Typicality effects in logically defined categories

LYLE E. BOURNE, JR.

University of Colorado, Boulder, Colorado 80309

Subjects learned to categorize geometric designs by some rule combining two features, *x* and *y*. When *xy* stimuli (stimuli with both features) are always positive, the rule (concept) is inclusive disjunctive; when *xy* stimuli are always negative, the rule is exclusive disjunctive. In three other experimental conditions, the probability of *xy* stimuli being positive during acquisition was .25, .50, and .75. In a variety of postacquisition tests, *xy* stimuli were chosen as prototypical despite their less than consistent occurrence in the positive category. Stimuli with one relevant feature (*x* or *y*), although consistently assigned to the positive category, were often evaluated as poorer examples of the concept. These results are interpreted in terms of a schema model in which information acquired during learning is organized according to probability density functions over feature dimensions. This theory appears to be general enough to accommodate the evidence from laboratory studies of both logical and natural concepts.

Conventional laboratory studies of concept learning have been criticized for their simplicity and artificiality and their failure to tell us anything useful about natural categorization processes or about category organization (Rosch, 1978). These experiments typically use simple stimuli that take on only a small number of discrete values (or features¹) on a small number of separable, orthogonal dimensions (Garner, 1978), with each dimension represented by the same number of values. The stimulus features and the categories to be learned are usually already well-known by the subject. Every instance of a particular category, for example, "red square," is as good as any other instance of the same category. In contrast to "laboratory" concepts, natural categories often appear to be ill-defined and not nearly so clear, or artificial, or internally unstructured. Instances of natural categories vary on a large number of not necessarily orthogonal dimensions. Furthermore, each dimension may have many or few and not necessarily the same number of values. In fact, many dimensions are continuous rather than discrete. One instance of a concept is not necessarily as good as any other. Indeed, natural categories often seem to be based on "typical" or "best" instances. According to Rosch (1978), these prototypes form a focus for the category and serve as the basis for organizing our perception of category instances. They can be considered memory representations of categories in analog format. In a variety of experiments, Rosch and colleagues (Rosch & Mervis, 1975; Rosch, Mervis, Gray,

Johnson, & Boyes-Braem, 1976) demonstrated the reality of prototypes within natural categories and a variety of empirical effects attributable to the varying degrees of membership of category instances.

There are, indeed, important limitations on the generality of traditional laboratory experiments insofar as the formation and knowledgeable use of natural categories is concerned. One need not despair of these criticisms, however, for traditional laboratory experiments by and large were not intended to deal directly with the formation and use of natural categories, although Bruner, Goodnow, and Austin (1956) make a strong claim to the contrary. They focus primarily on the formation of well-defined logical concepts, rather than ill-defined semantic concepts. Further, these traditional studies have attempted primarily to discover the processes (e.g., strategies) by means of which relevant attributes or relationships among relevant attributes are identified, or learned, rather than to examine the structure of these categories or their memory representation. But beyond that, the general theoretical position that has been the basis for traditional work might be able to accommodate phenomena like typicality that have been identified within natural concepts (Osherson & Smith, 1981). Thus, the existence of those phenomena in no way vitiates traditional theory or traditional methodology.

Consider the following arguments. Suppose people mentally represent concepts as dimensionalized schemata. The subject enters any new concept learning problem with a generalized schema containing empty slots corresponding to known dimensions of the stimuli to be categorized. The slots act as requests for information about these dimensions from the task (stimuli and feedback). As slots fill with information, some dimensions are identified as relevant, because of their contingency with feedback, and conversely, others are identified as irrelevant. At problem solution, the

This research was conducted within the Institute of Cognitive Science and was supported by Research Grant BNS 76-81416 from the National Science Foundation. The author wishes to thank Michelle Collins, Linda Liebelt, and Peder Johnson for various contributions to the project, ranging from construction of stimulus materials through extremely useful discussions about the theoretical ideas expressed. Of course, any errors of execution or logic are solely the responsibility of the author. This paper is Publication 104 of the Institute of Cognitive Science.

subject has a concept-specific schema that contains pertinent information on relevant and irrelevant stimulus dimensions in the appropriate slots (or could conceivably lack any slots for irrelevant information altogether). For an application of this model to text-based concepts, see Kozminsky, Kintsch, and Bourne (1981).

