JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 18, 441 461 (1979)

Polymorphous Concepts in Semantic Memory

JAMES A. HAMPTON

The City University, London

Two experiments tested a set of predictions derived from Smith, Shoben, and Rips' (1974) Characteristic Feature Hypothesis and two-stage categorization model. Experiment I elicited category definitions from subjects, and Experiment II measured categorization latencies. Neither a strong version of Smith et al.'s theory, which assumes that defining features are common to all category members, nor a weaker version, were well supported by the results. It is argued that an alternative feature-based model of category definitions, using the notion of a Polymorphous concept, can account for the results without making the distinction between defining and characteristic features, on which Smith et al.'s model relies.

Much interest in recent years has been given to the question of how noun categories such as *fruit* or *bird* are defined. Two main classes of model have been described in which category information is either (a) contained in a network linking different words or "concept nodes" (e.g., Collins & Quillian, 1972) or (b) encoded by a system of features or meaning components (also called attributes or properties). An influential model of the latter type, which attempted to create a quantifiable feature-based model of categories, was proposed by Smith, Shoben, and Rips (1974). The model is based on what they term the Characteristic Feature Hypothesis. This hypothesis makes a distinction between two types of features in the defintion of a category. *Defining features* provide the necessary and sufficient criteria for deciding whether any word is a category member; whereas the *characteristic features* only determine the typicality or centrality of category members, and the semantic relatedness of nonmembers. The distinction of the two types of feature is made by setting a cut-off point on a continuum of *definingness* along which the features are assumed to be ordered.

The two types of feature are further distinguished by the role they play in the model of categorization decisions proposed by Smith et al. (1974). The model has two stages. The first stage makes an overall "holistic" appraisal of the degree of feature overlap between the category and the item to be categorized. If a sufficient degree of match or mismatch is found, then a rapid Yes or No decision can be made. All the features, characteristic and defining, are involved in this first stage. If an intermediate degree of overlap is found in the first stage, then a second stage of processing proceeds to check only the defining features of the category against those of the instance, thus producing an accurate, but slower judgement of category membership.

The research described in this paper is concerned with feature-based models of category structure. It was intended that by eliciting from subjects the feature definitions of eight categories, a specific test could be made of Smith et al.'s (1974) model. It is

This research was supported by a postgraduate studentship awarded to the author at University College London by the Medical Research Council of Great Britain. The author wishes to thank Dr. David Green for his help and guidance at all stages of the research, and Herbert Clark, Peter Fonagy, John Gardiner, Stephen Lea, and Ed Smith for their helpful comments on earlier drafts of this paper. A shorter version of the paper was given at the Experimental Psychology Society conference at Bedford College, London, in January 1978. Requests for reprints should be sent to James A. Hampton, Department of Psychology, The City University, St. John Street, London, EC1V 4PB, England.

0022-5371/79/040441-21$02.00/0

Printed in Great Britain

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

one hour. The session was divided into two parts. In the first part, subjects were asked to give descriptions of eight noun categories. For each of the categories a set of seven different questions was used in order to encourage subjects to generate as many different properties as they could. Thus, for example, phrases such as "loosely speaking" and "technically speaking" (known as hedges (Lakoff, 1972)), were used in order to ask a subject to consider why some items might be only loosely speaking a kind of Furniture, and in particular what properties of an item would make it belong to the category or would exclude it from the category. Other questions required subjects to consider why items should be typical category members or borderline cases. In all questions involving the consideration of particular examples, in order to avoid biasing the subject, no particular examples were ever given by the experimenter. Thus, the subject would be required to produce his own examples. In the second half of the session, the subject was presented with a list of 30 words for each of the eight categories. The experimenter read out the name of one of the categories, followed by the list of words for that category. To each word, the subject made a vocal response according to a seven-point response scale, ranging from 1 for definite category members, through 4 for "unable to decide" or "don't know," to 7 for totally unrelated words. The seven possible responses were written out on a card which the subject kept in front of him while making the decisions.

In the first part of the experiment, the order of presentation of categories was balanced across subjects, so that each category occurred the same number of times in each ordinal position, and also followed each of the other categories the same number of times. For the second part, the subjects received the same order of presentation of the categories as previously, and the order of the words in each list was randomized, but was the same for each subject.

Materials. Eight category names were selected from the list used by Battig & Montague (1969) for collecting their norms of category members. Three informal criteria were used by the experimenter to guide the selection of categories:

(1) Members of the category could be similar to one another in a number of different ways, without there being an obvious property that they all had in common. (Wittgenstein referred to this characteristic as showing "family resemblances" (1953, pp. 31–34)).

(2) There appeared to be degrees of typicality among the category members, and there were other words which although outside the category, were still related to it in meaning.

(3) There were some cases which appeared to be doubtful or borderline members of the category.

The eight categories were *Kitchen Utensil, Furniture, Vehicle, Sport, Fruit, Vegetable, Fish,* and *Bird.* Most of the words for the categorization rating task in the second part of the experiment were selected from the category norms provided by Battig & Montague (1969). A wide range of typicality was aimed for, and several nonmembers of the categories were included, selected from the most infrequent responses to the category in the category norms. In some obvious cases, a translation from American into British terminology was made (for instance, *can-opener* to *tin-opener*, and *sidewalk* to *pavement*).

Results and discussion

The results are presented and discussed in four sections. The first section describes how a final list of the main features of each category was produced from the properties given as definitions of the categories. The second section describes the construction of a scale of rated membership for each list of category words. The third section examines in detail the relation between the features of each category and the rated category membership of the lists of words, and provides a test of the four

HAMPTON, JAMES A., **Polymorphous Concepts in Semantic Memory** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

predictions derived from Smith et al.'s (1974) model (using the strong version for the identification of defining features). A final section discusses the implications of the findings and considers them in the light of an alternative feature-based model.

