the feature value variable. The model will, therefore, be most sensitive for a <u>small</u> value of the diagnosticity weight. For the exemplar set {rose, daffodil, tulip, daisy, marigold, lily}, where all are typical of the simple concept, the model will be most sensitive when greater influence is given to the feature weights of the attribute associated with the modifier. This, in the model, is reflected in a very <u>large</u> weight for diagnosticity. The optimal diagnosticity weight, the value at which the correlation of the model and obtained typicality will be maximized, is then a <u>variable</u>, the value of which is a function of the range, for the two factors, of the particular set of exemplars used in the task.

The relationship of the variability of these factors to obtained typicality for the Selective Modification model may be summarized in and referred to as the diversity hypothesis. According to the diversity hypothesis, the performance of the Selective Modification model in predicting typicality is a function of the variability on associativeness with the target property for the set of exemplars considered, and variability in the typicality of the set of exemplars with respect to the simple concept. The diagnosticity weight is a variable whose magnitude serves as an indicator of the relative contribution of the two factors in predicted typicality given a particular exemplar set. The influence of either factor in computing typicality is a function of that factors' variability in the exemplar set.

A comparison of the evaluations of the respective exemplar sets used for Expt. 1 and the Smith et al. study provide support for the diversity hypothesis. Although the procedure used in the present study required subjects to identify exemplars perceived to be highly typical and atypical of the concept, selection of members was made from a list which had been qualified by the subject as containing only clear members of the category. The atypical exemplars were not, then, simply "animals that are atypical of canines", or "animals that are atypical of felines", but are "animals that are atypical canines" and "animals that are atypical felines". That is, the selections were conditional on prior category membership. If subjects followed the instructions given, there could be no non-members of the category on the list of atypical exemplars. The Smith, et al. study, in contrast, contained a very broad range of exemplars which included at least marginal, and perhaps in the judgment of some subjects, non-members (e.g. avocado as an instance of *fruit*, garlic as an instance of *vegetable*) of the category. The difference in judged typicality between the most and least representative exemplars with the concept, appears to be clearly greater for the exemplar sets used by Smith et al. This difference in range can be numerically compared through an examination of the typicality ratings for exemplars in simple concepts for the two studies. The range of the average typicality ratings can be expressed and compared as a percentage of scale. Averaged for the two categories examined, the difference between the best and worst examples of the simple concepts for the Smith et al. exemplars covered 65% of the scale. In comparison, the average range across subjects, as percent of scale, for the present study was 33.5% (N = 13, SD = 14.1). In judged typicality, then, the sets of exemplars used in the present study appear to be more homogeneous.

This difference in range of judged typicality is merely suggestive of limits in variability in computed typicality, similarity to the prototype. The diversity hypothesis interpretation of the very large values for the diagnosticity parameter for the present study is that the weight of influence in the similarity computation is strongly in favor of the target attribute. Given that the diversity hypothesis holds that the diagnosticity weight reflects the variability in the exemplar set of the two factors, this implies that the variability of typicality of exemplars in the simple concept is limited. The diversity hypothesis predicts that an inspection of the data for the present study will reveal a relation of judged typicality in the complex concept with target feature weight, and not with computed similarity to the simple concept representation.

To check for a relation for judged typicality between the complex concept and exemplar target feature weight, the value of the exemplar on the target feature for each trial, and judged typicality of the exemplar with respect to the target complex concept were correlated for each subject. The average correlation across subjects of target feature weight and judged typicality in the complex concepts was found to be .79 (*range.* = .55 - .93, $SD = .12$, $N = 13$, $p < .01$). To check for a relation of judged typicality between the complex concept and computed typicality of the exemplar in the simple concept, judged typicality of exemplars for each trial, and the computed similarity of the exemplar and average representations were correlated for each subject. As expected, no relation was found for these variables ($r = .02$, $N = 13$).

The failure to find a relation of judged typicality of exemplars in complex concepts and computed typicality of the exemplars in the simple concept could be due to either a lack of variability in values of computed similarity, or just to the fact that the exemplar and prototype representations were not related, by similarity, to the typicality judgments. To elaborate on the latter possibility, if the exemplar or prototype representation somehow failed to capture the important aspects of the concept, then a relation would not be found given any amount of range in their computed similarity.

The latter hypothesis is quite plausible. Smith et al. represented the simple concept prototype by the average of 15 exemplars, where only 4 were used here- two typical and two atypical. This small number might not be adequate for a good summary representation of the concept. Selecting high and low typicality exemplars for the category *bird* , for example, could result in the set {robin, crow, ostrich, penguin}, and it is easy to see that we might be suspicious of a concept representation for *bird* formed from their average.

Unlike birds, however, it does not seem that the categories used here include eccentric members that lack highly characteristic category properties (e.g. kiwis are featherless; penguins don't fly). Further, analyses of the directly rated exemplars and complex concepts, in which both types of representation were used in computing correlations and in predictions of the conjunction effect, suggest that the problem is not the representations. In these analyses, it will be recalled, none of the representations used were model-generated. The correlation of predicted and obtained typicality, then, shows that the exemplar representations capture important aspects of the exemplar concept. The problem, it appears, is the limited range in predicted typicality of the exemplars with respect to the simple concept.

Limited range in predicted typicality for a homogeneous set of exemplars in a complex concept follows from the nature of the Selective Modification model. For a set of exemplars and a few complex concepts, the Selective Modification has a limited capability to produce differences among the complex concept prototypes which would result in greater range, and presumably discriminability among computed similarities of exemplars. Complex concept representations generated by the Selective Modification model differ from one another by a small number of features each associated with a single attribute. They are identical in weight, in pairwise comparison, for the great majority of features. The effect of this commonality, for a given exemplar, is that if the target attribute is held out, the similarity of the exemplar to each of a set of complex concept representations (generated from the same simple concept) will be just about the same. If the set of exemplars are homogeneous in not differing radically across most properties, the range of computed similarity with the complex concept representations, based on non-target attributes, will be quite narrow. For this model, overall similarity, or typicality in the simple concept, becomes relevant only in the case where there is a broad range of diversity in the exemplar set. The requirement is not that there be a large number of exemplars, but that the ones considered are not much alike. The diversity hypothesis will be tested in Experiment 2, following evaluation of the correlational modification model.

Evaluation of the correlational modification model

To investigate correlational modification models, exemplar feature weight evaluations from a new group of subjects (Group 2) were used to build a matrix of correlations of features for each of the 2 categories (canine, feline). Group 2 subjects rated features for a number of instances for one of the categories, which were averaged and inspected for interproperty correlations. These correlations were then used to guide modification of the simple concept representation (AVGVEC) to create a new set of modified complex concept representations (MODCC's), which in turn were used as a basis for predicting the typicality ratings of Group 1.

Group 2: Exemplar-property database construction

Stimuli. The same categories, CANINE and FELINE, were used. Features were the same as those used by Group 1. Subjects assigned to a category rated the same 10 exemplars for that category (for feline: bobcat, tabby cat, cougar, jaguar, leopard, lion, panther, Persian cat, Siamese cat, tiger; for canine: bulldog, cocker spaniel, collie, coyote, dachshund, fox, German shepherd, Great Dane, St. Bernard, wolf).

Subjects. 10 MIT undergraduate volunteers were paid for their participation.

Procedure. Tasks were conducted in two sessions. In Session 1 subjects were asked to inspect the list of all category members to insure that all members were familiar and considered members of the category. The balance of their time for this session was occupied with tasks unrelated to the present experiment. In Session 2, subjects were instructed first to rate the superordinate concept for strength of association on the 49 features. They were advised that they would not be permitted to revise these values after completing that card. After this, subjects rated each of the 10 category exemplars for strength of property association on the set of features. All subjects evaluated features for the same set of instances, which were selected by the experimenter on the basis of general familiarity and diversity in the category.

Six subjects evaluated exemplars for the category canine and four subjects evaluated feature weights for the members of the category feline. The ten exemplars rated for each subject were normalized. Ratings for each exemplar were then averaged, feature by feature, across subjects. This resulted in two 10 X 49 (exemplar-feature) matrices. All pairs of features were then correlated over the 10 exemplars, producing a square (49 X 49, correlation-) matrix.

As a check on the richness of interproperty relatedness for the categories, a tally was made of significant correlations.

Insert Table 2 about here

Results of interproperty correlation analysis. Tallies of correlations by category are shown in Table 2. For both categories, a large number of correlations ($p < .05$) were observed between feature pairs (427 or 36.7% for feline , 332, or 28.5% for canine). The average percent of correlations exceeding p values (two-tailed) of .05 and .01 were 32.6%. and 18.8%, combining categories. These values, reflecting correlations between non-synonym pairs such as *timid* and *fierce* (1164 total pairs), are very close to those found by Malt & Smith (1984), of 33.5% and 18.3%.

Correlational modification

Generating a complex concept on the basis of these correlations requires using the correletions of the entire set of features with the feature encoded by the modifier. In order to compute correlational change for the entire set of properties of the simple concept prototype, it is necessary to know the direction in which to adjust each comparison property, and the magnitude of the adjustment. The correlation of each comparison feature with the target feature conveys direction as well as a measure of "confidence" with which the comparison feature weight should be changed. Strongly related features should be adjusted to the greater extent, with greater confidence, in association with the change in the target feature, than for weakly related features.

I adopted the following feature weight adjustment scheme; first, assume that the modifier adjusts the target feature to a maximal value (for the scale used here, 1), and second, use the correlation of the target feature with other features to proportionately adjust their weights in the average vector.

As in the Selective Modification evaluation, 10 new complex concept representations (MODCC's) were constructed for each of the 13 subjects in the initial group, using AVGVEC to represent the simple concept, the set of selected modifiers, and the Group 2 correlation matrix. To build a complex concept representation (MODCC), the correlation between the target and each comparison feature was used to increase or decrease the prototype feature weight. For a feature **j**, being modified by target feature **i**, the new weight $n^{*}_{ij}(P)$ is given by:

$$
\begin{aligned}
n^{*}_{ij}(P) &= n_{ij}(P) + (corr_{ij} * (1 - n_{ij}(P))) \quad \text{for } corr_{ij} > 0, \text{ and} \\
&= n_{ij}(P) + (corr_{ij} * n_{ij}(P)) \qquad \text{for } corr_{ij} <= 0 \\
&= 1 \quad \text{where} \quad i=j.
\end{aligned}
$$

For positive correlations, $corr_{ij}$ multiplies the difference of the original value ($n_{ij}(P)$) and 1, and the result is added to $n_{ij}(P)$, thereby increasing it by a proportion of the difference between the original weight and 1 (Figure 3). A correlation of .5 would, for example, increase a comparison feature weight (e.g. ".2") by half of the difference of the original weight and 1 (.6 = .2 + (1-.2) * .5). For negative correlations, $corr_{ij}$ multiplies $n_{ij}(P)$. The result, added to $n_{ij}(P)$ thereby diminishes it by a proportion of the difference between the original value and 0.

Insert Figure 3 about here

As noted before, the averaged exemplar simple concept representation (AVGVEC), exemplar representations (DRE's) and typicality data used in this analysis are those collected for Group 1. The entire contribution of the Group 2 data is in providing the property correlation matrix.

Analyses, again, were within subject. For each trial, a modified prototype representing the complex concept (MODCC) was generated via correlational modification and compared to the the exemplar representation (DRE). Similarity was computed using the ratio function, but since the correlational modification model does not make use of diagnosticity weights, the simpler version (Eq. 2) was used. The similarity for each instance/complex concept pair was correlated, as before, with obtained typicality judgments for the instance in the complex concept.