On this view, people become aware of or focus upon stimulus dimensions that exhibit, over a sample of stimuli, some contingency with the positive (or possibly negative, or contrast) category of a concept. Commonly occurring features within the range of possibilities on a given dimension become the focus of attention and enter into a person's understanding and representation of the concept. In attribute identification problems (Haygood & Bourne, 1965), subjects know (by instruction and pretraining) the general solution form or rule of the concept. Feature category contingencies that emerge over a series of instances permit the subject to identify the relevant features. Relevant features are entered into the rule to determine category membership for any new stimulus (Bourne, Ekstrand, Lovallo, Kellogg, Hiew, & Yaroush, 1976).

Thus, a person's representation of a concept might be described as some relationship (rule) among a set of probability functions, one for each identifiable dimension of the stimuli. Such a probabilistic characterization of a concept deals adequately with many of the phenomena identified in studies of natural concepts. For example, for X to be a defining dimension, $p(X)$, a probability density function, must be "significantly" nonrectangular for the category of positive instances. A probability density function generates the probability that any stimulus with value x_i on Dimension X is a positive instance. If it is nonrectangular, then it contains at least one value, x_m , with a maximal contingency with the concept. Now the definition of a prototype, at least for conjunctive and inclusive disjunctive concepts, is straightforward. The prototypical stimulus will be the stimulus that contains values x_m on all defining dimensions, X . Even if subjects have never seen the prototypical stimulus configuration, they will still identify it as the "best example" of the concept category, for this is the modal stimulus in the similarity space defined by all dimensions of the stimulus set. The degree of category membership is also straightforwardly calibrated. The degree of category membership for any stimulus will be a function of (1) how many x_m s it exhibits, (2) how close its value, x_i , on Dimension X is to x_m , in the similarity space, and finally, (3) which x_m s are involved. The last criterion takes into account the fact that the peak values on the various defining dimensions can themselves differ in probability. In this theory, the prototype does not serve as the focus of the category, as the (analog) representation of the category, or as having any other special status. In this view, the prototype is derivative of feature occurrence information collected schematically by the subject in the course of forming a category.

The position that this theory tries to capture is not unrelated to the arguments of Osherson and Smith (1981). Their claim is that prototype theory, in a general sense, is incompatible with two criteria of adequacy for theories of concepts, one concerning the relationship between complex concepts and their simpler conceptual constituents (rules and relevant features) and the other concerning the truth conditions on simple inclusive relations. One possibility for resolving these incompatibilities is to assert that prototype theory is limited and fundamentally incomplete. To extend prototype theory, Osherson and Smith distinguish between the core of a concept and its identification procedure; core is intensive and definitional, and identification relates to surface information about instances (e.g., features) on which membership decisions are made. Traditional (nonprototype) theories, which appear to prescribe rigid conceptual boundaries, capture the definitional or core aspect of a concept, whereas prototype theories relate most clearly to identification procedures operating on instances. Osherson and Smith suggest, as does schema frequency theory, that even when subjects acquire a well-defined logical concept, typicality phenomena that are based on the presence or absence of surface features might still be evident.

It is difficult to think of any single experiment that allows a direct comparison of a feature-contingency approach to category formation and the prototype approach in general form, let alone any of its specific realizations. The following experiment, however, might shed some light on the difference between these two formulations. The experiment hinges on what seems to be one of the assumptions of most prototype models, namely, that the prototypical instance maximizes the number of defining characteristics of the concept in question (e.g., Rosch & Mervis, 1975). Ideally, the prototype is a conjunction of all defining features. There appears to be no provision in current versions of prototype theory for what might be equivalent to a nonoccurrence marker, that is, a tag in memory to indicate that, under some circumstances, Feature x_m cannot occur in an exemplar. Suppose the category to be learned is characterized by a set of features, some of which show a low or nonexistent conjoint frequency. In other words, Features x and y are among the several that show a high-probability, contingent relationship with the category. It is the case, however, that if the positive instance has Feature x , it is unlikely to have Feature y . Alternatively, if it has Feature y , it is unlikely to have Feature x .