(1) *Derivation of the category features.* Since all subjects received the same seven questions about each category in the same order, no use could be made of the information of which properties were produced in response to which questions. Subjects' responses were therefore pooled over the seven questions for each category. By pooling responses across subjects, a large list of properties was obtained for each category, ordered according to the frequency with which they were given. In order to keep the experiment within manageable limits, it was necessary to reduce these lists to a reasonable size by excluding the least important properties. One obvious way to achieve this is to use the production frequency as a measure of the importance or salience of a property for defining a category. As a check on this procedure, an independent group of 16 judges was asked to rate the importance of each category property given by more than four of the 32 subjects. These properties were written out in alphabetical order, and the judges placed them in a rank order, according to how important they felt each property was for defining the category in question. There were between 13 and 22 properties to be ranked for each category. There was significant agreement between judges on the rankings (Kendall's coefficient of concordance, W, had values between .21 and .58 in all cases $p < .001$, Siegel 1956). The rankings were therefore averaged to give a mean rank for each property. Kendall's rank correlation coefficient tau (Siegel, 1956) between the mean ranked importance and the production frequency of each property was computed for each list of properties. All the correlations were in the predicted direction (mean tau $= 0.375$), with high frequency properties being ranked as more important, and six of the eight values were significant beyond the .05 level. It may therefore be concluded that production frequency may be used as a measure of the importance of a property for defining a category, thus providing evidence for the major assumption underlying the method adopted in this experiment. The results also provide support for Smith et al.'s (1974) prediction that the category features may be ordered according to some underlying variable called *definingness.*

The lists of properties for each category were therefore reduced by excluding those properties which had low scores on both measures. Properties given by fewer than eight of the 32 subjects which were also ranked in the less important half of the list of properties ranked, were excluded, leaving a final short list of properties which was then taken for the purposes of the experiment as the list of features for the category. There were between 8 and 16 features for each category in the final short list. They are shown in the Appendix.

(2) *Derivation of a scale of rated membership.* The ratings given to the lists of words in the second part of the experiment were summed over subjects, and linearly transformed to fit on to a scale ranging from 100, corresponding to unanimous agreement that a word is definitely a category member, to 0, indicating complete agreement that a word is definitely not a category member. A value of 50 corresponded to a response of "unable to decide," which was the response "4" on the 7-point rating scale. To justify summing over subjects, the subjects were divided into two groups at random in three different ways, and the correlation between the mean ratings for each group were calculated for each category. Spearman's rank correlation coefficient r_s (Siegel, 1956), had values $.67 < r_s < .96$, with a median of .94, $p < .001$, demonstrating a high degree of agreement between subjects.

Two other independent measures of degree of category membership were compared with the scale produced in the present experiment. Category Production Frequency taken from Battig & Montague (1969), which is the

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

frequency with which a word is produced as a member of the category, correlated with the scale significantly in all categories. Kendall's tau had values between .39 and .69, $p < .005$. The other measure was a rating of Typicality taken from published data from Rosch (1975), who used six of the eight categories included in the present experiment. In this case, correlations were higher ($.65 < \text{tau} < .80$, $p < .001$) indicating the similar nature of the two tasks of rating membership and rating typicality of members.

(3) *Testing the category definitions*. In the two previous sections, category features were operationally defined as the properties fulfilling certain criteria of production frequency and rated importance, and a scale of rated membership was constructed for the eight categories. It is therefore now possible to compare the two, and analyze the relation between them. In particular, it is possible to examine the complete set of features in order to identify those which are defining and those which are characteristic in Smith et al.'s (1974) terms. Adopting the strong version of their model, a defining feature is one which is common to all category members. The defining features are therefore those which are possessed by all the words with a rated membership of greater than 50 on the scale. (Additional evidence for placing the category boundary at 50 comes from the data reported in Experiment II. In that experiment a selection of the words rated in Experiment I were presented with others to subjects for a forced-choice Yes–No categorization decision. A comparison of rated membership with the number of Yes responses a word received showed a strong linear relationship ($r = 0.96$), with 50% Yes responses corresponding to a rated membership of 50.27 on the scale. The scale position 50 is therefore the most justifiable position for defining the category boundary. The effects of allowing the boundary position to vary for each category are discussed in a later section).

Characteristic features are those which determine typicality of membership. If we make an arbitrary division of category members such that typical members are those with a rated membership greater than 75, then the characteristic features are identifiable as those possessed by more typical members than atypical ones. (The exact level at which the criterion was placed is arbitrary, but small changes in the level did not significantly affect the choice of the characteristic features). In a similar way the range of nonmembers may be arbitrarily divided into related words (with rated membership between 25 and 50) and unrelated words (with rated membership less than 25), so that the features associated with the relatedness of nonmembers may be identified.

In order to make this analysis, it was necessary to decide which of the words possess which of the category features. To this end, four postgraduate students acted as judges, and decided for each word-feature pair whether the object named possessed the property or not. A five point scale was used from $+2$ = definitely yes, through 0 = uncertain, to -2 = definitely no. Judgments were summed across judges to give a score of between $+8$ and -8. For the categories of FISH and BIRD, only three judges were used. In addition to the 30 words used in Experiment I, each judge assigned scores to between 11 and 17 additional words which were used in Experiment II, described later.

In order to improve the accuracy of the analysis, these additional words used in Experiment II, were added to the lists of 30 words used for Experiment 1, by assigning them to one of the four quarters of the membership scale on the basis of the number of Yes responses they received in the categorization task used in that experiment. (The close relationship between the number of Yes responses and the rated-membership of words that were used in both experiments has already been mentioned). For each category there were therefore in all between 41 and 47 words distributed across the four quarters of

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

the membership scale and for each of these words, the degree to which they possessed each of the category features had been estimated by the judges.

(a) *Identifying the defining features.* The defining features were operationally defined as those with a positive feature score for all the words with a rated membership of greater than 50 (or with more than 50% Yes responses in Experiment II). Between two and seven defining features were identified for each category. These features are indicated in the Appendix by use of the Label D. For example, the category VEGETABLE had three defining features—"EDIBLE," "PLANT," and "IS CULTIVATED." Having thus chosen the defining features, we can now test the first prediction of Smith et al.'s (1974) model, namely that taken in conjunction, the defining features are not only necessary but also sufficient for category membership. Table 1 shows the words in each list which, while possessing all the defining features of the category, were not rated as members. It may be seen that in six of the eight categories, the conjunction of defining features is far from sufficient as a definition of category membership. Only for VEHICLES and BIRDS is there an adequate set of defining features, such that no nonmembers possess them all. The data do not therefore lend support to a strong version of Smith et al.'s (1974) model.