Across subjects, the average adjusted correlation of predicted to obtained typicality was found to be .59 ($N = 40$, *range.* = .30 - .75), $p < .001$. This result is very close to the correlation found earlier (.62) for typicality judgments predicted using the similarity of exemplars to directly rated complex concept representations (found in analyses used to establish method validity). This correlation, while not as impressive as the one found for the Selective Modification model, is important in that it was not obtained as the result of an artifactual, statistical effect. Because the correlation model uses no diagnosticity weight, it is not possible that the relative diversity of the particular set of exemplars used could have had a differential effect on the influence of either the exemplar's typicality or target attribute in the similarity computation. The independence of the model's performance from artifactual variables makes it a legitimate candidate for a model of conceptual combination.

Since the only difference between these analyses is the source of the complex concept representations, this close similarity in fit of the data suggests that the correlational model generates a representation that closely resembles the one produced through direct evaluation. Support here of the correlational model also provides further evidence that the failure of the Selective Modification model to demonstrate a contribution for overall exemplar-category similarity was not due to the quality of the simple concept representation (AVGVEC) since both models modified the same feature vector.

Several variations on this basic correlational model using slope information were examined to determine whether a better fit to the data could be obtained. Correlation, while providing an index of the strength of association of two sets of values, does not reflect the rate at which change along one dimension is associated with change in the other. This information is given by the slope of the regression line. Intelligence and agility may be highly correlated in mammals, yet the slope relating these measures may be shallow so that large differences in intelligence are associated with modest differences in agility. For this feature pair, correlation would be a misleading guide to modification. Slope therefore would be expected to be a more accurate guide to property modification.

One model exploited the Group 2 exemplar-property matrices to find the <u>slope</u> relating each feature pair, and adjusted feature weights as a function of slope. Other variations combined correlation and slope in weighting modificational feature adjustment. Surprisingly, for none of the models examined did the use of slope information improve the fit to data.

Another model examined the possibility that only features that were strongly related to the target feature are modified. In this model, correlations exceeding a certain magnitude affect change in the comparison feature. No threshold values were found at which the model performed better than the simple one.

A Chi-square analysis was run to see how well the correlational modification model predictions agreed with the direct typicality ratings for instances in simple and complex concepts in accounting for occurrence of the conjunction effect. As before, similarity of exemplar and both simple and model-generated complex concept representations were computed. The occurrence or non-occurrence of the conjunction effect was compared with it's occurrence based on judged typicality for the exemplar with respect to simple and complex concepts. Using a 2X2 table in which cells contain tabulations, across subjects, of occurrences (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) > 0) and nonoccurrences (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) <= 0) for computed and obtained typicality judgments, a X^2 goodness of fit test showed an association between the two measures ($X^2 = 45.46$, $df = 1$, $p < .001$).

Discussion : Experiment 1

The success of the correlational modification model provides evidence that correlations among properties of objects in the environment are used in a low level mechanism of reasoning. It shows that a simple mechanism can produce representations of complex concepts reflecting change across a number of features without requiring explicit linkage of features or access of higher level data structures. Perhaps most surprising about the correlational modification model is that it is very simple, requiring no external parameters.

The results of Experiment 1 motivated a second experiment to investigate the claims of the diversity hypothesis. Before testing these claims, however, it is worth looking more closely at the data with an eye to gaining a better understanding of the performance of the two models. Despite the strength of correlation found for the Selective Modification model, it failed, in the current context, in having been shown to be responsive to nothing more, in the representation generated, that the weights associated with the modifier-encoded attribute. In a broad sense, the relative success of the correlational model is ultimately due to its having generated representations of the complex concepts that come closer in approximating the mental representation it is modelling than those generated via the mechanisms of Selective Modification. The correlation obtained for judged typicality of exemplars in complex concepts and predicted typicality using all directly rated feature lists is evidence that the <u>directly rated complex concept</u> vectors are also a fair approximation of the mental representation. This, of course, assumes that subjects engaged in the rating task are in some sense "reading out" feature weights of the mental representation. By comparing representations generated by the modification models with the corresponding directly rated feature list, it might be possible to gain additional insight into the relation of similarity and typicality, and why the correlational model apparently worked where the Selective Modification model didn't. In the next chapter, a number of hypotheses will be advanced and examined in an effort to address these issues.

POST HOC ANALYSES

The Selective Modification model, in the context of Experiment 1, failed as a model of conceptual combination in being unable to demonstrate involvement of any part of the generated representation other than the modifier-denoted property in predicting typicality. The correlational modification model prevails, then, as the better model of conceptual composition. It was stated in the Introduction that if one model was shown to be better, the superior performance would be due to that model's having generated complex concept representations of greater fidelity to their mental correlates than the other. The availability, for each subject, of a set of directly rated complex concept representations makes it possible to examine this hypothesis.

In Experiment 1, computed similarity for each exemplar-complex concept pair was shown to be correlated with the subject's judgment of the representativeness of the exemplar in the complex concept. This relation was taken as evidence for the validity of the exemplar representations, which were directly rated by subjects, as approximations to the corresponding mental representation. This relation stands also as evidence for the validity of the directly rated complex concept representations as approximations of their mental correlates. On this assumption, it is meaningful to compare the outputs of the modification models, as themselves models of the complex concept, with these directly rated feature vectors.

The claim that the superior performance of one model is due to the greater fidelity of its generated representations to their mental correlates is in need of refinement. Intuitively, fidelity appears to be an alternate expression for similarity. Assuming that in the present context similarity is defined as extent of feature overlap, an example will suffice to distinguish representational fidelity and similarity. Imagine a pair of feature lists for the same set of properties. For each feature, the weight for the second list is equal to that of the first plus a constant. The "profile" of the features, for the two lists, is identical but offset by the constant. This agreement across features reflects a sense of fidelity in that the profile of one captures much of what is important in the other. Imagine a third feature list derived from the first, but in the following manner. On each attribute the value for the new list is again offset by a constant from the first, but the direction (positive or negative offset) is randomly determined. If the size of this second offset is less than the first, then it is clear that the overall deviation between features for the first and third lists will be less than for the first and second. The fidelity of the first and third, though, is clearly less than the first and second. Similarity, otherwise unspecified in meaning, is an ambiguous guide to some qualities of resemblance between objects.

Similarity as deviation

One sense of similarity can be characterized by the relative total deviation between feature weights for two weighted lists, each compared to a reference vector. A simple hypothesis regarding the modification models is that the correlational model succeeds by virtue of creating a complex concept vector that deviates less, across features, from the corresponding directly rated vector than the feature list generated by the Selective Modification model. This first hypothesis will be refered to as the deviation hypothesis.

To examine this hypothesis, the ten complex concept representations generated by each modification model were compared with their corresponding directly rated vectors for each subject. For each complex concept, the absolute difference for each pair of weights for each feature was summed across the property list. The average deviation per exemplar across subjects was 8.98 (N =13, *range* = 5.78 - 11.60, SD = 1.92) for the Selective Modification model and 10.06 for the correlational modification model (N = 13, *range* = 8.23 - 12.64, SD = 1.28). Surprisingly, the similarity, by this measure, of representations generated by Selective Modification was greater than for the correlational modification model (t = 3.12, df = 11, p < .005, 1 Tail). Clearly, then, the correlational model cannot be said to have generated representations that are more similar in the sense of absolute deviation.

Insert Tables 3 and 4 about here

Deviation provides an index of relative similarity. Feature overlap, in contrast, provides a direct index of similarity. The ratio function is an implementation of this sense of similarity. Although we expect feature overlap to be just a more refined measure of similarity, an examination of a few example vector comparisons will show that similarity as relative deviation is not a completely reliable guide to relative similarity via feature overlap. In Tables 3 and 4 examples show comparisons of pairs of vectors that have equal sum deviations of 18. Note, however, the difference in computed similarity. A review of the ratio function shows why this occurs.

$$\mathrm{TYP}(I,CC) = \sum \mathrm{MIN}(I_i, CC_i) / \sum \mathrm{MAX}(I_i, CC_i) \qquad (2)$$

Insert Tables 5 and 6 about here

For each feature pair in the first example, the denominator, (MAX, or 5) is greater relative to the numerator (MIN, or 2) than for the second example (8 to 5). For larger values, the relative difference for a given absolute deviation becomes smaller. In other words, large deviations at the top end of a scale, across property pairs for a feature list comparison may be associated with greater similarity than smaller deviations near the bottom of the scale. This is shown in Tables 5 and 6. Total deviation for the example in Table 5 is greater than for the Table 6 example (20 versus 12, or 75% more deviation). Despite the apparent handicap in deviation, the computed similarity for the Table 5 example is over 4 times greater than for the Table 6 example!

The lesson is that although the deviation hypothesis is false, it may still be the case that the similarity of correlational modification representations to directly rated vectors is greater than for those produced by Selective Modification. Having greater total deviation is consistent with being more similar in the sense of feature overlap. This could occur if the correlational modification model tends to "overshoot" on predicted feature values. The effect of this overshoot, as we've seen, could be to shift weights toward the upper range of the scale. This shift, as we have seen, could result in both greater deviation, and yet greater relative overlap of feature weight. This new hypothesis, the "feature overlap" hypothesis, claims that representations of the correlational model are more similar to directly rated complex concept representations than those generated by Selective Modification if feature overlap is the measure of similarity.

Similarity as feature overlap

To examine the feature overlap hypothesis, the ten complex concept representations generated by each modification model were compared with their corresponding directly rated vectors using the ratio similarity function for each subject. The average computed similarity to the average vector per exemplar across subjects was .668 ($N = 13$, *range* = .603 -.764, $SD = .051$) for the Selective Modification model and .645 for the correlational modification model ($N = 13$, *range* = .549 - .719, $SD = .046$). In another surprise, there was no significant difference in the relative similarity of the representations generated by the two models ($t = 1.32$, $df = 11$, ns, 1 Tail). The feature overlap hypothesis, like the deviation hypothesis, appears to be incorrect.

The conjecture that the success of the correlational modification model is explained by the model's ability to approximate mental representations appears to be false if the directly rated feature vectors for complex concepts are taken to represent fair images of the mental ones, and if similarity is understood exclusively in terms of either absolute deviation or feature overlap as measured by the min/max ratio. Another explanation is possible. Directly rated complex concept feature vectors and those generated by the correlation model might simply reflect the outcome of processes guided by distinctly different principles. Each principle might direct change for a given feature value, of different magnitude or opposing direction. The outcome vectors might each capture important yet somewhat different aspects of the information contained in the mental representation. If this explanation can be supported, it suggests that while both types of representations can be said to approximate the mental concept representations, neither can claim, on the present evidence, to do a better or more comprehensive job. It would further imply that the feature weight elicitation task, by which directly rated representations are constructed, engages processes other than interproperty correlation. The plausibility of this account is supported by the strength of the correlations for the two types of representations in fitting the data. The correlations for judged typicality of exemplars in complex concepts and similarity of exemplars to the two representations of the complex concepts were both very close in size, and while significant, not extremely strong. While each type of representation captured a significant amount of the variance, the variance accounted for by each model may be attributable to different sources.

Independent processes hypothesis

The hypothesis that representations created by correlational modification and directly rated complex concept representations are guided, in generation, by different principles that result in representations that each capture important yet different aspects of the mental correlates, will be called the "independent process" hypothesis. The guiding principle for the correlational modification model is pretty clear. On the direct ratings side, if subjects were not thinking in terms of correlations among features, what other principles could be a basis of determining property change? One possibility is that subjects were reasoning by reference to exemplars of the complex concept. Rather than correlating the properties of a broad set of feline-like exemplars for feature *ferocity*, for the example *fierce feline*, they might compute a value based on reference to one or a few highly representative examples of the complex concept. Another possibility is that subjects are engaging causal or theoretical knowledge embedding relations between properties, and changing feature weights of properties through a deductive process. It is not necessary to explain why each of these processes or strategies might direct change in property weight different from that of correlational modification, but only to note that alternate strategies are possible that could have such a result.