Conjoint probabilities are as straightforward to manipulate as are the probabilities of individual features. When the conjoint probability of two defining features is 0, their relationship can be described as an exclusive disjunction: "Members of the category must have (or are likely to have) Feature x or Feature y , but not both." The probability of conjoint occurrence, of course, can be manipulated from 0 through 1.0, the latter extreme being the inclusive disjunctive: "Members

of the category must have Feature *x* or Feature *y* or both." The question at issue is, what will the subject designate as the prototypical stimulus and how will he/she judge degree of category membership as conjoint probabilities change from 0 through intermediate values to 1.0? One possible reading of prototype theory is that, even in the case of 0 conjoint frequency (an exclusive disjunctive rule), the subject should choose, as the "best instance," the stimulus that contains both defining features, all other things being equal, and should rank other stimuli for category membership accordingly. In contrast, feature/schema theory predicts no fixed evaluation of instances except that instances with Feature *x* or Feature *y* are equally representative of the category; the selection of *xy* as the "best instance" should increase in probability as the conjoint occurrence of *xy* positive increases. But the more interesting questions are empirical. Will there be evidence for typicality functions when the core concept learned by the subject is logical, well-defined, and discrete? How will these functions vary with changes in the conjoint probability of relevant features in the positive category?

METHOD

Subjects

One hundred and twenty introductory psychology students participated in the experiment in partial fulfillment of a course requirement. The data of three additional subjects were discarded because of experimenter error or apparatus malfunction during a session.

Design

The design was a 5 by 2 by 2 factorial, including five percentage levels for *xy* instance positive (0, 25, 50, 75, 100), two stimulus populations (one consisting of 72 unique geometric designs and the other of only 57 from a total of 81 possible designs), and two different problems, each with a different pair of relevant features (one and red vs. large and triangle). All variables were manipulated between subjects, yielding 20 conditions. Six subjects were randomly assigned to each of these cells.

Procedure

Subjects were run individually during a 60-min session. They were told that a series of geometric designs would be presented on the screen in front of them. Each design was either a positive or a negative instance of a concept, as indicated by a signal light on a feedback-response panel. The four dimensions of the stimuli and their three values were described: color (red, yellow, black), size (small, medium, large), number (one, two, three), and shape (square, hexagon, triangle). The two relevant dimensions were named for subjects, and the general rule for classifying stimuli was described in truth table terms. In effect, subjects were told that one value or feature from each of two relevant dimensions determined stimulus-to-category assignments, that $\bar{x}\bar{y}$ and $\bar{x}y$ (stimuli with only one relevant feature) were always positive, that $\bar{x}\bar{y}$ (stimuli with neither relevant feature) were always negative, and the *xys* (stimuli with both relevant features) might be assigned to either category. Subjects were instructed that, even though the feedback for *xy* instances might not appear to be consistent, they should still try to use this information to decide whether or not *xy* instances were examples of the concept. The subjects, of course, had to determine from feedback which values or features on the two relevant dimensions defined the concept.

Two practice runs were given, using the alternative concept problem as an example (e.g., if "one red" was the actual concept, then "large triangle" was used for the practice trials). During the first practice run, subjects were asked to categorize positive and negative instances, knowing both the relevant features and the rule; no *xys* were shown. For the second run, the experimenter explained possible strategies to find the two relevant features. These strategies were mainly to use a process of elimination and to look for common features in positive instances.

The rest of the experiment consisted of an acquisition series of six study-test cycles, followed by four postacquisition tests. The study sequences consisted of eight slides shown for 6 sec each with feedback as to category, positive or negative. No response was required to these slides; the subjects were to use the feedback to try to identify the two relevant features. Two examples from each of the four truth table classes appeared in each study sequence. The test sequence required the subject to classify four slides, one of each truth table category, as either positive or negative, using designated buttons on the feedback-response panel. No feedback was given for test responses. All stimuli within a single study-test cycle were different. Because the size of truth table classes differed, however, it was necessary to repeat stimuli across cycles to meet the sequence constraints outlined above.