The fit of the conjunctive definitions could be improved in two ways. First, if additional "nearly necessary" features are added to the definition, on the condition that the number of additional nonmembers thereby excluded exceeds the number of category members which are also excluded, then an optimal solution with the minimum number of misplaced items can be obtained. This manipulation may be more in keeping with the weaker version of Smith et al.'s theory, in which defining features may be common to a large majority but not quite all category members. By this technique, the number of misplaced items drops from a

TABLE 1

WORDS POSSESSING ALL DEFINING FEATURES WITH RATED MEMBERSHIP LESS THAN 50.0 OR PERCENTAGE OF YES RESPONSES LESS THAN 50%[a]

Category	Words in Experiment I	Words in Experiment II
Kitchen Utensil	Thermometer (41.1), table (31.2), placemat (28.6)	Apron (44)
Furniture	Radio (49.5), ashtray (44.3), chairleg (44.1), fridge (42.2), trunk (39.1), sewing-machine (29.7), door (25.5), deskblotter (23.4), ladder (11.5), van (1.6)	Blackboard (37), swing (12), dustbin (6), pen (0), house (0)
Vehicle	—	—
Sport	Bar-football (44.2), tiddlywinks (42.7), camping (36.5), juggling (31.8), dancing (28.6), singing (11.4)	Yoga (25)
Fruit	Gherkin (27.6), marrow (23.4), cucumber (22.4), garlic (11.9), onion (9.4), mushroom (4.7)	Walnut (25), aubergine (19), peppercorn (6), pea (6), cauliflower (6), swede (0), potato (0)
Vegetable	Tomato (49.5), mango (48.2), rice (42.7), rhubarb (40.6), avocado (29.2), olive (28.6)	Tangerine (0), gooseberry (0)
Fish	Shrimp (33.3), lobster (24.5), tadpole (20.3)	—
Bird		—

[a] Values shown in parentheses represent rated membership for words in Experiment I and % Yes responses for words used only in Experiment II.

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal
Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

mean of 6.25 per category to an optimum level of 3.0 items per category. However, consideration of the particular items misplaced suggests that this lack of fit is still serious for the model. For example, the failure to exclude WALNUT from the category of FRUIT, carries the implication that other kinds of NUT would be equally misplaced if tested against the definition.

The alternative manipulation to improve the fit is to allow the scale position of the category boundary to deviate from 50. If a narrower definition of membership is employed, then more common features can be found, and so a better feature definition obtained. By this means, the mean number of misplaced items drops to 2.25 per category. However, the boundary level must be allowed to rise above 70 on the scale to achieve this degree of improvement. Such a wide deviation may be hard to justify.

(b) *Identifying the characteristic features.* In order to identify the characteristic features, the members of each category were divided into Typical members (with a rated membership of greater than 75) and Atypical members (with a rated membership between 50 and 75). Characteristic features were defined as those having significantly more words with positive scores belonging to the Typical end of the scale, using a Fisher Exact test with the 5% significance level as the criterion (Siegel, 1956). The features selected by this method are indicated in the Appendix by the label "C." For instance, characteristic features of VEHICLES are that they "HAVE WHEELS," and of BIRDS that they "SING." It can be seen in the Appendix that all eight categories had identifiable characteristic features, the number per category varying between two and six. The second prediction of Smith et al.'s (1974) model was therefore supported. The third prediction was that the characteristic features should be lower on some underlying dimension of definingness than the defining features. This prediction was tested using two different measures of the definingness of the features. First it was found that the defining features had a significantly higher production frequency than the characteristic features (a median of 24.5 compared with a median of 16.5, $p<.05$ on a Mann-Whitney U test). Second, in all eight categories the defining features were rated on average as more important than the characteristic features in the rating task described in Section (1) above ($p<.01$, sign test). Both predictions concerning the characteristic features were therefore supported.

However, an effect not predicted by the Characteristic Feature Hypothesis was also found. The mean feature scores of typical category members were significantly higher than those for atypical members even when just the defining features of the category are considered ($p<.01$, sign test). Thus typicality of a category member is not just a function of the *number* of characteristic features it possesses but is also associated with the *degree* to which it possesses the defining features. This result throws further doubt on the usefulness of making the distinction between the two kinds of feature. Thus an important element of the distinction is the notion that category membership and the typicality of members are determined by two different sets of features. The finding that defining features (as they have been operationally defined) are on the one hand not always adequate for defining category membership, and on the other hand are themselves associated with the typicality of members, is evidence against this part of the distinction.

(c) *Doubtful cases.* The fourth prediction derived from the model is that there should be no well-known words which do not have a clear-cut category membership. However, in every category except for BIRDS, the rated membership scale for the words used appeared to form a continuum with several words falling near the category borderline. (For the category BIRDS, the word TITMOUSE was on the borderline because of uncertainty about whether it is a mouse or a

HAMPTON, JAMES A., **Polymorphous Concepts in Semantic Memory** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

TABLE 2
BORDERLINE CASES OF CATEGORY MEMBERSHIP

Category	Just included	Just excluded
Kitchen utensil	Sink	Dustpan, sponge, apron*, thermometer
Furniture	Hammock, vase, picture, curtains, lights	Radio, ashtray, chairleg, fridge
Vehicle	Skates, surfboard, wheelbarrow, roller-skates	Crane*
Sport	Darts, skipping	Chess, bar-football, tiddleywinks
Fruit	Rhubarb, yam	Gourd
Vegetable	Seaweed	Tomato, gourd, mango, rice, rhubarb
Fish	Crayfish, ray*	Octopus, starfish
Bird	Titmouse	—

* Word used only in Experiment II.

tit. It is, in fact, the latter). Table 2 shows the words with a rated membership of between 40 and 60 (or the equivalent percentage of Yes responses for words used only in Experiment II). Thus the prediction of clear cut category boundaries derived no support from the data.

(d) *The relatedness of nonmembers.* A final issue concerns how a nonmember comes to be considered related to the category. By dividing the words which are not category members into related nonmembers (with a rated membership of between 25 and 50) and unrelated nonmembers (with a rated membership of less than 25), it is possible to see which of the features are associated with the related words. (As before, the exact level of the criterion is arbitrary but not critical to the selection of the features). Since there is an indefinitely large number of words which are unrelated to a category, the strategy adopted was to select those features which had more positive than negative feature scores within the class of related nonmembers. Six of the categories had a sufficient number of related nonmembers in the lists of words used to allow the features associated with relatedness to be selected. These features are indicated in the Appendix. In all, 35 features associated with relatedness were identified. Of these, 22 were also identified as defining features of the categories, four were characteristic, and nine were neither defining nor characteristic. These figures correspond to 76% of all defining features, 20% of all characteristic features, and 35% of the remaining features for those six categories. Therefore the chief determinant of the relatedness of nonmembers to the category is the number of defining features possessed by a word. This result is in agreement with notions of semantic relatedness that commonly define relatedness as common membership of some superordinate category (e.g., Collins & Quillian, 1969).

