Tenth Annual Conference of the Cognitive Science Society, August 17-19, 1988.

Basic Levels in Hierarchically Structured Categories

James E. Corter
Columbia

Mark A. Gluck
Stanford

Gordon H. Bower
Stanford

In a hierarchical set of categories, one level of the hierarchy is said to be "basic" if categories at that level seem to be used more easily (e.g. named faster) than categories at other levels. An experiment was conducted to determine whether the feature structure of a hierarchy of artificial categories (fictitious diseases) can induce "basic level" effects. The feature structure of this hierarchy was patterned after an experiment by Hoffman and Ziessler (1983), in which the middle level of the hierarchy was found to be "basic". In a learning task, subjects learned to diagnosis patient descriptions (consisting of a list of symptoms) in terms of the fictitious diseases. In a subsequent verification task, they were shown a patient description paired with a diagnosis, and asked to judge the correctness of the diagnosis. Mean reaction times were smallest for categories at the middle level of the hierarchy, showing that this level was indeed basic. In addition, analysis of data from the learning task showed that subjects made the highest proporation of correct diagnoses for diseases at the middle level, thus confirming the advantage of the middle level. These results provide additional evidence for basic level phenomena in the retrieval of categorical information from memory for hierarchically structured categories; they extend previous findings by showing basic level effects with categories described in terms of conceptual (verbal) rather than purely visual features.

INTRODUCTION

Certain categories seem to be learned and used more easily than others. One line of research that has shown this clearly is the work on "basic level" categories (Mervis & Rosch, 1981). Within a hierarchy of categories at various levels of generality (e.g. *sparrow-bird-animal*), one particular level of category (*bird* in this example) is usually found to be used most readily, and is referred to as the *basic level*. In laboratory studies the criterion variable most often used to determine which level of a hierarchy of categories is basic is the mean reaction time to name categories at that level, or to verify category-name pairings (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Jolicoeur, Gluck, & Kosslyn, 1984).

Rosch and colleagues (e.g., Rosch & Mervis, 1975) originally suggested that basic level categories are special because they capture significant regularities or patterns in the features associated with these categories. For example, basic level categories appear to be the most general (i.e., abstract) categories for which members have many common properties. We refer to this type of explanation of the basic level advantage as a *structural* theory, since it implies that certain categories are "basic" because of their structural properties, namely, the statistical associations between features and categories.

An alternative to such structural theories of the basic level phenomenon is the notion that *type* of feature is important in determining the basic level. For example, Tversky and Hemenway (1984) reported that when people are asked to give lists of features for various categories, there is a difference in the *type*

Correspondence should be addressed to: Dr. James E. Corter, Box 41, Teachers College, Columbia University, New York, NY 10027. This research was supported by NSF Grant BNS-8618049.

of features they give for superordinate and basic level categories. For superordinate categories, associates and functions are commonly listed. At and below the basic level, the most commonly mentioned features of concepts are *parts*, such as *wings* for birds and *engine* for cars.

Thus the existence of basic levels may be due to qualitative differences in the features associated with categories rather than to quantitative differences. Perhaps more abstract, superordinate categories (e.g. *vehicles*, *weapons*) aggregate objects by similar *functions* which are abstract and nonperceptual, whereas basic level categories rely more heavily on common perceptual properties. According to this *feature-type* hypothesis, the basic level findings simply show that perceptual features are more easily used than abstract features such as function.

An important experiment by Murphy and Smith (1982) demonstrated that basic-level effects could be obtained by manipulating the feature composition of a set of artificial stimuli. In this experiment subjects were taught a hierarchy of artificial categories (fictitious tools) that was designed so that the middle level should be basic. In later naming and verification tasks, their subjects responded quickest to categories at the middle level, thus confirming that this level was psychologically "basic". However, Murphy and Smith did not design their experiment to distinguish between the feature-structure and feature-type explanations. While their subordinate and basic level categories were distinguished by perceptual features, their superordinate categories were defined only in terms of function (tools used for pounding vs. tools used for cutting).