If different principles resulting in distinct feature weight profiles were engaged in generating directly rated and correlationally guided complex concept representations, it is expected that there would be limited agreement in <u>direction</u> of property weight change, across features, for the two processes. If different subsets of features were manipulated by each process, or if overlapping subsets were manipulated in different directions, then a feature by feature comparison of complex concept representations with the simple concept representation should reveal no overall agreement in the direction in which both processes modified feature weights.

The "independent process" hypothesis was tested by comparing the complex concept representations, from both direct feature evaluation, and correlational modification to the averaged feature list representing the simple concept. For each subject, directly rated and correlationally composed representations for the 10 complex concepts were compared, feature by feature, with the values for the simple concept. A 2 X 2 table was used to represent frequencies of agreement and disagreement in direction of change across features relative to the simple concept. Models could agree, for a feature, on positive change or negative-or-no change. They could disagree on directions, one model directing positive change and the other negative-or-no change. A X^2 goodness of fit test was carried out on the tallies for each subject. Results showed the two types of representation to be in agreement for 12 of 13 subjects ($df = 1, p < .05$), and in strong agreement for 11 of 13 ($df = 1, p < .01$). These results are stronger than those found for the same analysis using representations generated by the Selective Modification model. While the Selective Modification model did not change the values of many properties, it agreed with the directly rated vector on the few that were changed. While the computed X^2 showed agreement for 10 of 13 subjects ($df = 1, p < .05$) for the Selective Modification model, it found strong agreement ($p < .01$) for just 4 subjects. A comparison of analyses for the two modification models show that the strength of association for the correlational model exceeds that of the Selective Modification model for 12 of 13 subjects. On this evidence, then, it seems safe to reject the "independent process" hypothesis.

The general direction of change across features directed by processes generating directly evaluated complex concept representations and those generated by correlational modification is the same. This result provides a basis of confidence that both types of representation reflect the same patterns of change in feature modification. This result is consistant with a weaker form of the independent process hypothesis. Independent processes may produce very different representations, but they also may produce very similar ones. Evidence that the representations are not very different in the direction of change brought about by a modification process only counts as support for there being a common underlying mechanism, or strategy. It is not evidence against different mechanisms producing the same result. Processes other than correlational modification, such as the ones mentioned above, could guide complex concept generation and produce representations that agree overall with those produced by correlational change, but which do so via very different strategies or mechanisms. The task at hand, however, is to investigate correlational modification as a candidate for this process.

Correlational modification agrees with directly rated feature vectors in the properties modified and the direction of change of those properties. The representations generated via correlational modification, however, are less similar, by measure of deviation or feature overlap, to the directly rated representations than those generated by the Selective Modification model. And yet the results of analyses showed the great majority of attributes for the Selective Modification model to have no effect on the prediction of typicality. The earlier discussion on fidelity suggests a way of reconciling these results. If the correlational modification model has a tendency to "overshoot", to overestimate the amount of change in feature weight, for properties, then it might be the case that the generated representations resemble the feature weight profile of the directly rated vectors, but with exaggerated property weights. The representations generated by correlational modification, then, might just be caricatures of the directly rated complex concept representations. This sense of similarity, as similarity of patterns among weights of features, provides a basis for another hypothesis. The "feature profile" hypothesis claims that the representations generated via correlational modification are more similar to the corresponding directly rated feature vectors than those generated via Selective Modification if similarity is interpreted to mean relative magnitude of feature weights (among properties of the vector) between the vectors being compared.

Caricature hypothesis

A simple test provided some support that representations of correlational modification were in fact caricatures, or exaggerated imitations of the directly rated feature vectors. If computed deviations for each feature, by the correlational model, were systematically greater than those reflected by the endogenous process generating the directly rated representations, then the average percent of change in property weight, across properties, of the representations of each process and the simple concept vector, should differ. The average percentage of change among feature weights of the correlational model representations, from the average vector, according to the feature profile hypothesis, should be greater.

For each subject, the set of complex concept representations that were directly rated, and those generated via correlational modification, were each compared with the simple concept vector. Percentage of change in value for each pair of weights for each feature, were summed for each exemplar. Percent change as determined in correlational modification, it should be recalled, is a function of the direction of change, and the property weight scale. For positive correlations, the actual increase in feature weight is the product of correlation and the difference between the "old" feature weight and 1, the maximum scale value. A given positive correlational value has greater impact on small feature weights than large ones. For negative correlations, the actual deviation is the absolute value of the product of the correlation and the "old" feature weight. A given negative correlational value has a greater impact on large feature weights than small. Because of the confounds of size of property weight change and direction of correlation, percentage of change, as the term is used is for correlational modification, is a more accurate measure than absolute deviation. To compute percentage of change, for each feature, the absolute difference between simple and complex concept weights is compared to the weight of the feature for the simple concept. The ratio is the percent of increase or decrease for the simple concept feature weight. For directly rated vectors, the percentage of change for each feature was computed from the feature values for the simple and complex concept vectors. The average percentage of change per exemplar was then computed for both kinds of representation, by subject. An analysis revealed that the mean deviation per exemplar for correlational model representations was higher (M = 19.34, *range* = 15.68 - 23.37, SD = 2.711) than for the directly rated feature vectors (M = 17.36 , *range* = 12.07 - 22.08, SD=3.227). This difference was significant (t = 2.05, df = 11, $p < .05$, 1 Tail, correlated groups), supporting the explanation that the processes underlying the generation of both representations were both moving the same properties in the same direction, but that the correlational mechanism was directing a greater extent of change.

Making the case that representations produced by correlational modification are more similar to the directly rated feature vectors than those produced by the Selective Modification model requires that it is recognized that the representations generated by Selective Modification model cannot have extensive resemblance to the profile of the directly rated complex concept vectors by the measure of average property change, since only a handful of feature values deviate at all from the simple concept vector. Even if the weight shift resulted in a large change in property value for the target feature, the percent change for the great majority of features, compared to the average vector, will be 0. The overall average change cannot therefore be very large. An analysis for representations of the Selective Modification model measuring percent of change showed the average to be below 2% per feature per exemplar across subjects. This figure is far below those found for directly rated and correlational modification complex concept representations (17.36%, and 19.34% respectively).

The second requirement to make the case for a "feature profile" advantage for correlational modification representations is to show that the agreement in direction of property change, shown earlier for the correlational modification and direct rated feature vectors, is greater than the agreement of the Selective Modification model representations and directly rated feature vectors. The analysis described earlier in testing the "independent principle" hypothesis was done again, but this time using representations generated by the Selective Modification model instead of the correlational modification model. The goal, it will be recalled, is to see whether changes in properties in the simple cocept vector brought about through a modification process match the direction of change found through inspection of the directly rated simple and complex feature vectors. Again, X^2 goodness of fit tests were carried out on 2 X 2 tables for each subject. The results showed weaker agreement, across subjects, for the Selective Modification model representations than those generated by correlational modification. Results showed the directly rated and Selective Modification representation, to be in agreement for 10 of 13 subjects ($df = 1$, $p < .05$), and in strong agreement for only 4 of 13 ($df = 1$, $p < .01$). This pattern is not as strong as was found for the correlational model. Moreover, the magnitude of the X^2

consistantly showed greater agreement for directly rated and correlational representations than Selective Modification representations (12 of 13 subjects).

In investigating the internal structure of conceptual representations, then, similarity may be profitably viewed in terms of agreement among patterns of property weights rather than only in gross terms of summed deviation and shared features. In future investigations, these notions of similarity over profiles of properties should be refined.

EXPERIMENT 2 : TEST OF THE DIVERSITY HYPOTHESIS

The results of Experiment 1 for the Selective Modification model motivated a test of the diversity hypothesis. In this experiment, the procedure and for the most part, materials, of the first experiment will be replicated, but using a more diverse set of exemplars. Instead of evaluating feature weights for a small number of exemplars and a large number of complex concepts as in Experiment 1, subjects each evaluated feature weights for a larger number of exemplars and just a single complex concept.

Method

Subjects

Fifteen M.I.T. undergraduates were paid for their participation in this experiment.

Materials

The concept from which complex concepts were formed was *dog* . To assure adequate diversity in the set, the pool of exemplars included non-dog members, both canines and non-canines (cat, coyote, fox, grizzly bear, hyena, tiger, weasel, wolf) as well as dog exemplars (bassett hound, boxer, bulldog, Chihuahua, cocker spaniel, collie, dachshund, dalmatian, German shepherd, golden retriever, Great Dane, greyhound, Irish setter, lhasa apso, otter, pointer, poodle, Scottish terrier, Siberian husky, St. Bernard).

Non-dog exemplars were selected on the basis of physical resemblance to dogs. The purpose of including among the instance set only non-members that resemble dogs was to encourage serious and meaningful consideration on the part of subjects of the request to judge the typicality of such instances in the category *dog* or some complex combination involving *dog*. It was believed that subjects might think otters bad examples of dogs, but not a lot worse than, say, Chihuahuas. They might well, on the other hand, reject as meaningless the request to judge the typicality of a blue whale in the category *dog*.

Structure and presentation of the materials was very similar to Experiment 1.

Procedure

The experiment involved two tasks: typicality ratings and feature weight evaluations. The tasks were completed in two sessions each lasting less than an hour. To establish that the exemplars were familiar to subjects, in the first session subjects were presented with a list of exemplars and were instructed to eliminate unfamiliar ones. The list contained non-members and several common dog types of varying size and disposition (both types listed above). Ten of the animals composing the set of exemplars for the subject were drawn from this list. These were selected at random following the subject's identification of best examples of the complex concept.

Insert Table 7 about here

For each subject, one of a subset of the 49 features was randomly selected to serve as a modifier. The subset was selected by the experimenter on the basis of the property having a broad range in represented values among the set of instances (Table 7). All modifier properties were worded so that they could complete the phrase "dog that ...". The single complex concept for each subject was always presented in the form of this phrase.

To assure adequate range of diversity among instances for the compound, subjects were next presented with a more extensive list of members of the category *dog* (Afghan hound, bassett hound, beagle, boxer, bulldog, Chihuahua, cocker spaniel, collie, dachshund, dalmatian, German shepherd, golden retriever, Great Dane, greyhound, Irish setter, lhasa apso, Pekingese, pointer, poodle, Scottish terrier, sheepdog, springer spaniel, Siberian husky, St. Bernard) and were asked to select or name three members that were <u>most representative</u> of the complex concept.

To these three best examples of the complex concept, 10 familiar animals were added. Five familiar category members not selected as best examples and five familiar non-category members were randomly selected from the list to complete the set of exemplars which the subject used for subsequent tasks.

Typicality task

Typicality ratings were evaluated in rounds; the first round assessed typicality of exemplars with respect to the simple concept, and the second assessed typicality of instances with respect to the complex concept. All 13 instances were evaluated for each round, and the order of presentation for each round was randomly determined. Subjects were instructed to rate "how good an example, how representative the instance is" of the named category.

Property evaluation

The 49 features evaluated for each concept and exemplar were the same as those used in Experiment 1. Altogether fifteen cards were evaluated; one card was evaluated for each of the 13 exemplars, and one card each was used for evaluating the simple and complex concepts. Subjects were instructed to enter values indicating the association strength of each feature with the respective concept. Instructions also indicated that association strength for each concept or instance was to be evaluated in the context of mammals. The inclusion in the rating set of non-dog exemplars provided a built-in reminder that property evaluation was not to be judged with respect to just members of the category *dog*.