(4) *The status of the characteristic feature hypothesis.* The model proposed by Smith et al. (1974) is supported by the present data in the following respects:

(a) The features of a category can be placed on a continuum of definingness such that those produced most frequently are also rated as more important for defining the category.

(b) Features fitting the description of characteristic features, being associated with the Typicality of category members, were identified in all eight categories. They also had a lower production frequency and a lower rated importance than features which were necessary for category membership.

However, the following results are inconsistent with the model in its strong form:

(c) In six of the categories considered, there was no set of defining features which taken in conjunction were sufficient for category membership.

(d) It was found that the Typicality of members was also associated with the degree to which they possessed the features which were common to all members (i.e., the defining features).

(e) In seven of the categories, there were clear examples of words falling close to the border of the category, showing that the category boundaries are not clear cut.

This last result confirms findings reported by McCloskey and Glucksberg (1978) that there is considerable disagreement between subjects and even inconsistency within subjects concerning items situated close to the border of a category (as determined by rated typicality).

The pattern of results presented above may be used to argue that the distinction between characteristic and defining features may be neither feasible nor useful. This conclusion can be drawn on the following grounds. If we adopt the strong version of Smith et al.'s model, such that defining features are those which are common to all category members, then in the present data there were not enough features of this kind found for them to be a sufficient definition of category membership, that is for them to exclude all nonmembers. (BIRDS—the category most commonly used for illustrative purposes (Collins & Quillian, 1969; Smith et al., 1974)—is an exception to this result). Thus this strong version is given little support by the results. It is possible, of course, that certain essential features which might have provided an adequate definition were not elicited by the methods used, or were excluded as being given too infrequently by subjects. However it would be hard to reconcile this argument with the notion that defining features are the most important for defining the category, given that rated importance is correlated with production frequency. (It is also in principle impossible to fully test the sufficiency of a definition, since a counterexample consisting of a nonmember with all the defining features may still be found).

If we adopt the weaker version of the model, then it may be asked whether the distinction between defining and characteristic features is still a useful one to make. Thus the results presented above indicate that typicality is not only a function of the characteristic features, but also varies with the degree to which an item possesses features common to the whole class (which *must* be defining features for either version of the model). Thus typicality is not the sole province of characteristic features. Similarly the role of the defining features for determining an "error-free" judgment of category membership is placed in doubt by the finding, supporting McCloskey and Glucksberg (1978), that category membership is poorly defined, with several cases falling on the boundary. With the weaker version, it becomes difficult to know how to identify defining features. Perhaps they could be identified as those for which a difference in value between two items on the feature in question, with all other features held constant, can on occasion result in one item being a category member and the other item not. Such an operational definition would be extremely impracticable to operate. Also one could never be sure that a particular feature might not be defining, if only the right pair of examples could be found.

As a way out of these difficulties, an alternative model will be described, which is similar in many respects to Smith et al.'s (1974) model, except that no dichotomy of features is made.

An alternative model based on features. The usefulness of distinguishing *defining* from *characteristic* features was thrown in doubt by the results of Experiment I. However it is possible to have a feature-based model of category definitions, which makes no use of such a distinction. Given the lists of category features produced in Experiment I, a simpler model can be tested.

A Polymorphous Concept can be defined as one in which an instance is classified as belonging to a certain class, if and only if it possesses at least a certain number of a set of features, none of which need be necessary or sufficient in itself. Such concepts are common

HAMPTON, JAMES A., **Polymorphous Concepts in Semantic Memory** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

in biology (Sneath & Sokal, 1973), and have also been tested in concept formation tasks (Dennis, Hampton, & Lea, 1973; Lea & Harrison, 1978) and in the field of ethnosemantics (Hunn, 1976).

(The term "polymorphous" is taken from Ryle (1951). It should perhaps more correctly be termed "polythetic" to be in keeping with the terminology of taxonomists (Sneath & Sokal, 1973), as polymorphous might also be taken to mean "having more than one prototypical form"—a meaning not intended in the present context).

If in addition to the above formulation, we allow for differential weighting of the features, then the polymorphous concept becomes essentially equivalent to the idea of a prototype concept, developed by Rosch (1975). Thus Rosch & Mervis (1975) showed that the members of a category such as Furniture bear "family resemblances" to each other, such that no simple conjunction of features was sufficient for defining the category. The polymorphous concept model proposes that these resemblances which constitute the prototypical category member are stored as features constituting the meaning of the category name. In both models, the typicality of a member is determined by the number of category, or prototype features that it possesses, with a possible weighting for their differential definingness or cue validity.

The notion of a polymorphous concept can easily be applied to the data described above. Category membership may be treated as a continuous scale, where the position of a word is defined by the number of category features it possesses. Thus highly typical members would be at the top of the scale, having a large proportion of the category features; whereas unrelated words would receive a low score, not having any features in common with the category. When applied to the present data the number of category features possessed by a word should therefore correlate positively with the scale of rated membership described above. The observed correlation between the two scales was indeed high and positive for each category, (Kendall's tau, $.61 < \text{tau} < .78$, in all cases $p < .001$, $N = 30$).

These values indicate that the number of features possessed may be a better predictor of rated membership than is the category production frequency (from Battig & Montague, 1969). Thus category production frequency correlated positively with rated membership, with Kendall's tau ranging across categories from .39 to .68, ($p < .005$, $N = 30$). The difference between the two sets of correlation coefficients is significant across categories on a Wilcoxon matched-pairs signed-ranks test ($p < .05$) (Siegel, 1956), although obviously the unknown reliability of the two measures makes such comparisons suggestive rather than conclusive.