This confounding of level with type of feature was avoided in experiments by Hoffman and Ziessler (1983). Different groups of subjects were taught one of three category hierarchies. Exemplars of the categories were schematic line drawings of rocket ships. In one hierarchy, the feature-category associations were such that the top level was expected to be basic, in the second hierarchy so that the middle level would be basic, and in the third so that the bottom level would be basic. These predictions were confirmed by subjects' reaction times to name the categories and to verify category-name pairings. In each hierarchy the categories at the basic level were named fastest, regardless of whether the basic level corresponded to the top, middle, or bottom level of the hierarchy.

The results of these last two experiments demonstrate that feature structure and/or feature type are sufficient to cause a basic-level effect. However, because feature-structure and feature-type were confounded by Murphy and Smith, only the Hoffman and Ziessler experiments demonstrate that feature structure alone is sufficient to cause a basic-level effect. It is important that this critical experiment be replicated and extended. In particular, since pictorial stimuli were used in both previous studies, the question remains of whether it is only visual features that can induce basic-level effects. Therefore, the present experiment was designed to see if basic level effects could be obtained with non-pictorial stimuli, specifically, categories defined in terms of verbal features. If so, this would indicate that the basic level effects are not purely a perceptual phenomenon, but rather arise from very general cognitive mechanisms. Such a finding would thus constitute strong evidence that a purely structural explanation of basic level phenomena is tenable.

EXPERIMENT 1

A set of materials was designed to mimic the feature structure of Hoffman and Ziessler's Hierarchy 2. However, our categories were defined in terms of verbal or conceptual features rather than perceptual ones. Specifically, the categories were diseases and instances of the categories were individual patients with a set of symptoms characteristic of the disease. In a training task subjects learned to diagnose individual patients, each described as a set of symptoms, in terms of which disease the patient had.

Following the training task a verification test was given in which subjects were presented with a variety of patient descriptions paired with a diagnosis, and were asked to indicate whether that diagnosis was correct. The measure of primary interest was the average reaction time to verify diagnoses for categories at various levels of the hierarchy.

Materials

The concepts to be learned were fictitious diseases. Instances of the diseases were individual patient descriptions, each consisting of a list of three symptoms. An example patient description is: "blotchy rash, swollen gums, red eyes". One symptom always pertained to gums, one to eyes, and one to rash. Each of these symptom "dimensions" was substitutive in nature, having four possible values. An individual patient description contained one value of each of the three symptom dimensions. The four possible values of the rash dimension were: *blotchy, spotted, itchy, scaly*. For gums they were *swollen, discolored, bleeding, sore*, and for eyes they were *puffy, sunken, red, burning*.

The set of diseases to be learned were defined at several levels of generality. Specifically, the concepts formed a hierarchy with eight bottom-level, four middle-level, and two top-level categories. The hierarchy of disease names is shown in Table 1. The feature structure (i.e. the symptom-disease associations) of these categories was adopted from Hoffman and Ziessler's (1983) Hierarchy 2, which was designed so that the middle level was expected to be psychologically basic. The stimuli in Table 1 were made up to have this same structure, with symptom dimensions replacing the perceptual feature dimensions of the Hoffman & Ziessler stimuli.

TABLE 1: The Disease Hierarchy and its Feature Structure

Categories			Feature Dimensions		
Top	Middle	Bottom	Gums	Eyes	Rash
queritism	-- burlosis --<	-- jirenza	1	1	1
		-- malenza	1	2	1
	-- cretosis --<	-- gilenza	2	3	2
		-- surenza	2	4	2
philitism	-- midosis --<	-- habenza	3	1	2
		-- kelenza	3	2	2
	-- nitosis --<	-- tumenza	4	3	1
		-- valenza	4	4	1

For each subject the correspondence between features and structure was randomly assigned. Also, middle-level and bottom-level category names were randomly assigned to the feature structure shown.

Method

Subjects were students and other members of the Columbia University community who were paid $6 per hour for their participation. In all, 27 subjects were tested. Subjects were informed that each patient description would consist of one rash symptom, one eyes symptom, and one gums symptom. It was explained that just as real diseases could be diagnosed at more or less general levels (e.g. jaundice, hepatitis, hepatitis A), so the present diseases would be identified sometimes in more general and sometimes in more specific terms. However, the actual hierarchy of diseases was never shown to the subject.