The following postacquisition tests were given, in the same order to all subjects.

Typicality rank ordering. Four stimulus cards, one example from each truth table class, were laid before the subject. He/she was asked to rank the cards from the best example of the concept to the worst. Six sets of four cards were rank ordered. Randomized across the six sets were two novel stimuli (not seen during acquisition) from each truth table class.

Speeded classification. The subject was asked to classify 20 slides as examples or nonexamples of the concept learned as rapidly as possible. There were five stimuli, one of them being novel (used neither during acquisition or the rank-ordering task), for each truth table class. A ready signal was given before the first slide. All slides remained on the screen until the subject responded. The interslide interval was approximately 1 sec.

Frequency estimates. On a scale from 0 to 50, subjects were asked to estimate how likely each of 15 named values or value combinations was to be assigned to the positive category of the concept. The six individual values and the nine value pairs from the two relevant dimensions were presented in random order for estimation.

Pairwise comparisons. Subjects were given 24 pairs of stimuli and were asked to choose the better example of the concept in each pair. The 24 different comparisons represented six truth table combinations, each occurring four times. The six combinations were the product of each truth table class paired with every other class (e.g., *xy* vs. $\bar{x}\bar{y}$).

Apparatus

A Kodak Carousel slide projector was interfaced with solid state circuitry designed to control sequencing, timing, and feedback. Slide sequences and feedback were read from punched tape. A printout timer recorded all responses and response times in milliseconds.

RESULTS

Acquisition

For the results of posttest to be meaningful, it is important that all subjects, regardless of experimental condition, reach a comparable and stable level of response, at least to those stimuli for which feedback is systematic. In our analysis, we examined responses to *xy* stimuli separately from the remaining stimulus classes, $\bar{x}\bar{y}$, $\bar{x}y$, and $x\bar{y}$. Feedback with respect to the proper

response for each of the last three classes was regular for all experimental conditions during study trials. Feedback for xy instances differed across conditions, being stable for Condition 0 (xy instances always negative) and Condition 100 (xy instances always positive), but variable for Conditions 25, 50, and 75. Unless otherwise indicated, all effects described below are reliable at $p < .01$.

Percent correct responses over the six tests during acquisition were analyzed for $x\bar{y}$, $\bar{x}y$, and $\bar{x}\bar{y}$ classes only. There were three significant main effects, ordinal position of test [$F(5,700) = 21.76$], truth table class [$F(2,1700) = 5.04$], and condition [$F(4,100) = 3.74$]. The ordinal position effect merely confirms the existence of an acquisition function. On the average, percent correct responses was 47% on the first test and 100% on Test 6. The difference among truth table classes is attributable primarily to performance on $\bar{x}\bar{y}$ instances being significantly better than performance on $x\bar{y}$ and $\bar{x}y$ instances during early tests. The condition effect demonstrates that acquisition took place more rapidly when feedback was consistent (Conditions 0 and 100) than when feedback was inconsistent (Conditions 25, 50, and 75) for xy instances. Further, Condition 100, in which xys were called positive 100% of the time, led to more rapid acquisition than Condition 0, in which xys were called negative 100% of the time. Recall that Condition 100 creates an inclusive disjunctive concept and that Condition 0 creates an exclusive disjunction. Exclusive disjunctions are well-known to be more difficult to learn and to use than inclusive disjunctions (Salatas & Bourne, 1974). While not critical, this result is clearly consistent with the general tenets of schema frequency theory in that the representation of an inclusive disjunction is based on two-point density functions (for Dimensions X and Y), which should be easier to establish than the corresponding distribution functions required by an exclusive disjunction.

There were two important interactions, and the pertinent data are shown in Tables 1 and 2. First, class of stimulus instance interacted with ordinal position of acquisition test [$F(10,1700) = 3.80$]. The interaction results from the fact that percent correct response on Test 1 to $\bar{x}\bar{y}$ instances was 61% but only 40% for $x\bar{y}$ and $\bar{x}y$ instances. Subjects achieved the same level of performance, 100% correct responses, on

Table 1
Probability of a Correct Response Averaged
Across Conditions During Acquisition

Class*	Ordinal Position of Test					
	1	2	3	4	5	6
$x\bar{y}$	.405	.533	.732	.866	.975	1.000
$\bar{x}y$	.393	.518	.755	.838	.985	1.000
$\bar{x}\bar{y}$	.605	.714	.819	.895	1.000	1.000

*Truth table class.