The card for the simple concept *dog* was rated first. The complex concept was evaluated next. Values entered for *dog* were copied forward into corresponding locations of the complex concept card under columns which were marked "New" to serve as default values, and also into a second column, marked "Old", which was to serve as a reference and was therefore not able to be edited. Subjects were instructed to freely change values in the "New" column. Cards for the exemplars were ordered randomly. Subjects were permitted to return to and edit previously rated cards, but were advised to complete the cards in the order presented.

Subject qualification

The typicality ratings of exemplars in complex concepts, and selection of best examplars of the complex concept jointly provided a basis for screening subjects. To provide a check subjects were not interpreting these different tasks, and that the modifiers selected were seen by the subject as salient for the category, a "2/3" rule was adopted. It was required that of the exemplars selected as best examples of the complex concept, two of the three were required to be among the three most highly rated in typicality for that concept. Two subjects were rejected because they failed this criteria, leaving 13 subjects.

Results

Categorical responses

A concern, noted above, was that subjects might reject the typicality task for trials involving non-dog exemplars. It is taken to be rejection of the task in case a subject enters a "0" rating for all trials involving non-member exemplars. Pilot studies showed that subjects who were asked to judge the typicality of non-members were more inclined to adopt a categorical response mode (across the board "0" ratings) than subjects asked to judge "representativeness". Therefore, subjects were instructed to judge typicality, and, in fact, no subjects made categorical responses to the non-member exemplars.

The effect of diversity of context on typicality judgment

The diversity hypothesis predicts that several effects will follow from an increase in diversity of the exemplar set. First, though, it is necessary to show that the method of choosing exemplars resulted in an increase in diversity, as reflected in greater range in judged and computed typicality for exemplars vis a vis simple concepts.

Given that this evidence of greater diversity is found, four predictions follow from the analysis given above of the major factors driving the Selective Modification model. First, judged typicality of exemplars with respect to the complex concept should be shown to be related to their computed typicality with respect to the simple concept. The judged typicality of *ocelot* in *fierce feline,* for example, should be shown to be related to how typical *ocelot* is of *feline,* where the latter measure of typicality is computed from the similarity of the two concept representations. Second, judged typicality with respect to the complex concept should be shown to be related to the target feature weights of the exemplars. How representative *ocelot* is of *fierce feline* should be related to it's *ferocity*. Third, given demonstrated association of judged typicality of exemplars with respect to the complex concept with target feature weight and typicality with respect to the simple concept, it is expected that correlations will be maximized for the Selective Modification model where the diagnosticity parameter reflects a contribution of both target attribute and similarity to the simple concept representation. Finally, the diversity hypothesis leads to the prediction that a change in the range of diversity of exemplars considered should cause a change in the value of the diagnosticity weight for which the correlation of predicted and obtained typicality for exemplars in complex concepts is maximized. Specifically, narrowing of the range of diversity should result in an increase in the optimal diagnosticity weight.

In the first analysis, range of scale used for judged typicality of instances in the simple concept is compared for the two experiments for evidence that the inclusion of nonmembers in the exemplar set increases the perceived diversity of the set. Range of judged typicality is just the difference between the maximum and minimum values obtained for the set of exemplars evaluated for representativeness in the simple concept. Inclusion of non-members of the category among the set of exemplars succeeded in increasing the range of judged typicality for all 13 simple concepts which included both members and non-members of the category. Across subjects, the average range of judged typicality of instances in the simple concept, for Experiment 2, was found to be 89.2% (N = 13, SD = 12.4) of scale. This was significantly greater than the range for Experiment 1 exemplars ($t = 10.69$, $df = 24$, $p < .0005$), which was 33.5% of scale. The range of averaged judged typicality for the Smith, et al. exemplars in the simple concept, it will be recalled, was 65% of scale. The presence of non-members in the set did not affect the range for exemplars which were category members. The average range for dog exemplars was 35.3% of scale (N=13, SD=.22, *range* =.02 - .75).

Given the effectiveness of the exemplar set manipulation in increasing diversity, it is possible to check for a relation of judged typicality in complex concepts to the computed relatedness of exemplars to the simple concept. To do this, judged typicality for each exemplar with respect to the complex concept was paired with computed typicality, which is the similarity of the exemplar and simple concept vector. The simple concept vector, for these analyses, was computed as the average over the representations of the 8 true category members only. Whereas no relation of these variables was found for the Expt. 1 data, they were strongly correlated in the Expt. 2 data for most subjects (r = .72, *range* = -.32 - .94, $N = 13$, $p < .01$).

To check for a relation of judged typicality with respect to the complex concept to exemplar target feature weight, these values, for the set of exemplars for each subject, were correlated. The average correlation for 13 subjects (r = .52, *range* = -.22 - .89, $N = 13$, $p < .05$) is significant. This relation, while weaker than the .78 of Experiment 1, is not significantly different.

Given the larger range of typicality among the exemplars, and relation of judged typicality in the complex concept with exemplar target feature weight and computed exemplar typicality in the simple concept, it is expected that performance of the Selective Modification model should not, as in Experiment 1, indicate a dominance of the target attribute through extremely large diagnosticity weights.

Evaluation of the Selective Modification model

The simple concept representation produced from averaging over the exemplars was again modified in accordance with the Selective Modification model's weight-shift mechanism. The diagnosticity parameter, **d**, was used to increase the weight of the target attribute in the similarity computation. The weights assigned to features of the target attribute, as before, were shifted to the modifier-encoded feature in constructing a complex concept representation.

Instance feature lists were compared to this diagnosticity-enhanced modified prototype, and the set of predicted typicality values correlated with obtained ratings. The average correlation was .86 (*range* = .70 - .95, $N = 13$, $p <$ .001). More importantly, maxima were found for all subjects but one at modest values of **d**, the average value of **d** for these subjects being 6.4. This value is comparable to that of the Smith et al. diagnosticity parameter, 6.2 (average for both categories). The maximum correlation for the remaining subject, reminiscent of the first experiment, approached a limit with increasingly large values of d. The importance of the target attribute reflected by the average diagnosticity value is greater than other attributes, as is expected, but is not so large as to cancel the influence of the non-target attributes. The influence of the non-target attributes can be seen in the fact that the correlation of typicality and modifier-encoded feature alone was only .52. Clearly, in the present context, the target attribute for the instance was not dominant.

As expected, then, inclusion in the exemplar set of marginal and non-members of the category resulted in performance of the Selective Modification model that comes much closer that reported by Smith et al. The strength of correlations and magnitude of the observed diagnosticity weight were comparable for the two studies. This result, in conjunction with the earlier results showing relations between judged typicality and the two factors, exemplar typicality and property weight for modifier-encoded feature, provide strong support for the diversity hypothesis.

This model was examined to see how well it predicted conjunction effects. As before, a 2X2 table was constructed, in which cells indicated frequencies, across subjects, of occurrences (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) > 0) and nonoccurrences (SIM(DRE,DRCC) - SIM(DRE,AVGVEC) <= 0) of the conjunction effect for computed (predicted) and obtained typicality judgments. The Chi square analysis showed the model to do a fair job fitting the results from direct judgments ($X^2 = 8.96$, $df = 1$, $p < .01$).

<u>Effect of exemplar set on diagnosticity weight</u>

The last prediction concerns variation in the diversity of the exemplar set and its impact on the optimal diagnosticity weight. If the diversity hypothesis is correct, and range in diversity among exemplars is the critical factor in balancing the influence of overall category typicality and target attribute for the Selective Modification model, it should be possible to show an increase in the diagnosticity weight given restriction of the range of typicality in the exemplar set. Such a restriction can be simply achieved by eliminating the non-dog exemplars from the analyses. This analysis assumes that since the best examples of the complex concept are represented in the set of category members, there is adequate range on the modifier-encoded feature among exemplars.

For an analysis of the Selective Modification model using just the 8 dog exemplars per subject, maximal correlations were found for all but 3 subjects. The average correlation was unchanged ($r = .87$, $N = 8$, $p < .01$), but the average value for the diagnosticity parameter was found to increase to 18.2. Parameter weights were larger, across subjects, under conditions of restricted range of typicality ($t = 3.16$, $df = 6$, $p < .02$), as was expected.

Evaluation of the correlational modification model

As in Experiment 1, feature weights for the complex concept were adjusted from those of the simple concept vector using correlations between the modifier-encoded and comparison feature. Because of the greater number of exemplar vectors available for each subject in the Experiment 2 design, correlations were computed within subject. Since typicality for exemplars was judged in a context that included peripheral and non-members, all exemplars were used for computing correlations[5].

As in Expt. 1, simple concept representations were modified by correlations of each comparison feature with the target feature. Each exemplar representation was compared with the complex concept representation using the ratio similarity function. Similarity, for exemplars, was correlated with judged typicality of the exemplar in the complex concept. Results of Experiment 1, for the correlational modification model, were replicated. The average adjusted correlation for the 13 subjects was .79 (*range* = .33 - .91, $df = 11$, $p < .01$). While the effect size is greater than was found in Experiment 1, the difference in correlations for this model was not statistically significant. A 2X2 Chi square analysis of predicted and obtained occurrence and non-occurrence of the conjunction effect showed the model to agree with direct judgments of typicality ($X^2 = 6.92$, $df = 1, p < .01$).

The replication of the earlier results demonstrate the reliability of the correlational modification model. The consistancy in performance shows the model not to be dependent on task-specific variables.

To investigate the effect of restricting the range of typicality on the correlation model, an analysis analogous to the one performed using the Selective Modification model for just category members was run. Limiting the typicality

---

5. A large number of analyses showed no difference in the performance of the correlation model if only representations for category members were used to generate correlations.

of the exemplar set did not reduce the fit of the model-predicted judgments to the obtained judgments. On the contrary, removal of non-dog exemplars resulted in an increase in the strength of the correlation, the average r = .96 (N = 13, range = .83 - .93).

The success of the correlational model warrants a closer inspection of the assumptions and conditions of its performance. The major assumption of the model is that modification involves input into the process of far more than representations of the constituent concepts of the complex compound. The correlations that guide prototype property weight adjustment are assumed to be derived from feature values of exemplars that are not constituents of the compound.

Further fleshing out of the model requires recognizing that correlational modification is a process that probably occurs "on the fly" in reasoning and comprehension. Modification, it will be argued, is not mediated by access of interproperty correlations that are stored in long term memory. Rather, each encounter with a novel noun phrase requires that correlations be computed and utilized. This, of course, implies that a set of exemplars is somehow identified by the system as a basis for generating correlations. In the next chapter, arguments and evidence will be offered in support of the view that 1). correlational modification is carried out "on the fly", 2). the exemplar reference set is a spontaneously generated, novel set of exemplars, and 3). the membership of the reference set is not constrained to membership in the category denoted by the simple concept.

CORRELATIONAL REFERENCE SET

Much of the appeal of the Selective Modification model is due to the fact that it is pretty close to being a truly compositional model of conceptual combination. Compositionality, it will be recalled, yields a representation for a complex concept from a blending of the representations of the constituent concepts. An explicit condition of compositionality is that the new representation require no other input than the representations of the constituent concepts. The requirement for qualifying the Selective Modification model as compositional, in this sense, is with solemn respect for the intelligence possibly required in guiding the "feature selection mechanism". For example, "gold" as a modifier can refer to either the <u>material</u> from which an object is composed, or to the <u>color</u> of the object. A feature selection mechanism should make a non-random guess as to which property values should be impacted for a given simple concept. Making correct decisions regarding the mapping of modifiers onto attributes and properties may require information that is not contained in the constituent representations. Even if, however, some "external-to-composition" processing was needed to identify the appropriate property and feature to associate with an adjective modifier, the mechanisms of the Selective Modification model, per se, may be viewed as taking as input nothing more than the representation for the noun concept, and the specifier for the encoded attribute. With some qualification, then, the Selective Modification represents a truly compositional model.