The question of weighting the features is clearly also of interest. Rosch & Mervis (1975) suggest that the features are weighted by their cue validity for determing category membership. Similarly, Smith et al. (1974) use the notion of weighting, but only in the second stage of the decision process, where defining features have maximal weight, and characteristic features have zero weight. Many taxonomists however (e.g., Sneath & Sokal, 1973) find the notion of weighting of polythetic classifications to be problematic methodologically. Two forms of weighting are relevant for a polymorphous concept—(a) the degree to which an instance possesses the category feature and (b) the degree to which the feature is defining of the category and so given weight in the sum of features. With regard to the first form of weighting, the degree to which an item possesses a category feature, this weighting did prove important for the fit of the model. In every category except one, the value of Kendall's tau improved when the weighting was used as compared with a simple plus, minus or zero judgement of whether each word possessed each feature. The mean value of tau rose from .622 to .674 taken across all eight categories. For the second form of weighting—the definingness of the features –

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

it is clearly important to use an independent measure of feature definingness, if a valid test of the model is to be made. Two such measures were available in the data of the experiment rated importance and production frequency. When weighted with either of these two measures, however, there was no significant improvement in the fit of the model (mean taus were .643 weighted by median ranked importance, and .660 weighted by production frequency). This failure to improve the fit does not of course imply that some transformation or function of one or both measures might not be more successful. If the weighting had improved the fit, then this might have been taken as some evidence for Smith et al.'s model, since giving more weight to the features rated most important should approximate more closely to the situation in which the most important (the defining features) receive a weight of + 1, and all others a weight of zero. However the evidence was not forthcoming. In the light of these results, a scale weighted for degree of instance possession, but unweighted for definingness was used in the further testing of the model in Experiment II.

What are the advantages of using a polymorphous concept model of category definitions? With respect to Smith et al.'s (1974) model, the advantage is that with just one derived scale one may predict the typicality of category members, the membership in the category of any item, the existence of borderline cases of membership and the relatedness of nonmembers. So long as Smith et al. are able to motivate their distinction of two kinds of feature in terms of the functions they perform, then their model may be preferred as making more specific predictions and hence being the more refutable theory. However, once the distinction of functions becomes blurred, and the procedures for identifying the defining features become vaguely defined, then one may argue on grounds of parsimony that the distinction is no longer helpful. In order to demonstrate the generality of the polymorphous concept model, the following experiment aims to show that the derived scale the number of category features possessed by a word—can also be used to predict the speed with which people can make categorization decisions. Categorization latency was one of the main ways in which Smith et al.'s (1974) model was tested. It is therefore clearly important to demonstrate that the proposed alternative model can account for such data. In addition a particular test of Smith et al.'s (1974) model is possible which relies less heavily on the assumption of a strong version of their notion of a defining feature.

EXPERIMENT II

An important aspect of any model of category definitions should be its ability to account for two well-established findings concerning the time it takes to categorize a word. The first finding is that the more typical a word is of a category, as measured by a number of different variables such as production frequency (Wilkins, 1971; Loftus, 1973) or ratings of typicality (Smith et al., 1974) then the faster subjects will be to decide that the word is a category member. The second finding is that the more related a noncategory word is to a category, then the slower subjects will be to decide that it is not a member. Since for the Polymorphous Concept model, both typicality of members and relatedness of nonmembers are determined by the number of category features possessed by a word, the model would require that the speed of a categorization response should be closely related to this variable. More specifically, Yes responses should be faster and No responses slower the greater the number of category features possessed by a word.

From Smith et al.'s (1974) model a further prediction can be derived. They provide a very specific two-stage model of the categorization process. The first stage is a rapid holistic appraisal of the overlap between the features of the word and those of the category. If

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

sufficient agreement or disagreement is found then a response is produced. If not, a second stage is required in which just the defining features of the category are checked, thus producing a slower, more accurate response. A crucial part of the model is that most errors will occur in the first stage, and indeed Smith et al. (1974) found that erroneous responses were faster than responses to atypical members.

It was noted in Experiment I that the features which are common to all category members are also common to several nonmembers, in all except two categories. The procedure adopted for finding the defining features was therefore clearly inadequate, if we assume the correctness of Smith et al.'s (1974) model. However, with *either* version of their model, those features which are common to all category members are defining features. In other words, Experiment I identified some, but not all of the defining features of the categories. Consider then the case of a subject who responds Yes in a categorization task to a word which *lacks* one or more of the defining features identified by Experiment I. By the two-stage model such a response must by definition be an error and is therefore most likely to be produced by the first stage of processing. Such responses should therefore be as fast as other responses originating from the first stage—namely Yes responses to very typical category members, and No responses to unrelated nonmembers. This inference is only strictly true for the strong version of the model, since by the weak version one may occasionally get Yes responses in the second stage where a word is lacking one of the defining features. However one would expect a significant proportion of such responses to be first stage errors even on the weaker version of the model. This prediction can be tested by measuring the categorization latencies for a set of words ranged across the scale of category membership. If the extent to which each word possesses each of the category features is known, then the Yes responses made to words lacking at least one defining feature can be identified and their mean latency compared to other kinds of Yes response to determine which stage of processing they might have resulted from.

The polymorphous concept model makes the different prediction that Yes response latency will depend on the degree of feature overlap, *regardless* of which features are possessed. Words lacking a defining feature are likely to have a low number of category features, and Yes responses to such words should therefore be slow.

Method

Subjects. Eight male and eight female undergraduate volunteers from University College, London were paid to be subjects in the experiment.

Design and procedure. The experiment aimed to measure the time taken to categorize words, and to relate this variable to the number of category features that a word possessed.

Words and category names were presented on the display scope of a PDP-12 computer, and subjects responded by depressing one of two response keys, which were placed by each of the subject's hands. Response times were measured by the computer from the onset of the display to the nearest one hundredth of a second. The list of words for each category were presented as a block of trials, with one word being displayed on each trial. The eight category lists were presented in the same balanced order as in Experiment I. Each subject received a different random order of the words in each list. Half the subjects responded Yes with their preferred hand, and half with their nonpreferred hand.

The procedure was as follows. A question such as "Are these words kinds of Furniture?" was first displayed for 3 seconds. A list of 26 words corresponding to the named category, and including typical and atypical category members and related and unrelated nonmembers of the category, was then displayed one at a time in the centre of the screen. The subject's response on the Yes or the No key terminated the display and brought up the next word after a delay of a few seconds while the teletype printed out the response and latency for that trial. At the end of the list of 26 words, there was a pause and the subject himself initiated the start of the next block of trials by pressing a third key. Each new block of trials began with a new category question. Ten practice trials with a different category were given. Subjects were asked to report errors they were aware of making at the end of each list of words. Reported errors were excluded from the latency analysis, but were used in corrected form in the analysis of Yes and No response frequencies. Instructions stressed that subjects should respond as fast as they could, but that it was important that they should avoid making any errors. The subjects were also told to respond only on the basis of their knowledge of the words' meanings as used in everyday language.