Table 2
Probability of a Correct Response on $x\bar{y}$, $\bar{x}y$, and $\bar{x}\bar{y}$
Instances During Acquisition

Condi- tion	Ordinal Position of Test					
	1	2	3	4	5	6
0	.485	.591	.787	.869	.995	1.000
25	.477	.573	.726	.822	.983	1.000
50	.459	.567	.728	.808	.977	1.000
75	.468	.500	.752	.822	.987	1.000
100	.475	.683	.808	.990	1.000	1.000

Test 6. The interaction, then, evidences the convergence of performance functions for truth table classes over trials. Condition interacted with ordinal position [$F(20,1700) = 3.07$]. This interaction is due essentially to the fact that performance improved more rapidly over tests for Condition 0 and Condition 100, compared with the intermediate conditions. For all groups, percent correct response was approximately 47% on Test 1 and 100% on Test 6. Conditions 0 and 100 diverged and achieved higher performance levels than Conditions 25, 50, and 75 on intermediate tests.

An analysis of percent correct response on $x\bar{y}$, $\bar{x}y$, and $\bar{x}\bar{y}$ classes during the last two tests only revealed no significant effects. This outcome suggests that subjects in all conditions and over all stimulus classes achieved the same acquisition level and the same understanding of the concept, at least as it applies to stimuli for which feedback was consistent.

Over the entire series of acquisition tests, xys were called positive instances on .35, .59, .70, .80, and .96 of the test trials by subjects in Conditions 0, 25, 50, 75, and 100, respectively. These differences are highly reliable [$F(4,100) = 5.44$]. There was a strong initial tendency to call xys positive by subjects in all conditions, the proportion averaging .81 on the first test administered. Subjects learned, however, to adjust this response in light of feedback provided over study trials. Thus, on the last two tests, the proportion of xys called positive was .01, .31, .59, .79, and 1.0 by subjects in Conditions 0, 25, 50, 75, and 100, respectively. These differences were also statistically reliable [$F(4,100) = 14.80$]. In an analysis of these data, ordinal position of tests [$F(5,500) = 3.17$] and the interaction of Conditions by Ordinal Position [$F(20,500) = 5.44$] were statistically significant. These performance functions are reminiscent of probability matching functions observed in a variety of simpler learning tasks (Estes, 1976).

Posttests

Different stimulus populations were sampled during acquisition to provide differing degrees of access to old or previously seen vs. new instances on the posttests. Throughout the analyses of posttest data, no differences

were detectable between these populations or between new and old stimuli, possibly because so few novel stimuli were available from either population. But, because there were no differences, all data and analyses reported will ignore the differences between old and new stimuli and between stimulus populations.

Typicality rank ordering. For each subject, we computed the average rank of all xy , $x\bar{y}$, $\bar{x}y$, and $\bar{x}\bar{y}$ instances presented in the six sets of four cards that subjects were asked to rank order. In an analysis of variance on these data, the difference among conditions contributed practically no variance, as one would expect, given that the sum of rank was a constant across conditions. The difference among classes of stimuli was highly significant, with xy stimuli, on the average, assuming the top rank, $\bar{x}\bar{y}$ stimuli falling at the lowest rank, and $x\bar{y}$ and $\bar{x}y$ stimuli ranking in between [$F(3,300) = 966.26$]. Large as it was, the effect was rendered relatively uninteresting by the existence of a Stimulus Class by Condition interaction [$F(12,300) = 3.18$], shown in Table 3. The $\bar{x}\bar{y}$ stimuli obtained a relatively constant and low rank across conditions. The mean rank of xy instances increased, and, compensatingly, the rank of $x\bar{y}$ and $\bar{x}y$ instances, which did not differ, decreased across conditions. For Condition 100, subjects were almost entirely consistent in their ranks, placing xy at the top, $\bar{x}\bar{y}$ at the bottom, and $x\bar{y}$ and $\bar{x}y$ at Ranks 2 and 3. In Condition 0, $x\bar{y}$ and $\bar{x}y$ obtained, on the average, tied ranks, with xy instances next and $\bar{x}\bar{y}$ instances ranked lowest. Ranks changed systematically as the percent of xys in the positive category increased. Thus, even with relatively modest feedback to the effect that xy instances are sometimes positive, subjects, on the average, ranked xy instances as the best example of the concept that had been learned.