Interproperty correlations as prestored links

In providing an explanation for their results on multiple property change, Medin & Shoben (1988) theorized that interproperty relations may be prestored, structural parts of the conceptual representation. A model of correlational modification taking as input concepts whose representation included links between related properties might with some justification be construed compositional. It is not necessary, after all, that interproperty correlations be computed over exemplars. They might reflect the relative frequency and salience of associations among features that we, as observers, routinely notice in objects in the world.

There are advantages to a representational scheme involving prestored links between features. Access to interproperty relations would be rapid if no intervening computations are required. The problem of determining the source of the correlations could be distinguished from the mechanisms by which complex concept representations are generated.

Data from Experiment 1 suggest that the amount of storage capacity required may not be very large. Correlational analyses (Group 2) showed that correlations may be found for a third of all feature pairs. Also, analyses of variants of the correlational model showed that performance was not changed when only large correlations were used in modification.

Finally, there is an intuitive sense that salient relations that we encounter in the world between features are noticed in a manner that does not seem to be mediated by correlation over exemplars or instances.

One type of example presents a problem, however, for prestorage models. This type of example is one in which the modifier denotes a feature that 1). is not already associated with the simple concept, and 2). implies changes in a number of other features. An example of this is the concept *jungle-dwelling dog*. It may be allowed that before this encounter, most readers have not thought of this concept, so that the modifier can be assumed to have no prior association with the simple concept. Upon thinking of this concept, however, it is apparent that more is likely to be true of an instance of this concept than just the unusual ability. The diet of a jungle-dwelling dog would probably not include much canned dog food, and might include things like snakes.

It should be noticed that a model such as the Selective Modification model cannot account for this example. Even if a mechanism was added for purposes of appending a new feature to the representation (e.g. jungle-dwelling) of the simple concept, the model would have no means of recognizing or adjusting related properties. A prestorage model would fail for similar reasons. Even if properties related to the novel modifier were part of the simple concept representation, the prestorage model would need some means of connecting the modifier-denoted feature with these existing features.

The requirement for a model to "look outside" of the constituents of the complex concept means that a model that can account for these examples must be non-compositional. Information of some kind that is distinct from that which is contained in the constituent representations must be input to a model for producing a representation for the compound. One strategy would be to assume that many interproperty relations are prestored, and to propose models to account for the newly generated correlations. Once, however, a new model is required for this special purpose, it is more parsimonious to propose that all correlations are generated by the same process. If correlations are assumed to uniformly be derived, then there is no requirement that some of them be dealt with separately, and it is possible to posit conceptual representations that are not encumbered with the additional structure.

The case for an online, non-compositional model calls for an account of the new correlation-producing process. One possibility, suggested by Medin & Shoben (1988) is that correlations are derived from causal theories. This suggestion appeals to our intuition that interpreting something like "lip-reading dog" requires generating a causal "story" as an explanation for the talent. There are currently no explicit models describing the manner in which causal theories are represented or how their content is exploited for this type of application.

The results of both present experiments show, however, that it is possible to derive interproperty correlations that are psychologically valid from a set of exemplars without reference to causal theories. It then becomes necessary to provide an account of the means by which the set of exemplars that might have been the basis for learning or computing such correlations were identified. The first step, in providing this account, is to recognize that not everything is relevant to the problem of generating a particular complex concept representation. The population of Croatia is just not relevant to generating a representation of *cold toast*. The set of exemplars which formed a basis for the estimate of these correlations were, in both experiments, not simply a random set of concepts, but were in some manner relevant to the problem. In the next section, models for generating sets of relevant exemplars for complex concept generation will be examined.

The correlational reference set

A natural assumption, in thinking about the set of exemplars that are relevant to a modification process, is that the set is coextensive with the category denoted by the simple concept. What is relevant, by this assumption, to making a representation of *leather-bound book* out of a representation of the prototype for *book*, for example, are exemplars of books and nothing else. This assumption accords with our intuition that many things (e.g. suitcase, fish bowl, linoleum) are not relevant to the determination of feature weights for a representation of *leather-bound book*, and an important shared feature of this set is that they are not books.

The apparatus of a generally accepted characterization of prototype theory (Osherson & Smith, 1981) by which category membership is determined for a domain of entities through similarity to a prototype, provides a ready mechanism for constructing a correlational reference set consistent with this idea. Reference set members for a complex concept are just those domain entities that are members of the category of objects denoted by the simple concept.

An alternative hypothesis regarding the construction of reference sets, is that members of the set are a superset of the category corresponding to the simple concept. One motivation for this view is the observation that the line between members and nonmembers of categories is often fuzzy, and that in many cases nonmembers of a category appear to be relevant to the judgment of categorization or representativeness of an instance. *Fox* and *wolf* are not members of the concept *dog* , yet seem relevant to determining how good an example an instance is of a fierce dog. A second motivation for this view is the observation that complex concepts can incorporate modifiers that encode properties that no members of the simple concept category possess. For example, we have no difficulty in understanding the concept *hopping chipmunk,* although none of us are likely to have representations of chipmunks that hop. Such an animal might be expected to have long rear feet and large, muscular legs, exhibit hopping behavior, etc. A reference set identification model based on correlations derived exclusively from category members would have no basis for predicting these changes in weight for relevant features. A set identification model based on the neighborhood of objects around *chipmunk* , on the other hand, might include in the reference set *rabbit,* which does hop.

The same similarity-to-prototype and threshold model proposed for a reference set identification model above would capture the idea of a broad neighborhood of exemplars for a given target concept simply by manipulation of the threshold.

A problem with the "broad neighborhood" plan is that there is no way of knowing whether the chosen threshold value will succeed in assembling a reference set which includes exemplars that are relevant to prototype modification if those exemplars are not members of the category denoted by the simple concept. For the *chipmunk* example, a threshold value which is too constrained might include chipmunk-like things, yet exclude hopping rodent-like creatures like *rabbit* and *gerbil*. The selected set using this strict threshold, let it be assumed, would have little or no variability in value of the target feature across members. If there is no variation in values for this property there will be no correlation with other properties and no modification outside of the change in value for the target property which is mandated by the model.

No variation and negligible correlations are only a problem when there are no exemplars outside of the identified set which might be relevant. Much, if not most of the time, this result (no correlated feature change) is desirable. It is important to note that although evidence provided by Medin & Shoben (1988) and the present data support the case for property weight change for features outside of the target property, this is not evidence that extensive feature weight change happens <u>all</u> the time for <u>any</u> concept and modifier. Recall that the correlational analysis (Expt. 1, Group 2) showed that approximately a third of interproperty correlations, for the set of features examined, were significant in strength across the set of exemplars, this means that two thirds of the feature pairs failed to be strongly related. This can mean either that the average feature is associated with a few others, or that most features are associated with fleeting few others, <u>and</u> a minority of properties are richly interrelated. A review of the feature list suggests the latter possibility. Many features (e.g. "color" attributes for the *dog* category) are not related in any interesting way to other properties for members of or non-member objects near the simple category. These features may be useful for some other purpose, such as identifying an instance as an instance of the exemplar concept. For the "color" example, there seems to be just about nothing that is associated, across types of dogs, with their color. If no other properties are related to a color modifier, then the best representation of the complex concept prototype is just the simple concept prototype, biased, among color attributes, toward the target color.

In perhaps the majority of cases, then, the correct response of a model of conceptual composition would be to generate new representations that <u>don't</u> deviate in feature weight from the simple concept representation outside of the target attribute. The Selective Modification model is ideally designed to account for complex concepts with modifiers encoding this type of attribute. As stated above, a problem arises in cases where there exist relevant exemplars that are not "found" by the simple threshold set identification mechanism. The solution is not an arbitrary loosening of the threshold, since there is no assurance that the relevant exemplars will be included in the set. For *hopping chipmunk*, a less restrictive threshold function might include *rabbit*, but not *kangaroo*.

A solution to this problem is to make the set identification mechanism sensitive to differences in weight, across exemplars, for the modifier-encoded feature. One implementation would be to have the mechanism select for exemplars that are similar to the simple concept prototype, <u>and</u> that are highly weighted on the target feature. This could be expressed by a function that combines the two values for each instance (I):

$k = \alpha I_j + (1-\alpha)(SIM(I, P))$, where j is the value of the modifier-encoded feature

The value assigned for each object in the domain reflects a combined weight for the object on the two factors. For *hopping chipmunk*, each concept would get a number that indicates both how chipmunk-like the concept is, and how much it's associated with hopping. Rabbits would get higher values than kangaroo since rabbits more closely resemble chipmunks. A parameter, α, determines the relative contribution of the factors. Our intuition suggests that the value of the parameter would more highly weight similarity to the simple concept than weight of an instance on the target feature. *Pogo stick*, for example, is highly associated with the property "hopping", yet is not a good model for modifying chipmunks.

The set of relevant exemplars, on the basis of this model, would be those exemplars that are the most highly weighted on **k**. Criteria for set inclusion in this scheme is either a set size constraint, or a threshold, or a combination of the two.

Data from Experiment 2 was examined to see if evidence could be found for the set identification mechanism for complex concepts described. For each subject, a range of threshold and set size combinations could be used for set identification. Correlations for the modification model would then be generated from the produced sets. Evidence for the mechanism would be a demonstrated improvement in the fit of the correlational modification model to the data across subjects using sets of exemplars (from those included in the experiment) identified using this model.

Given the already high correlations found for most subjects using the model based on a maximal set of all 13 exemplars, a dramatic improvement is not possible. Systematic incremental improvement across subjects, while weak evidence, is the best that can be hoped for given the data at hand.

For each subject, values (k) were computed for each of the 13 exemplars using the exemplar and simple concept representations (averaged) and the modifier. Similarity of the exemplar and simple concept representations was computed using the ratio function. The parameter (α) was varied in the interval (0,1), and the set size was varied to include from 3 to 13 exemplars. Representations for the set of exemplars identified, the simple concept vector and modifier were then input to the correlational modification model. As in previous analyses, representations of exemplars and generated complex concepts were compared using the ratio similarity function. Similarity of exemplars with the modified complex concept representation was correlated with judged typicality of the exemplars in the complex concept. The set size and parameter weight for the reference set selection function at which the fit of the correlation model was optimized, for each subject, was sought.

Improvements in correlations of predicted and obtained typicality, using the reference set identification function, were found for all 13 subjects. For all subjects, therefore, the maximum correlation was found to occur at a set size of less than 13. The average adjusted correlation, across subjects, was .88 (*range* = .803 - .935, $p < .0005$, for 13 trials). This result provides evidence for this set identification mechanism. The average optimal set size, across subjects, was 7.31 (*range* = 3 -12, *SD* = 2.71). For no subjects did the mechanism identify all and only members of the simple category. A subset of dogs were selected by 7 subjects, and a superset for 5 subjects. The remaining subject had a maximum correlation at a set size of 6, of which half were dogs. The value of the parameter α for which correlations were maximized was .31 (*range* = 0 - .9, *SD* = .30). This value, which multiples the target feature weight, corresponds to a predominant weighting of the "typicality" term, the similarity of the exemplar to the simple concept. This result is not surprising; it is more important to relevant exemplars that they resemble the simple concept than that they possess a large weight on the target feature (e.g. *pogo stick* is a poor set member for *hopping chipmunk*). To see if there was evidence for a threshold function, the computed **k** values, across exemplars, were inspected for members of the optimal set size for each subject, and the minimum value identified. In the scale or (0,1), the average of the minimum **k** values was .58 (*range* = .27 - .81, *SD* = .175). The range and variability in minimum **k** values suggest that threshold values may be an idiosyncratic variable.