Stimuli were presented, and responses collected, on an IBM AT microcomputer. The training phase involved blocks of two different tasks: study and testing. In a given block diagnoses were to be made only at a single level of the hierarchy. In a study block, the subject saw a succession of patient descriptions, each paired with the correct diagnosis. Immediately after this a testing block was given at the same level. Here the subject saw patient descriptions and had to identify them with names of diseases at that level. A response was entered by typing the first letter of the disease name. Corrective feedback was given following the subject's response. Both the study and testing were self-paced, although subjects going extremely slow during the first few study blocks were urged to speed up. Blocks were of different lengths, depending on the level of categorization, in order to equalize the frequency of usage of each category (disease) name. Top-level blocks consisted of eight trials (one presentation of all eight patterns of symptoms), middle-level blocks had 16 trials (two cycles through the eight patterns), and bottom-level blocks had 32 (four cycles). Thus, each category name was experienced four times per block.

Overall, the structure of the training phase was as follows. Subjects were given a study block followed by a test block at one level, then a pair of blocks at another level (i.e., a study followed by a test block), then a pair at the third level. Then three more pairs of blocks were given, using a different ordering of levels. At this point the subject had seen six study blocks and six test blocks. Six more test blocks followed. If by this time the subject was performing at criterion (90% correct diagnoses) on all three levels, the training was terminated. Otherwise the subject repeated sets of three test blocks (one at each level) until this overall criterion was attained. Three different orderings of blocks were counterbalanced across subjects. One order presented the bottom level first, one the middle level, and one the top level. The orders were: M-T-B-T-B-M-B-M-T-..., T-B-M-B-M-T-M-T-B-..., B-M-T-M-T-B-T-B-M-...

Once subjects had learned the category hierarchy to criterion, their use of these categories was tested in a verification task. In the verification task subjects were presented with a patient description, paired with a correct or an incorrect diagnosis. The subject responded "yes" or "no" by means of specially labelled keys on the keyboard. Simple feedback on the validity of the subject's response ("correct" or "incorrect") was given. Half the trials were true ("yes") trials, and half were false ("no") trials. The reaction time and correctness of each response was recorded. There were 12 blocks of 24 trials each. As in the training phase, in a block of trials diagnoses (both correct and incorrect) were made only at a single level. The same three orderings of levels described above were used in this task.

Results

Six subjects (out of 27) did not learn the categories to criterion within the allotted time (2.5 hours) and therefore were not given the verification task. Thus complete data was available for 21 subjects.

The measure of primary interest for identifying the basic level is the mean reaction time to verify diagnoses at each of the three levels in the verification task. Responses should be quickest for categories (diseases) at the basic level. In the verification task there were four blocks at each level. The first of these was not analyzed. Based on an examination of the distribution of reaction times for individual subjects, responses with latencies greater than 15 seconds were considered to be outliers and were excluded from further analysis. Less than 1% of the responses were eliminated by this rule. The mean reaction times for each level (averaged over the three analyzed blocks) are shown in Table 2.

TABLE 2: Mean and S.D. of Verification Reaction Times by Level of Categorization

	TOP	MIDDLE	BOTTOM	OVERALL
MEAN	3.115	2.567	3.045	2.909
S.D.	1.313	0.679	0.858	1.001

The hypothesis that the middle level would be verified fastest was tested with a special contrast in a repeated-measures ANOVA. The hypothesis was confirmed, $F(1,20) = 14.20$, $p < .01$. A comparison of the top and bottom levels was not significant, $F(1,20) = 0.05$. A majority of subjects (11 out of 21) showed the expected pattern (that is, verified the middle-level categories fastest). Five subjects verified the top-level categories fastest, and five the bottom level categories.

The learning task can also provide useful data on the relative "goodness" of categories at the three levels of the hierarchy, in the form of the proportion of correct category identifications at each level of the hierarchy. If the middle level is indeed basic for this structure, then a higher proportion of middle-level categories should be correctly identified, compared with top-level or bottom-level categories. This hypothesis was confirmed by the data. Across all blocks, the mean proportion correct was .824 for the top level, .924 for the middle level, and .878 for the bottom level. The advantage of the middle level over the top and bottom levels was tested by a special contrast in a repeated measures ANOVA, and was significant at the .01 level, $F(1,20) = 8.96$. Top and bottom levels did not differ significantly, $F(1,20) = 1.51$. The ordering of levels by proportion correct did not change when the proportions were corrected for guessing by the formula $p' = p - \frac{(1-p)}{(k-1)}$, where p' is the corrected proportion, p is the uncorrected proportion of correct identifications, and k is the number of alternative categories at that level. The corrected proportions for the top, middle, and bottom levels were .648, .899, and .861.