Speeded classification. Performance on this task was relatively unenlightening. A correct classification of xy instances would place them in the negative category in Condition 0 and in the positive category in Condition 100. The proper classification of xy instances for intermediate conditions was "indeterminant." As in the final two tests administered during acquisition, however, subjects tended to divide xy instances between the positive and negative category according to the percentages of feedback administered during acquisition. Error rates can be determined only for Condition 0

Table 4
Mean Estimates of Frequency Within Positive Instances

Condition	Single Features		Two-Feature Combinations		
	R	I	Both R	One R	Both I
0	30.2	19.1	5.7	39.0	.5
25	37.2	20.6	20.5	45.2	.0
50	43.7	20.0	30.9	48.5	1.3
75	47.7	18.8	41.6	49.2	.0
100	49.5	19.5	50.0	48.8	.8

(viz., .1) and Condition 100 (viz., .03) for xy instances. The proportions of xy instances called positive were .31, .57, and .81 for Conditions 25, 50, and 75, respectively. Reaction times to xy stimuli on correct response trials in Condition 0 (xy negative) and in Condition 100 (xy positive) were 3.86 and 3.21 sec, respectively. Reaction times taken over all trials showed no significant trend attributable to conditions. Reaction time in Condition 100 (3.41 sec) was, however, significantly faster than the pooled reaction times for the remaining conditions (3.75 sec) [$F(1,100) = 3.52$].

Frequency estimates. Subjects were required to estimate on a scale from 0 to 50 how often each of 15 features or feature combinations had occurred in positive instances during acquisition. All individual features and combinations on the two relevant dimensions were presented for estimation. The mean obtained values over five conditions are presented in Table 4. In an analysis of variance of these data, there were two effects of significance, features (or feature combination) [$F(14,1400) = 160.86$] and the interaction of features and conditions [$F(56,1400) = 3.32$]. The meaning of these effects is readily apparent from Table 4. Frequency estimates increased systematically over Conditions 0-100 for the individual relevant features and for the combination of the two relevant features, the combination showing a steeper rise because of the lower estimate in Condition 0. There was no significant trend in non-relevant individual values on relevant dimension or in combinations of one relevant and one irrelevant value. When neither value of a pair was relevant, estimates on the average were close to 0. These estimates are essentially what one would expect, given that all subjects mastered the concept during acquisition. Estimated frequencies were sensitive to objective frequencies, both for individual features and combinations. There was a definite tendency, however, to overestimate feature category contingencies, which is consistent both with the acquisition data and with probability matching functions (Estes, 1976).

It is potentially of some importance to ask how accurate these estimates were and whether or not accuracy correlated in any way with concept acquisition. We constructed an accuracy measure by converting frequency estimates to a probability and then taking the difference between estimated and actual probability

Table 3
Typicality Rank Order of Stimulus Instances Across Conditions

Stimulus Class	Condition				
	0	25	50	75	100
xy	2.89	1.58	1.32	1.11	1.00
$x\bar{y}$	1.81	2.12	2.21	2.58	2.52
$\bar{x}y$	1.72	2.25	2.31	2.32	2.51
$\bar{x}\bar{y}$	3.83	3.99	4.00	3.98	3.97

of a certain event, given the sequence of acquisition stimuli presented. Of major interest in the analysis of these data were the correlations between accuracy of estimation and number of errors committed on acquisition test trials. Two correlations were statistically reliable, one based on the occurrence of either relevant feature [$r = .390$, $t(118) = 3.17$], and the other on the occurrence of both relevant features [$r = .330$, $t(118) = 2.95$], suggesting that the accurate accumulation of frequency information bears some relationship to the ability to form relatively complex concepts.