Given that most people are probably acquainted with a far larger number of dogs and that typicality in the concept dog seems an important factor in judged typicality of instances in complex concepts containing the concept, it is surprising that the correlations were not maximized at set sizes that included at least the 8 true members. It is, in fact, surprising that the strength of the correlations were higher using essentially half of the database, and as few as 3 exemplars as reference set input to the modification model. The implication is that the processor, in conceptual composition, does not tend to take a "broad view" of exemplars in composing new representations, but rather bases the new creation on a rather narrow slice of the domain, in which only a subset of the members of the simple category are considered.

The set identification mechanism does not limit membership to exemplars that are true members of the simple category. Results of this analysis confirm that non-members can be relevant, useful sources of information. Correlations were maximized for 7 subjects when the selected sets included non-members of the category *dog*.

To summarize, response to the charge that the correlational modification model is incomplete in not detailing the mechanisms by which reference sets are determined, one such mechanism has here been described and demonstrated. In response to the charge that such a mechanism must involve access to information external to the set of domain exemplars and prototypes and the modifier-encoded property, the described function makes use of typicality, target feature weight and two parameters in identifying a set of relevant exemplars from a domain. This mechanism has been shown to do work, even under constrained conditions.

Further study needs to be directed at a range of issues raised with respect to this model of a reference set selection mechanism. In particular, the relation between set size and threshold needs to be explored as well as the stability of the α parameter for subjects across types of category.

GENERAL DISCUSSION

Selective Modification reviewed

Evidence from the two experiments support the case for a relation between the set of exemplars considered and value of the diagnosticity weight of the Selective Modification model for predicting typicality of instances in complex concepts. The "variable diagnosticity weight" assumption of the diversity model provides an explanation for the behavior of the diagnosticity weights in Expt. 1 and resolves the results the apparent conflict between of these experiments and those of Smith et al.

This dependence of the model on an artifactual variable is a difficulty that seriously impugns it as a model of a psychological process. Given the dependency of diagnosticity on characteristics of a presented set of exemplars, for example, it is hard to explain how we are able to confidently generate evaluations of single objects in the absence of an explicit exemplar set. Attempts to patch the model by limiting the range of variability of the diagnosticity weight or positing a default diagnosticity weight for use in context-free judgments will not solve other problems, such as the following:

The Selective Modification model, as a psychological model, implies that a change in the composition of the exemplar set considered will have an effect on the judged typicality of an exemplar in a complex concept. This is different from the claim that the <u>relative</u> judged typicality of an exemplar will vary as a function of context. For example, a Great Dane may be judged the best example of *small dog* if the set of exemplars considered was {Great Dane, pony, mule, deer, jaguar}, where it would be a lousy example relative to the set {Great Dane, collie, dachshund, toy poodle, Chihuahua}. Our intuition suggests that *Great Dane* is an equally lousy example of *small dog* in either case. The evaluation of *Great Dane* in *small dog* , per se, doesn't appear to change with respect to the exemplar set considered, but just the <u>relative</u> evaluation. According to the Selective Modification model, the actual strength of the judgment would change with the set of exemplars presented. Since members of the first set vary more on typicality with respect to *dog,* than on *size,* the model would result in greater influence of typicality. *Great Dane,* being a pretty good example of a dog, would be predicted to be pretty typical of the complex concept. Computed typicality for *Great Dane* in the second set would be very low, since much greater influence would be shifted to target property weight (size) among exemplars. This model would lead to the prediction that an observer's evaluation of the typicality of an exemplar would change upon sequential addition of exemplars to the set. While relevant data is not available, it does not seem likely that judgments of typicality are influenced by the composition of presented exemplars to the degree implied.

The variability of the diagnosticity weight in the Selective Modification model is due to its reliance on two factors in the exemplar set, relation of exemplars to the simple concept, and the weight of modifier-encoded attribute. Each factor is related to information associated with one of the constituents of the complex concept. If any model aspiring to the status of a "closed operation" depends on just this information, it is difficult to see how it would be able to avoid the range-dependency of the Selective Modification model. If it cannot avoid the range-dependency, that model, like the dynamic Selective Modification model, would not be closed.

Compositionality

Osherson & Smith (1981) discussed compositionality as a constraint on a model of complex concepts. A challenge of a model, termed the "compositionality problem" (Smith & Osherson, 1988), is to account for the productivity of complex concepts by reference to mechanisms that access no information other than that which is contained in the representations of the constituents of the compound. Various arguments in the preceding pages suggest that models of conceptual combination require information beyond that which is contained in the constituent concept representations.

The problem with the Selective Modification model is it's inability to account for multiple property change for some adjective-noun combinations. The problem with a prestorage correlational model, another candidate for compositionality, is the productivity requirement.

Productivity, as a criterium for a model of conceptual combination, does not mean that a model should account for combinations involving modifiers that will correspond with features present in the representation of the simple noun concept. It means that a model will account for any modifier that can meaningfully be joined with the noun concept. Concepts cannot represent features for every possible modifier, let alone their possible interproperty relations. To restate an earlier argument, if a modifier encodes an attribute that is not associated with the simple concept, and this attribute implies value change for other attributes, a model will require a means of determining new interproperty information, whether through correlational derivation or through some other means. This interproperty information will have to be determined on the basis of information outside of the modifier and noun constituents.

The demise of models based on strict compositionality does not imply that the notions of closed operation and modification must be thrown out, or that we are lead inexorably to embrace exemplar models or causal theories as necessary inputs to a mechanism. A broader notion of compositionality can be adopted which is less restrictive than limiting input to a representation-generating process to those representations of the constituent concepts of the noun phrase. Underlying this new sense is the idea that the output representation of a generative mechanism is distinct from any of the input representations. The correlational modification model, which is not compositional in the strict sense, is compositional in this new sense. Not all possible models are compositional in even this more liberal sense. Exemplar models, for example, that equate concepts with representations of one or a set of (prestored) exemplars are not, in this sense, compositional.

Compositionality, viewed as a process that generates a representation that is genuinely novel, makes no claim with respect to the openness of the generative system. In the narrowest interpretation of "closed operation", no information beyond that contained in the constituent concepts of a compound is permitted as input to a generative mechanism. The concept of system openness, however, can be liberalized to refer to access of limited kinds of information. A system would be considered closed, then, to input of any nonspecified information type. The correlational modification model can be considered a closed system if input to the model is assumed to be restricted to just simple concepts. Were the model found to be influenced by types of information other than those found in or derivable from simple concept representations, it would count as evidence against an explicit assumption of the model. A less constrained model of conceptual combination (Shoben & Medin, 1988) is one in which the generative mechanism is compositional and open "all the way up". The mechanism, in generating complex concept representations, can be guided by access to causal theories embedding interproperty relations, schemas or general world knowledge. A range of models exist between the relatively closed correlational modification model and unlimited access models.

Correlational Modification reviewed

Results of the present experiments provide support for the correlational modification model as a model of psychological processes underlying typicality judgment of instances in complex concepts. The model was able to predict the typicality of exemplars in complex concepts and occurrences of the conjunction effect across conditions of varying task context, and is not dependent on characteristics of the particular set of input exemplars as was the Selective Modification model.

The correlational modification model has a number of virtues. First, the model utilizes simple mechanisms and simply structured concept representations. Like Selective Modification, a mechanism is required to pick out the property in the prototype representation that is encoded by the modifier. Unlike the mechanism of the Selective Modification model, the correlational mechanism does not require that features be explicitly identified with their attribute (e.g. **color** : red, yellow, white, brown, etc.). The correlational selection mechanism requires only that the target feature be identified. Since possession of one feature of a property (e.g. *jungle- dwelling* , as a feature of the property *habitat*), is often negatively associated with possession of other features of the same attribute (e.g. *desert dwelling* as an feature of the attribute *habitat*), the correlational model makes the appropriate adjustment to these related features without reference to any other representational structure. This means that the correlational model does not, as the Selective Modification does, require a structured representation of a concept as input. Features do not need to be explicitly attached to attributes since the model refers only to feature-level information.

A second virtue of the correlational model is that it shows that explicit, prestored linkage of properties is not a necessary feature of the cognitive architecture. Through the online derivation of correlations, relations between features which are implicit in the exemplar representations are made explicit. In the discussion of compositionality, it was argued that pragmatic considerations weigh against prestorage of interproperty relations. The reason was that productivity requires that a model be able to account for any meaningful combination of modifier and noun concept. Since it is not possible to store, for each concept, a feature corresponding to every possible modifier, it is impossible to have relations stored for them. The correlational model shows how interproperty relations for novel modifiers can be derived.

Another virtue of the correlation model is that is shows how interproperty relations can be accessed without appeal to causal theories or higher order knowledge structures that might embed theories. The only input required for the correlational mechanism is the modifier, a prototype, and a set of relevant exemplars. It is, of course, required that the model provide an account of the process whereby the set of exemplars used as input to the correlations is selected.

In addition to the typicality phenomena discussed, the correlational modification model accounts for other results. Medin & Shoben (1988, Expt. 2) show that for a set of related modifiers, the pattern of similarity judgments can change across noun contexts. For example, given the modifiers "gold", "silver" and "brass", subjects view "gold" and "brass" as more similar when the three are combined with "railing", yet judge "gold" and "silver" as more similar when each is combined with "coin". Where either modification model would predict the pattern of judgments for gold and brass "railing" given the overlap of selected color features for *gold* and *brass* , only the correlational modification model accounts for the second pattern in which the modifiers refer to material rather than color. When "gold", "silver" and "brass" are considered as material features of coins, they would likely be associated with properties like "weight" and "value". Correlations among these properties would render the representations for *gold coin* and *silver coin* more similar. While a causal or theoretical story can be generated to support the pattern of judgments, it does not appear that they are necessary for deriving the judgments.

The diversity hypothesis, through directing our attention to the interaction of the major factors influencing the Selective Modification model, suggests that these factors should somehow figure into any model of conceptual combination. After all, when the conditions of the hypothesis were satisfied, the Selective Modification model did an admirable job of accounting for the data. So it is a little mysterious that factors that weigh so heavily in the performance and operation of one model have apparently limited impact on another when both account equally for the data. The correlational modification model showed a stronger average correlation under conditions of greater diversity in the exemplar set, although the difference between correlations for the two experiments was not significant. In analyses using the correlational modification model in which the influence of the target property in the similarity computation was allowed to vary, analogously to the diagnosticity enhancement mechanism of the Selective Modification model, there was virtually no difference in correlations. One interpretation of this result is that the influence of the target property is, through the correlational modification mechanism, translated throughout the entire set of features of the representation. Since the entire set of features reflect the impact of the modifier, it gains nothing to enhance the influence of the single modifier-encoded feature. The influence of the attribute "temperament" for *fierce feline* , for example, does not have to be differentially weighted in similarity since the impact of the modifier is translated through small changes, across the entire set of features.