Materials. The same eight categories were used as in Experiment I. For each category a list of words was required such that category members of varying typicality and nonmembers of varying relatedness would all be included. A large number of words for each category were rated by the experimenter for the extent to which they possessed each of the category features. From these words, 26 words were then selected at random for each category, within the constraint of covering the range of the feature scale of membership described by the polymorphous concept. Some small adjustments to this selection were made after a pilot study, in order to achieve approximately equal numbers of Yes and No categorization responses over the whole of each list. The words finally chosen were then given to the same four judges used in Experiment I, so that a measure of the extent to which each word possessed each of the category features could be obtained. By summing these ratings across the category features, each word was therefore given a score corresponding to the degree of feature overlap it had in relation to the category. This variable the number of category features possessed was the theoretically derived variable which was predicted to be significantly correlated with categorization latency.

Results and Discussion

(a) *Response frequencies.* As reported already in Experiment I, for the words used in both experiments, the number of Yes responses produced correlated well with the rated membership of the words ($r = 0.96$). The number of Yes responses was also highly correlated with the number of category features possessed by a word (Kendall's tau, $0.72 < \text{tau} < 0.85$, in all cases $p < .001$, $N = 26$), and less highly correlated with the category production frequency of the words (from Battig & Montague, 1969, $0.54 < \text{tau} < 0.82$, in all cases $p < .001$, $N = 26$).

(b) *Error responses.* Unlike many previous experiments (e.g., Smith et al., 1974) a response was only counted as an error if the subject reported it as such. (Subjects were under instructions to report errors that they were aware of making). Errors of this type accounted for fewer than 1% of trials, and were excluded from the analysis of latencies. Because some errors may not have been reported, the additional precaution was taken of excluding responses in which the subject was the only one of the sixteen to respond in a particular way.

(c) *Latencies.* The main interest of the experiment is in the categorization latencies. Latencies for Yes and No responses were analysed separately and mean latencies were calculated for each word in each category list. By including means for responses given by a minority of subjects it is possible to provide data for Yes responses to related nonmembers, and for No responses made to atypical members. Mean latencies for words varied widely. The fastest were in the region of 750–800 msecs (e.g. APPLE—FRUIT had a mean of 752 msec for Yes responses). The slowest were well over 3 seconds (e.g. SURFBOARD—VEHICLE had a mean of 3345 msec for Yes responses ($N = 8$) and a mean of 3010 msecs for No responses ($N = 8$)). The means were then subjected to a correlational analysis. The prediction of the polymorphus concept model was that the number of category features possessed by a word should be a good predictor of the time taken to respond Yes and the time to respond No when categorizing the word. Kendall's tau was calculated, for each category list, between the mean response latency and the number of category features possessed by each word. The values of tau are shown in Table 3, where it may be seen that in all categories except KITCHEN UTENSILS the prediction of significant positive correlations for No responses and negative correlations for Yes responses was supported by the data. Mean values of tau were $-.43$ for Yes responses and $+.55$ for No responses. Thus both the typicality and the relatedness effects on categorization latency can at least in part be accounted for in terms of the number of category features possessed by a word.

TABLE 3

CORRELATIONS (KENDALL'S TAU) OF MEAN RESPONSE LATENCIES WITH SUM OF CATEGORY FEATURES

Category	Sum of features	Number of words
Yes responses		
Kitchen utensil	−0.075	16
Furniture	−0.672**	17
Vehicle	0.532**	16
Sport	−0.383*	15
Fruit	−0.400*	13
Vegetable	−0.442*	14
Fish	−0.406*	17
Bird	−0.508**	14
Mean	−0.427	
No responses		
Kitchen utensil	0.649**	19
Furniture	0.463**	17
Vehicle	0.637**	19
Sport	0.372**	18
Fruit	0.577**	16
Vegetable	0.643**	18
Fish	0.603**	18
Bird	0.456**	14
Mean	0.550	

Significance levels $*p < .05$, $**p < .01$.

Because subjects who were faster overall were also more likely to make the more frequently given response (Kendall's tau = .308, $N = 16$, $p < .05$) an additional analysis was necessary. This analysis was required to test the possibility that the observed correlation of latency with number of features might be entirely due to the slower subjects contributing disproportionately to the less frequent responses. Considering each subject individually it was found that for every subject the mean response times to words to which the subject gave a minority response were longer than the mean response times for words on which all subjects were unamimous. Thus the relation between number of features and latency holds good for each subject separately as well as for the group. When considering individual differences it was also noted that the broader the subject's definition of the category, the faster were his Yes responses as compared with his No responses (Kendall's tau = 0.430, $p < .05$, $N = 16$). The implications of this finding are discussed below.

This experiment was also designed to test a prediction derived from Smith et al.'s (1974) two-stage categorization model, concerning the relation between latency and the defining features. The prediction was that Yes responses given to words which lack one or more of the defining features should for the most part result from the first stage of the decision process. This is because such words must be category nonmembers (since those defining features that were identified in Experiment 1 were required to be common to all members). Thus a Yes response to such a word is an error, and errors are made, according to Smith et al.'s model, almost entirely by the first stage of holistic comparison. Such responses should therefore be nearly as fast as

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

responses to very typical members, and considerably faster than responses to the less typical category members which are assumed to require the second stage of the decision process for a correct response to be made. The mean response latency for Yes responses to words lacking at least one defining feature was compared with mean latencies of responses to very typical members (those with all 16 subjects responding Yes), and mean latencies for Yes responses to atypical members (those with between 9 and 12 Yes responses). Six of the eight categories gave sufficient data for this analysis to be possible. The results are shown in Table 4, where it can be seen that in all six categories, Yes responses to words lacking a defining feature took longer (2124 msec) than those to typical members (989 msec), and took as long if not longer than those responses to atypical members (1520 msec). The first but not the second comparison was significant (sign test across categories $p < .05$).