Pairwise comparisons. We used as primary data the number of times each member of a particular pair was chosen as the better example of the concept learned. Reducing the various stimuli to their truth table classifications, there were six possible comparisons. Those six comparisons were analyzed independently in canonical order as a within-subjects variable. For purposes of more efficient comparison, however, we combined certain pairs in the presentation of data in Table 5. To be specific, because there was no reason to expect any difference between the xy - $x\bar{y}$ and the xy - $\bar{x}y$ comparison or between the $\bar{x}y$ - $x\bar{y}$ and the $\bar{x}y$ - $\bar{x}y$ comparison, these pairs were combined. The values presented in Table 5 reflect the average frequency with which the first member of each pair was selected as a better example of the concept.

In the analysis of variance, three effects of interest were statistically reliable: condition [$F(4,100) = 3.58$], pair [$F(5,500) = 131.83$], and the interaction of these variables [$F(20,500) = 2.77$]. These data can be described as follows. Subjects are indifferent to the comparison between $x\bar{y}$ and $\bar{x}y$ stimuli, choosing one or another with roughly equal frequency regardless of experimental conditions. Also regardless of condition, subjects picked with high frequency either the $x\bar{y}$ or the $\bar{x}y$ as a better example than $\bar{x}\bar{y}$ stimuli. Condition, as one would expect, produced its major effect on comparisons involving xy stimuli. Even in Condition 0, xy stimuli were picked as better examples than $\bar{x}\bar{y}$ stimuli, although both were equally nonexemplary of the concept. This preference for xy stimuli increased as a percentage of xys placed into the positive category increased across conditions. In the comparison of xy with $x\bar{y}$ or $\bar{x}y$ stimuli, there was an important crossover. In Condition 0, $x\bar{y}$ or $\bar{x}y$ stimuli were chosen more frequently as

better instances than xy , although the choice of xy as a better instance had some significant value even in Condition 0, in which xy instances were always negative. The point of indifference between xy and $x\bar{y}$ or $\bar{x}y$ instances lay between Conditions 50 and 75, despite the fact that $x\bar{y}$ and $\bar{x}y$ instances were always positive and xy instances were positive only 50%-75% of the time. In Condition 100, xy instances were almost always chosen in preference to $x\bar{y}$ and $\bar{x}y$ instances, even though all three instance classes appeared with the same regularity in the positive category.

DISCUSSION

A concept based on some well-defined rule, even a rule that integrates only a pair of relevant stimulus features, yields certain phenomena not unlike those observed by Rosch (1978) for so-called natural concepts. Subjects in Conditions 0 and 100 of the present experiment learned easily and to a high criterion to place stimuli into positive and negative categories based on a logical rule that integrates two stimulus features. Despite their training and their knowledge of the exclusive disjunctive rule, subjects in Condition 0 nevertheless exhibited a tendency to respond to a stimulus with both features (always negative under the rule) as a better example than a stimulus with neither feature (also negative under the rule). That tendency was even more obvious in other conditions, in which at least some xy instances were assigned to the positive category. The potency of this conjunction of relevant features is clearly revealed in various posttest performances, even when only a small percentage of stimuli with both critical features is called positive during training.

Thus, we conclude that, even when the subject learns a well-defined rule based on primitive stimuli, there are underlying processes that generate performance on the surface reflecting a dimension of instance goodness. This conclusion clearly relates to the argument of Osherson and Smith (1981) that prototype theory is fundamentally limited. Subjects might acquire a well formed core concept and, at the same time, base membership responses on the presence and absence of surface features. Processes involved in this phenomenon might derive from the formation of a concept-specific schema during training. The subject enters the problem with a generalized schema containing slots that act in the task as requests for information. Some of the slots are identified by known stimulus dimensions, established during careful and detailed instructions given at the outset. One kind of information, but by no means the only kind, that can be requested and recorded in these slots corresponds to category contingencies or frequency of feature occurrence within categories. As the subject records this information, dimensional distinctions begin to emerge and to determine feature relevance. Under all conditions of the present experiment, Dimen-