Conjunction effects

The power and elegance of the Selective Modification model was shown in its ability (Smith, et al., 1988) to not only account for the conjunction effect, but to explain two related typicality effects. To understand the first effect it is necessary to recognize the distinction between compatible and incompatible conjunctions (Smith & Osherson, 1984). <u>Compatible</u> conjunctions are those for which the modifier "denotes a likely value of the object denoted by the noun" (Smith, et al., 1988), such as *red apple*, and <u>incompatible</u> conjunctions, such as *brown apple*, are those for which the modifier "denotes an unlikely value". Smith & Osherson (1984) found a difference in the magnitude of conjunction effect involving these types; the conjunction effect was greater for incompatible conjunctions. Schematically, for this example, where plain text terms designate instances and italicized terms designate concepts, this says:

TYP(brown apple, *brown apple*) - TYP(brown apple, *apple*) >

TYP(red apple, *red apple*) - TYP(red apple, *apple*)

To see how the Selective Modification accounts for this effect, examine first the pair of judgments involving <u>red</u> apples and note that any difference in the two is solely attributable to the color property. This is because the same instance, a red apple, is being compared in both cases, and the weight of features for all non-color properties of *apple* and *red apple* are identical. Overall similarity (excluding color) will be the same, then. Since *apple* is associated with being red, and the red apple is exclusively weighted on *red*, the right hand term (TYP(red apple, *apple*)) will be pretty strong. The computed similarity for the left hand term will be even stronger because all of the color weight will shift to *red* in the modified complex concept representation. This difference, we will guess, is small. Things are a little different for the brown apple case. Again, since both terms compare a brown apple to representations differing only in weights for one property, any difference must be due to the modifier-denoted property. Since *apple* is not associated with the color brown, there will be little similarity on color features in the right hand term. The similarity for these features, thanks to weight shifting to the brown feature in the complex concept, will be much larger. The effect, as explained by the model, trades on the prior weights among modifier-encoded attribute features for the noun concept and the weight shift mechanism.

The correlational modification model predicts a pattern of feature weight change among color attributes for *red apple* and *brown apple* similar to that of the Selective Modification model. This comes about because redness is negatively related to being other colors. In addition, modification guided by correlation predicts changes in feature values for non-color features. *Brown apple* might be associated with having a rough, ruddy surface, being not very round, and being mushy to the touch. The correlational model would make these adjustments in the new complex concept representation. Representations of good examples of red apples and brown apples would resemble their conceptual counterparts more than would be predicted by the Selective Modification model. The similarity of good instances in their respective complex concept would probably therefore be about the same. If both left hand terms are the same, the difference must reside in the right hand terms. If one considers that both *apple* and *red apple* tend to be associated with being red in color, smooth, firm and round, the effect falls out of the difference in similarity for the two instances in apple. Brown apples are just awful examples of apples, and not just because brown is an unlikely value on color for *apple* . Along with being brown go other properties that also associated with unlikely values for *apple*.

Besides providing an explanation for this conjunction effect, it should be noted that since no mechanism is involved that requires knowing which features are attached to a given attribute like the weight shift mechanism, and since there are no diagnosticity weights, there is no need for a structured concept representation. Correlational modification operates on a simple list of weighted features. Another dividend of the correlational modification model is being able to work with far simpler representational structures.

The last typicality effect involving complex concepts discussed by Smith & Osherson, 1984 is called the reverse conjunction effect. It was shown that the judged typicality of instances with <u>no</u> association with the feature denoted by the modifier tends to be less than even incompatible instances. Such an exemplar-property pair is *brown apple,* since apples are not normally associated with the color brown. So, for example-

TYP(brown apple, *apple*) > TYP(brown apple, *red apple*)

As described in the previous example, correlational modification, in generating red apple, would diminish the property weight of non-red color features, as would the weight shift mechanism of the Selective Modification model. Rather than relying, though, on the enhanced color diagnosticity weight to dilute the similarity of the right hand term, the correlational account would predict that adjusted values for non-color properties related with redness tend, overall to resemble less the representation of the brown apple. The correlational modification model, then, is able to provide a coherent account of these typicality effects.

Adverbial modification

In addition to demonstrating support for the Selective Modification model in predicting typicality of instances in adjective-noun compounds, Smith et al. provided evidence in support of model's claim to account for adverbial modification. The adverbs in this case are adjectival modifiers (e.g. very red apple). The mechanism by which the adverb impacts the modification process is through diagnosticity enhancement. It is argued that adverbs act primarily to draw attention to the property of the object denoted by the adjectival modifier, and that the proper translation of this intent, in the model, is through manipulation of diagnosticity rather than indices of property salience. Adverbs such as *very* or *slightly*, then, impact the process by modulating the extent of change of the diagnosticity weight of the adjective-selected property.

The current version of the correlational modification model has no mechanism for implementing change in response to adverbial modification. This is because there are no diagnosticity weights, per se, and therefore no mechanisms of diagnosticity enhancement, and because the proposed mechanism adjusts the target feature weight to a maximum value already.

One implementational strategy is to adopt distinct diagnosticity weights and a mechanism that modulates the value of those weights associated with the target feature. If the argument that the entire set of feature weights bear the impression of the target property carries any force, then a mechanism that translates the impact of the adverb across all properties seems a better strategy. The intuition here is that enhancement of the adjectival modifier increases the salience of the target property, but also has an impact on related properties. Godzilla, for example, is more than a very large lizard; he (?) is a proportionately more aggressive and destructive lizard. This hypothesis, that adverbial modification has extended impact, can be examined empirically. Failure to find support for the hypothesis would be indirect evidence for some scheme involving diagnosticity weights.

Computational costs

The correlational modification model involves a far greater commitment of cognitive resources than the Selective Modification model. Measures of interproperty association must either be stored in memory as part of the concept representation, or generated online. It was argued in the last chapter that given the broad requirements of the productivity characteristic of complex concepts, interproperty correlations are most likely generated online. Evidence was also provided, in the context of the reference set identification model, that interproperty correlation may be based on a relatively small number of exemplars. Utilization of a small number of exemplars may reflect the relative high level of cognitive costs entailed by reference set construction and interproperty correlation. Reference set size may also be a function of the likelihood that the modifier denoted by the adjective is correlated with other properties. It was noted that many properties are significantly related to few or no other properties among exemplars of the reference set. In cost-benefit terms, use of larger set sizes may not be worth the trouble if in most cases the correlational mechanism has marginal impact in the generation of a new representation. There may be other reasons why it is more efficient or useful in predictive value to base modification on correlations derived from a small set. It has been suggested that some reasoning tasks may take advantage of higher order relations among properties (Wattenmaker, Dewey, Murphy & Medin, 1986, Medin, Wattenmaker & Hampson, 1987). Among birds, for example, the properties "wing length", "talon size" and "tendency to eat meat" may covary, so that knowing a value for one would predict values on the others. Small sets constructed by use of the described mechanism may be used to extract higher order relations among properties.

Success of the reference set identification model in improving the fit of the correlational model to obtained typicality judgment, combined with the finding that the improvement was associated with small set sizes, leads to the possibility that sets of relevant exemplars can be exploited, without mediation of a correlational process, in the generation of a complex concept representation. If some type of set identification mechanism is required for conceptual combination, and if such a mechanism can reliably pick out a small set of highly relevant exemplars, then it could be the case that judgment of typicality is a relation of an exemplar and the reference set. Typicality of an exemplar, given a reference set, might be the greatest (maximum) similarity of the exemplar to any set member, or the similarity of the exemplar to an average of set members. Some examples suggest problems for such models.

Suppose, for example, we were using the reference set to identify a set of exemplars relevant to the judgment "How typical is a terrier of a burrowing dog?", and the set identified was { dog, fox, weasel, hamster, gopher }. A problem with equating typicality with the maximum similarity of an instance to a set is that instances that are not very typical of the simple concept may be very similar to some member of the set. So where terrier is most similar to fox, beavers are very similar to gophers, and they (beavers) are not good examples of *burrowing dog.*

The other possibility mentioned was equating the prototype for the complex concept with the average of the set members. The following example presents problems for this formulation. For the judgment "How typical is a bluejay of an arctic sparrow?", our set identification mechanism might generate the set {sparrow, penguin, seal, polar bear }. Clearly we don't mean by *arctic sparrow* the average of these animals. An average of the set exemplars, across features would violate property values that are thought to be essential to sparrowhood, such as having beaks and wings. The correlational model survive an example like this since the only features that would be associated among the set of exemplars are non-essential properties associated with arctic habitats.

Multiple processes

In the course of building a case for the correlational modification model, a range of alternative models of conceptual combination have been considered, from causal theoretic models to exemplar based models. In some cases these alternative models were set aside for lack of explicit computational expressions which to test. Other models were shown to be unable to account for our intuitions with respect to particular types of examples. Examination of the broad range of types of examples that count as instances of just adjective-noun compounds, particularly with regard for the variable uses in language of such expressions, suggests that a complete account will comprise multiple processes.

It is clear from inspection of a number of examples that conceptual combination sometimes requires causal reasoning. An example using a familiar noun concept is *lip-reading dog*. While a set identification mechanism such as the one described earlier may be able to generate a set of dog-like objects and lip reading people, an understanding of the concept seems to demand an account of the circumstances that would have led to a non-human animal coming to acquire an un-doglike behavior. A causal story linking the lip-reading ability to its instantiation in a canine allows us to deduce characteristics about the animal's intelligence, probable relationship with humans, etc. that would not be available from interproperty association among the set of relevant exemplars.

In contrast to the *lip-reading dog* example is the example of *brown dog,* in which the modifier encodes a feature that it not likely to be associated with other features. The set identification model may easily find a set of exemplars that are doglike and tend to be brown. The representation which is the outcome of the correlational process, if no other (non-color) features are correlated with the target feature, will be essentially identical to one generated by the Selective Modification model, since no features besides "color" ones will have been affected. Correlational modification, while not producing a representation which is flawed in any manner, has gone to significant computational expense to accomplish what could have been achieved very simply.

An approach to a multiple process account would be to distinguish cases of complex concepts on the bases of interproperty correlationss and associations of the modifier-encoded property available among a set of relevant exemplars. If the feature is associated with a number of the set exemplars, but is not correlated with other features (e.g. *brown dog*), then a computationally simple model may account for our judgments. If the modifier-encoded property is not associated with many of the exemplars in the set (e.g. for *lip-reading dog,* dog exemplars are negligibly associated with the property), then a process involving accessand exploitation of causal theory would be recruited. In the range of cases in which exemplars are associated with the modifier-encoded feature and where these properties are correlated with other properties, a process like the correlational modification model may be recruited.

This proposal assumes, speculatively, that any path chosen to generate the complex concept representation does so via inspection of a set of exemplar representations. If this assumption is true, then it may be possible to experimentally distinguish these cases through differences in the processing times required for generating representations via each path.

SUMMARY

According to prototype theory, in composing concepts, the observer exploits the tendency for natural objects to be packaged in clusters of features that are often correlated. Evidence has been presented that supports the view that the correlational structure underlying the properties of at least some natural objects may be exploited in the composition of complex concepts. The findings therefore support the claim that changes in the value of one property in a concept representation, are often accompanied by changes in others (Medin &

Shoben, 1988). The presented model of correlational modification shows that composition of a complex concept representation can require nothing more than a set of exemplars and simple concept prototype, a correlation function, and a function for adjusting feature weights. Further work should attempt to extend the capabilities of models using independent exemplars and simple modification mechanisms.

REFERENCES

Battig, W.F. & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80, Part 2, 1-46.

Cohen, B. & Murphy, G.L. (1984). Models of concepts. Cognitive Science, 8, 27-60.

Gregson, R.A.M. (1975). Psychometrics of similarity. New York: Academic Press.

Hampton, J.A. (1987). Inheritance of attributes in natural concept categories. Memory & Cognition, 15, 55-71.

Malt, B.C. & Smith, E.E. (1984). Correlated properties in natural categories. Journal of Verbal Learning and Verbal Behavior, 23, 250-2x9.

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

Medin, D.L. & Schaffer, M.M. (1978). A context theory of classification learning. Psychological Review, 85, 207-238.

Medin, D.L. & Shoben, E.J. (1988). Context and structure in conceptual combination. Cognitive Psychology, 20, 158-190.

Medin, D.L., Wattenmaker, W.D. & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness & category construction. Cognitive Psychology, 19, 242-279.

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

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill.

Murphy, G.L. (1988). Comprehending complex concepts. Cognitive Science, 12, 529-562.