The mean proportion correct for a level is calculated over all blocks of a subject's data. Thus, it is an average of a wide range of values, which generally increase toward unity. It is natural to ask if the relative advantage of the middle level varies as a function of block, perhaps decreasing as all levels are learned to criterion. Figure 1 presents the mean proportion correct for each level, calculated by block for the first four blocks (which were the only blocks experienced by every subject).

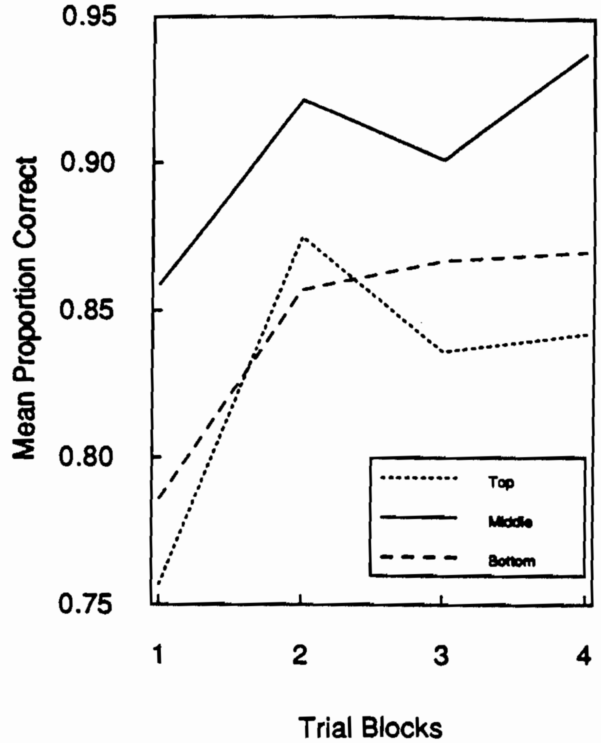

Figure 1. Mean proportion correct for levels by block (first four blocks) from Experiment 1.

It can be seen from Figure 1 that the middle level exhibits an advantage over the top and bottom levels in all blocks. A drop in accuracy for the top and middle levels from Block 2 to Block 3 is evident, no doubt due to the fact that each of the first two test blocks was immediately preceded by a study block.

Analyses were also done on the learning data from the six subjects who did not learn the diseases to criterion in the allotted time. Average proportion correct for these subjects showed the same pattern (highest accuracy on the middle-level categories) as the subjects who successfully learned the categories to criterion.

SUMMARY

In the present experiment naive subjects were taught a hierarchical set of artificial categories, then tested in a verification task. Mean reaction times in the verification task indicated that the middle level of the stimulus hierachy was basic, since diagnoses were made most quickly at that level. In addition, fewest errors were made on the middle level during the training task, indicating that the middle level was easiest to learn.

The results of the present experiment demonstrate that the feature *structure* alone of a hierarchical set of categories is sufficient to induce a basic level effect, thus replicating the experiments of Hoffman and Ziessler (1983). These results demonstrate that explanations of basic levels based on feature *type* are not necessary. In addition, the present experiment extends previous findings by demonstrating basic level effects with categories described in terms of verbal rather than schematic visual features. Accordingly, the present results provide evidence that the basic level effects are not purely a perceptual phenomenon, but rather reflect more general characteristics of human learning and memory. This provides an additional empirical constraint on models of human category learning. We are currently working on an extension of our adaptive network model (Gluck & Bower, 1988) to account for basic level effects, but these results are at too early a stage to report here.

REFERENCES

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

Hoffmann, J., & Ziessler, C. (1983). Objectidentifikation in künstlichen Begriffshierarchien. *Zeitschrift für Psychologie, 194*, 135-167.

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*, 243-275.

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

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21*, 1-20.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-603.

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

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169-193.