Table 5
Average Number of Times the First Member of Each Pair
Was Chosen as the Better Example of the Concept

Pair	Condition				
	0	25	50	75	100
xy - $\bar{x}y$ ($x\bar{y}$)	1.091	1.625	2.137	2.833	3.542
xy - $\bar{x}y$	2.954	3.500	3.875	4.000	4.000
$x\bar{y}$ - $\bar{x}y$	2.037	2.208	2.054	1.833	1.922
$x\bar{y}$ ($\bar{x}y$)- $\bar{x}\bar{y}$	3.708	3.792	3.955	3.833	3.542

sions X and Y, individually and/or in combination, are distinguished from all other dimensions by nonrectangular frequency distributions (over the three values of each) in the positive category. Instructions allow the subject to encode Dimensions X and Y as relevant. Study instances establish the frequency or probability with which particular features, x_i and y_i , on these dimensions occur within the positive category. Frequency differentials allow the subject to identify the relevant features on each dimension. These features can be plugged into the classification rule given during instructions to define the core concept. Note that, in this study, there is uncertainty in the application of this rule only with respect to xy instances, and only there in Conditions 25, 50, and 75.

At least for the disjunctive rule used in this study, such an encoding of the concept generates naturally the phenomena of "best instance" and typicality that appear ubiquitously in studies of natural concepts. The typicality gradient is determined, as implied by Neumann (1977), by variations in the actual frequency with which all values on each dimension and value combinations were assigned to the positive category during acquisition. "Goodness" of any stimulus is indexed by the likelihood (summed frequencies) of its particular features appearing in the positive category, taken over all study instances. Both schema theory and the position of Osherson and Smith (1981) imply that it is not concept fuzziness that generates these observations. Rather, it is the application of a clear and parametrically well-defined core concept to instances whose feature structure varies probabilistically so as to produce these effects.

Contrary to Rosch (1978), we argue that simplicity or logical structure in concepts on conceptual tasks does not produce unnatural phenomena. Rather, the phenomena are essentially the same in natural and logical concepts. Furthermore, we claim that the notion of typicality, rather than being fundamental to concepts, is derivative of more basic knowledge, which often can best be examined in the laboratory.

REFERENCES

- BRUNER, J. S., GOODNOW, J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- BOURNE, L. E., JR., EKSTRAND, B. R., LOVALLO, W. R., KELLOGG, R. T., HIEW, C. C., & YARROUSH, R. A. Frequency analysis of attribute identification. *Journal of Experimental Psychology: General*, 1975, **106**, 294-312.
- ESTES, W. K. The cognitive side of probability learning. *Psychological Review*, 1976, **83**, 37-64.
- GARNER, W. R. Aspects of a stimulus: Features, dimensions and configurations. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. New York: Wiley, 1978.
- HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, **72**, 175-195.
- KOZMINSKY, E., KINTSCH, W., & BOURNE, L. E., JR. Decision making with texts: Information analysis and schema acquisition. *Journal of Experimental Psychology: General*, in press.
- NEUMANN, P. G. Visual prototype formation with discontinuous representations of dimension of variability. *Memory & Cognition*, 1977, **5**, 187-197.
- OSHERSON, D. N., & SMITH, E. E. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 1981, **9**, 35-58.
- ROSCH, E. Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. New York: Wiley, 1978.
- ROSCH, E., & MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, **7**, 573-605.
- ROSCH, E., MERVIS, C. B., GRAY, W. D., JOHNSON, D. M., & BOYES-BRAEM, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, **8**, 382-439.
- SALATAS, H., & BOURNE, L. E., JR. Learning conceptual rules: III. Processes contributing to rule difficulty. *Memory & Cognition*, 1974, **2**, 549-553.

NOTE

1. Because stimuli in this experiment are dimensionalized, we often use the term "value" (on a dimension) to refer to a feature of a stimulus. Indeed, stimulus value and stimulus feature are mutually substitutable throughout this paper.

(Received for publication June 5, 1981;
revision accepted October 26, 1981.)