Murphy, G.L. & Medin, D.L. (1985). The role of theories in conceptual coherence. Psychological Review, 92, 289-316.

Murphy, G.L. & Wisniewski, E.J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), Advances in cognitive science, Vol. 2: Theory and Application. Chichester : Ellis Horwood.

Osherson, D.N. & Smith, E.E. (1981). On the adequacy of prototype theory. Cognition, 9, 35-58.

Osherson, D.N. & Smith, E.E. (1982). Gradedness and conceptual combination. Cognition, 12, 299-318.

Rips, L.J. (1988). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony (Ed.s), Similarity, analogy and thought. NY: Cambridge University Press.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), Cognitive development and acquisition of language (pp. 111-144). New York: Academic Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Ed.s), Cognition and categorization. Hillsdale, NJ: Erlbaum.

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

Rosch, E., Mervis, C., Gray, D., Johnson, D. & Boyes-Braehm, P. (1976). Basic objects in natural categories. Cognitive Psychology, 3, 382-439.

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

Smith, E.E. & Osherson, D.N. (1984). Conceptual combination with prototype concepts. Cognitive Science, 8, 337-361.

Smith, E.E. & Osherson, D.N. (1988). Compositionality and typicality. In S. Schiffer & S. Steele (Ed.s). Cognition and Representation. Boulder, CO: Westview Press, 37-52.

Smith, E.E., Osherson, D.N., Rips, L.J. & Keane, M. (1988). Combining prototypes: A Selective Modification Model. Cognitive Science, 12, 485-527.

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

Tversky, A. & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic & A. Tversky (Ed.s), Judgment under uncertainty: Heuritics and biases. New York: Cambridge University Press.

Wattenmaker, W.D., Dewey, G.I., Murphy, T.D. & Medin, D.L. (1986). Linear separability and concept learning: Context, relational properties and concept naturalness. Cognitive Psychology, 18, 158-194.

Table 1

List of features and attributes

Attribute	Feature
Nose/muzzle size	large nose/muzzle, small nose/muzzle
	ear size
Paw size	large paw, small paw
Tail length	long tail, short tail
Overall dimension	height (overall), length (overall)
Overall shape	lean, bulbous
Hair length	long hair, short hair
Perceptual Acuity	visual acuity, auditory acuity, olfactory acuity
Social unit	solitary, family, pack
Color	black, brown, tawny/yellow, red/orange
Markings	patches, spots, stripes
Habitat	forest, mountains, jungle, plains
Diet	large game, small game, canned/dry food
Activity level	active, inactive
Speed	fast, slow
Strength	strong, weak
	noisiness
	cleanliness
	intelligence
	territorial
	predatory
	domesticity
Temperament	ferocity, timidity, excitability, friendliness

Table 2

Interproperty Correlation Frequency from Normalized Average Matrix

	Level of significance	
Concept	$p < .05$	$p < .01$
Canine	332 (28.5%)	169 (14.5%)
Feline	427 (36.7%)	269 (23.1%)

Note. Total property pairs numbered 1164.

Table 3

<u>Sample feature weights showing a relationship of deviation and similarity</u>

	Feature						Sum
Prototype	1	2	3	4	5	6	
Directly Rated	5	5	5	5	5	5	
Modified	2	2	2	2	2	2	
Deviation	3	3	3	3	3	3	18
Minimum	2	2	2	2	2	2	12
Maximum	5	5	5	5	5	5	30

Similarity (Min./Max) = .40

Table 4

Sample feature weights showing a relationship of deviation and similarity

	Feature						Sum
Prototype	1	2	3	4	5	6	
Directly Rated	5	5	5	5	5	5	
Modified	8	8	8	8	8	8	
Deviation	3	3	3	3	3	3	18
Minimum	5	5	5	5	5	5	30
Maximum	8	8	8	8	8	8	48

Similarity (Min./Max) = .625

Table 5

Sample feature weights showing a relationship of deviation and similarity

	Feature						Sum
Prototype	1	2	3	4	5	6	
Directly Rated	6	6	6	6	6	6	
Modified	9	10	10	9	9	9	
Deviation	3	4	4	3	3	3	20
Minimum	6	6	6	6	6	6	36
Maximum	9	10	10	9	9	9	56

Similarity (Min./Max) = .643

Table 6

Sample feature weights showing a relationship of deviation and similarity

	Feature						Sum
Prototype	1	2	3	4	5	6	
Directly Rated	2	2	2	2	2	2	
Modified	0	0	0	0	4	0	
Deviation	2	2	2	2	2	2	12
Minimum	0	0	0	0	2	0	2
Maximum	2	2	2	2	4	2	14

Similarity (Min./Max) = .143

Table 7

List of modifiers and number of subjects who were assigned each : Experiment 2

Modifier	N. subjects
tall	1
long	0
lean	0
bulbous	0
ferocious	4
timid	1
excitable	0
friendly	0
active	1
fast	2
slow	0
strong	0
weak	1
noisy	2
intelligent	1

- Ex. Typicality of apple in red fruit-

  - A representation for "fruit" is accessed

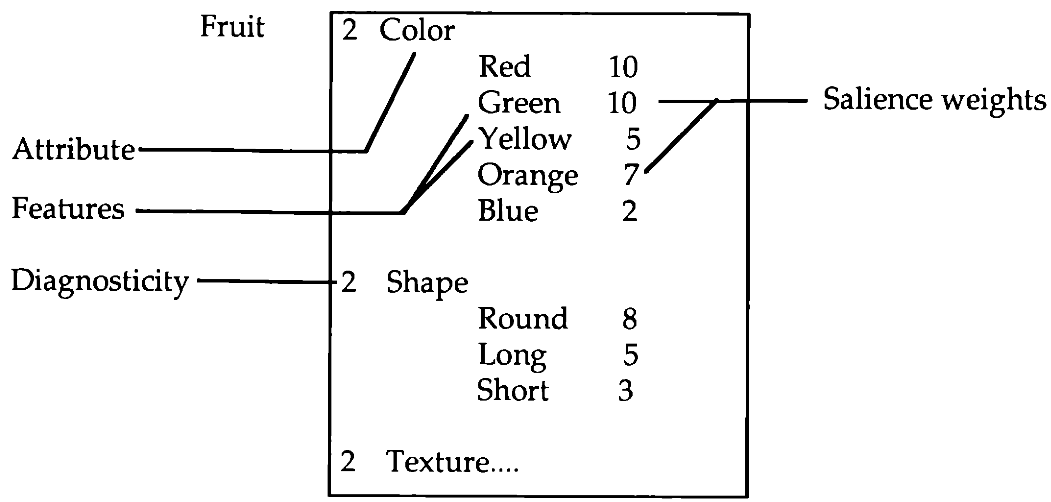


Figure 1. Selective Modification Model - Representation

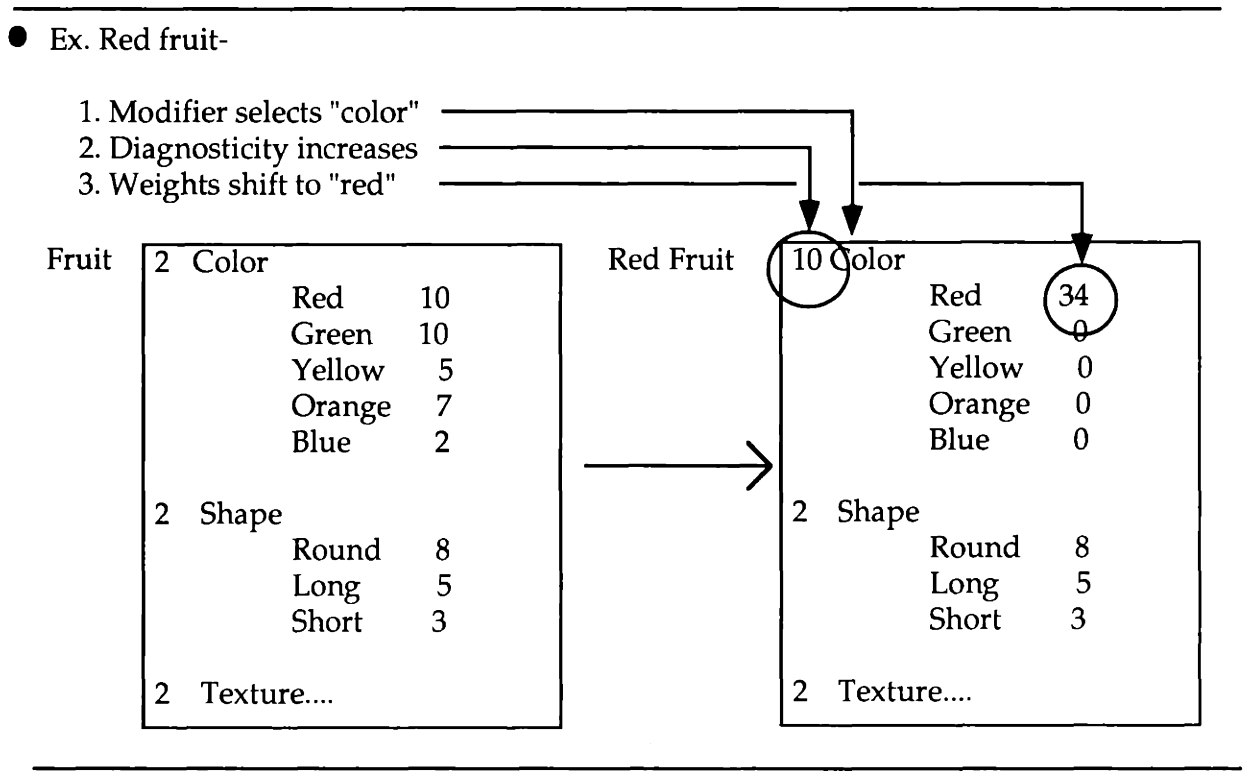
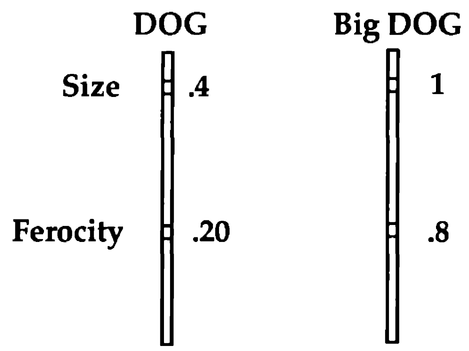


Figure 2. Selective Modification Model mechanisms



1. Target attribute increases to maximum.

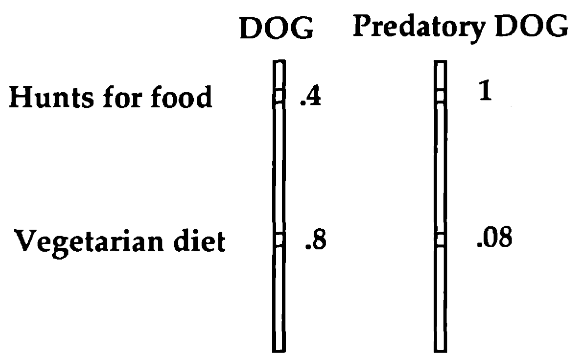
For positive correlations-

2. For size, ferocity, correlated across exemplars, r=.75

P2'= P2+ (1 - P2) * Corr(P1,P2)

.8 = .20 + (1 - .20) * .75 = .2 + .6

For positive correlations, new feature weight changes as a proportion of the difference of its original weight and 1.



For negative correlations-

2. For size, ferocity, correlated across exemplars, r=-.9

P2'= P2 - (P2 * Corr(P1,P2))

.08 = .8 - (.8 * -.9) = .8 - .72

For negative correlations, new feature weight changes as a proportion of the difference of its original weight and 0.

Figure 3. Mechanisms of correlational modification