The interpretation of this result must be treated with caution. Closer consideration of Smith et al.'s model reveals that although the prediction, if upheld, would be seen to support the model, yet it is not a necessary prediction. In other words, the model can be made to account for the results. There are in fact three possible reasons for a subject saying Yes to a word lacking a defining feature as a result of second stage processing. First, his own individual definition may not include that particular feature. Second, he may consider that the item in question *does* possess that feature. Third, it may be a case of the weak interpretation at work, where occasionally Yes responses are produced in the second stage without a word having to possess all the defining features. In this case the problem becomes one of determining the expected relative frequency with which one of these three situations may occur. In particular, is it reasonable to suppose that most of the 51 responses of this type were due to one of these three reasons? (They represent about 3% of the Yes responses made). The first two explanations raise a question that is problematic for most models of semantic memory as presently formulated namely that individuals' concepts may differ in important respects. As yet no experiments have attempted to relate an individual's own definition of his concepts to his categorization performance, although the notion of a variable criterion level for positive judgments has been suggested as an individual

TABLE 4

Mean Latencies of Yes Responses Given to (a) Category Members Possessing All Defining Features and Given 16 Yes Responses; (b) Category Members Possessing All Defining Features and Given Between 9 and 12 Yes Responses; and (c) Nonmembers Lacking at Least One Defining Feature and Given Less Than 8 Yes Responses

Category	(a) Words with all defining features given 16 Yes responses		(b) Words with all defining features given 9–12 Yes responses		(c) Words lacking a defining feature given less than 8 Yes responses	
	M	*N*	*M*	*N*	*M*	*N*
Kitchen utensil	1.032	32	1.229	24	1.105	13
Furniture	0.945	80	1.666	41	4.810	4
Vehicle	1.175	96	1.569	35	2.727	12
Sport	0.997	112	1.075	11	2.527	3
Fruit	0.869	96	1.274	11	1.860	1
Fish	0.925	80	1.861	20	1.809	18
Weighted mean	0.989	496	1.520	142	2.124	51

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

difference (Gardner, 1953). The third explanation highlights the fact that given a weaker version of defining features, the second stage of Smith et al.'s (1974) processing model is capable alone of predicting the latency results. If not all the defining features need be possessed for a positive response to be made, then the checking of them need not be exhaustive. A sampling strategy could be used with a random-walk choice model, so that when sufficient positive or negative information had been accumulated a corresponding response could be made. Such a model would be supported by the finding reported above that the more Yes responses a subject made, then the faster were his Yes responses as compared with his No responses.

A different argument might point to the dissimilarities between the data of the present experiment and those of Smith et al. (1974) in the speed with which subjects were responding. Smith et al.'s subjects were responding much faster (mean latencies around 600 msec) and making many more errors than the subjects in Experiment II. They were therefore making a different kind of error from the slow and carefully considered "misclassifications" found in the present experiment. It is entirely possible that with different constraints on expected speed, the subjects in the two experiments were using different processing strategies based on different points on the speed–error trade-off. Thus if subjects are responding very fast, errors are made with very short reaction times, whereas if they are being more slow and careful the errors are in effect just more extreme cases of the slow responses given to atypical members. Thus a response is only an "error" because it goes against the majority of the group's decision. In spite of this account of the difference between the two experiments, the presence of error rates as high as 52% (for the No responses in the category FRUIT) in Smith et al.'s data, suggest that the definition of what counts as an error should perhaps more reasonably be left to the subject to decide. Thus a recommended procedure might be to present the subject after the experimental session with a list of his responses and obtain his decision as to their correctness.

The results of this experiment can best be summarized as follows: When subjects are making relatively careful category decisions, the more category features that a word possesses (regardless of whether the features are common to all members or not), the faster will be a Yes response and the slower will be a No response. This result held true regardless of whether a word was judged overall as a category member or not. There was no evidence that infrequently made responses were "errors" resulting from a hasty first impression in the way described by Smith et al.'s processing model.

GENERAL DISCUSSION

The results of the two experiments described here suggest that an important aspect of the definition of concepts such as FRUIT or FURNITURE is that there are *no* features which taken in conjunction provide a necessary and sufficient definition. Such concepts appear to be fairly widespread and have been discussed extensively by philosophers of language (Wittgenstein, 1953; Ryle, 1951). Indeed Needham (1975) advocates a much wider use of such systems of classification for the sciences. The important point is that this lack of a "common-feature" definition, need not imply that the concepts have no definition of any kind. The Polymorphous concept does allow the exact determination of category membership by fixing a critical number of features which a category member must possess. It seems likely however that such a criterion is quite flexible, depending on the demands of the task, and that different people might set a more or less strict criterion level. Hence instances may fall on the borderline of a category, as illustrated by the extensive disagreement between subjects in Experiments I and II, concerning the categorization of several words.

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

While in certain respects failing to support the Characteristic Feature Hypothesis of Smith et al. (1974) the results are consistent with the simpler polymorphous concept model of category definitions. In a series of papers (Rosch, 1975; Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 1976), Rosch and her colleagues have analyzed how instances of a category are more or less typical because of the degree to which they resemble the other members of the category. She argues from this that a notion of an "ideal prototype" is set up in memory, as the hypothetical or real object which possesses most resemblance to the members of the category. It can be seen that the properties elicited in Experiment I as category features serve the function of giving a description of just such a prototype. The Polymorphous concept and the notion of a prototype concept are therefore closely related. The present model can be seen as developing prototype theory in two ways, first by showing that the "family resemblance" features can be elicited as definitions of the category concept, (Rosch & Mervis (1975) used properties generated to category instances only), and second by demonstrating that the sum of category features possessed by a word can serve not only to predict its typicality, membership or relatedness to the category, but also the time it takes people to decide that it is or is not a category member.

In this way, the model described here is not intended primarily to be an alternative to that proposed by Rosch (1975), but rather as a way of bringing out the differences between Rosch and Smith et al. (1974).

Thus while limited support for the approach of Smith et al. was found in the evidence for a dimension of feature definingness, and for the existence of features possessed by the most typical items of a category but not by the less typical ones, the crucial distinction between defining and characteristic features was not found to be well supported. A strong version failed to find support in the data, whereas a weaker version was found to be less parsimonious than a simpler model that makes no dichotomy between features that are more or less defining.

APPENDIX: FINAL LIST OF CATEGORY FEATURES SELECTED IN EXPERIMENT 1

No.	Feature	Defining or characteristic	Associated with relatedness	Production frequency
(a)	Kitchen utensil			
1	Has a specific function, is used by humans	D	+	32
2	Is connected with food	D		30
3	Is found in the kitchen	D	+	29
4	Is for the preparation or cooking of food	C		27
5	Is made of metal in part	—		18
6	Is portable, can be lifted and held	C	+	18
7	Is *only* found in the kitchen	—		13
8	Has a handle	C		13
9	Is efficient, does a job well	D	+	11
10	Is made of wood in part	—		10
11	Is relatively small	C	+	10
12	Is manufactured, man-made	D	+	10
13	Is a container	—		9
(b)	Furniture			
1	Has a specific function, is used by humans	D	+	32
2	Is found in buildings	—	+	30

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

No.	Feature	Defining or characteristic	Associated with relatedness	Production frequency
3	Is made of wood	C		20
4	Is for sitting on or in			18
5	Is attractive	—		16
6	Is for comfort, convenience or satisfaction	D	+	16
7	Is found in the home	—	+	15
8	Is man-made, manufactured	D	+	15
9	Is for putting things on or in			13
10	Is not just decorative	D	+	12
11	Has legs	C		10
(c)	Vehicle			
1	Carries people or things	D	+	32
2	Can move	D	+	32
3	Moves along	D	+	32
4	Has wheels	C		31
5	Carries people	—		28
6	Is powered, has an engine, uses fuel	C		28
7	Is self-propelled, has some means of propulsion	C		21
8	Is used for transport	—	+	17
9	Is steered, has a driver controlling direction	—	+	16
10	Has a space for passengers or goods	C		10
11	Moves faster than a person on his own	D		6
12	Man-made	D		4
(d)	Sport			
1	Is physical, connected with the body	D	+	31
2	Is a human activity	D	+	30
3	Is a pastime, recreation, done for leisure	D	+	29
4	Is competitive	C	+	25
5	Involves physical exertion, is energetic	—		24
6	Is enjoyed when done	D	+	22
7	Is not a solitary activity	—	+	18
8	Has teams	C		17
9	Has spectators	C		16
10	Involves physical skill	D	+	14
11	Has rules	C	+	10
12	Uses special equipment		+	9
13	Takes place outdoors			8
14	Is exercise			8
(c)	Fruit			
1	Is a plant, organic, vegetation	D	(insufficient data)	31
2	Is edible, is eaten	D		30
3	Contains seeds	C		27
4	Grows above ground, on bushes or trees	C		26
5	Is juicy, thirst quenching	C		17
6	Is brightly coloured	—		16
7	Is sweet	C		15
8	Has an outer layer of skin or peel			13
9	Is round	C		9
10	Is eaten as a dessert, snack, or on its own	C		8
11	Is a protection for seeds	C		7

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory*** , Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

No.	Feature	Defining or characteristic	Associated with relatedness	Production frequency
(f)	Vegetable			
1	Is edible, is eaten	D	+	32
2	Is a plant, organic, vegetation	D	+	30
3	Grows underground or close to the ground	—		22
4	Is green	C		18
5	Is cooked	—		13
6	Is eaten for dinner, with meat, salt etc.	—		12
7	Has leaves and stalks	C		10
8	Is not sweet, is savoury or tasteless	—		10
9	Is cultivated	D	+	9
(g)	Fish			
1	Is alive	D		32
2	Lives in water	D	+	32
3	Has fins	C		31
4	Swims	—	+	27
5	Has gills			26
6	Has scales	C		23
7	Has a tail	D		20
8	Breathes water	—	+	18
9	Is streamlined	—	+	16
10	Has eyes large, round and at the side	D		13
11	Dies out of water	D	+	12
12	Is cold-blooded	D	+	11
13	Has no limbs	—		11
14	Is a long oval 'fish-like' shape	C		9
15	Has a mouth	D		9
16	Is edible	C		8
(h)	Bird			
1	Is alive	D	(insufficient data)	32
2	Flies	C		32
3	Has feathers	D		30
4	Has a beak or bill	D		26
5	Has wings	D		25
6	Has legs and feet	D		24
7	Lays eggs	D		20
8	Has just two legs	D		18
9	Builds nests			15
10	Sings, cheeps, etc.	C		14
11	Has claws	C		12
12	Is very lightweight	C		8

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monograph*, 1969, **80**, (3, Pt. 2).

Bock, M. The influence of instructions on feature selection in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 183–191.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, **82**, 407–428.

HAMPTON, JAMES A., ***Polymorphous Concepts in Semantic Memory***, Journal of Verbal Learning and Verbal Behavior, 18:4 (1979:Aug.) p.441

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–248.

Collins, A. M., & Quillian, M. R. How to make a language user. In Tulving E. & Donaldson W. (Eds.), *Organization of Memory*. New York: Academic Press, 1972.

Dennis, I., Hampton, J. A., & Lea, S. E. G. New problem in concept formation. *Nature*, 1973, **243**, 101–102.

Gardner, R. W. Cognitive styles in categorizing behavior. *Journal of Personality*, 1953, **22**, 214–233.

Glass, A. L., & Holyoak, K. J. Alternative conceptions of semantic memory. *Cognition:* 1975, **3**, 313–339.

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

Hunn, E. Toward a perceptual model of folk biological classification. *American Ethnologist*, 1976, **3**, 508–524.

Lakoff, G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Papers from the eighth regional meeting, Chicago linguistic society*. Chicago: Univ. of Chicago Linguistics Department, 1972.

Lea, S. E. G., & Harrison, S. N. Discrimination of polymorphous stimulus sets by pigeons. *Quarterly Journal of Experimental Psychology*, 1978, **30**, 521–537.

Loftus, E. F. Category dominance, instance dominance, and categorization time. *Journal of Experimental Psychology*, 1973, **97**, 70–74.

McCloskey, M., & Glucksberg, S. Natural categories: Well defined or fuzzy sets? *Memory and Cognition*, 1978, **6**, 462–472.

Needham, R. Polythetic classification: Convergence and consequences. *Man*, 1975, **10**, 349–369.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 1975, **104**, 192–232.

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

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

Ryle, G. Thinking and language. *Proceedings of the Aristotelian Society, Supplementary Series*, 1951, **25**, 65–82.

Siegel, S. *Non-parametric statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review*, 1974, **81**, 214–241.

Sneath, P. H. A., & Sokal, R. R. *Numerical taxonomy*. San Francisco: Freeman, 1973.

Wilkins, A. T. Conjoint frequency, category size and categorization time. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 382–385.

Wittgenstein, L. *Philosophical investigations* (translated by G. E. M. Anscombe). Oxford, England: Blackwell, 1953.

Reference Note

1. Hampton, J. A. *An experimental study of concepts in language*. Unpublished doctoral thesis, University of London, 1976.

(Received 6 July, 1978)