

Category representations and their implications for category structure

ROBIN A. BARR

Ball State University, Muncie, Indiana

and

LESLIE J. CAPLAN

National Institute of Mental Health, Bethesda, Maryland

In a series of experiments and reanalyses of previous research, we tested the hypothesis that categories that are primarily represented by extrinsic features (i.e., those that are relations between two or more entities) would yield more graded structures than would categories primarily represented by intrinsic features (i.e., those features true of an item considered in isolation). These predictions were confirmed. Extrinsically represented categories showed (1) less agreement across subjects on membership judgments, (2) more graded membership in a membership judgment task, and (3) smaller differences between gradients of typicality and of membership judgments.

Over the last 15 years, research into natural language category representation has been strongly influenced by the work of Rosch and her colleagues (e.g., Rosch, 1973, 1975; Rosch & Mervis, 1975). The essential finding in this research was that many natural language categories show graded membership functions. Some members are reliably rated as better examples of a category than are others (Rips, Shoben, & Smith, 1973; Rosch, 1973), and some members are produced and verified as category members more often and more quickly than are others (Battig & Montague, 1969; Glass, Holyoak, & O'Dell, 1974; Rips et al., 1973; Rosch, 1973; Wilkins, 1971). In addition, such categories include "borderline" members, that is, items over which subjects disagree with each other regarding category membership, and over which they are inconsistent from day to day in assigning membership (McCloskey & Glucksberg, 1978).

These demonstrations of graded membership generated many theories proposing that category representations are fuzzy. In other words, no membership criteria exist that clearly discriminate all members from all nonmembers. Such theories included Rosch's own "prototype" approach (Rosch, 1973; Rosch & Mervis, 1975), various "probabilistic" theories (e.g., McCloskey & Glucksberg,

1979), and some "exemplar" theories (e.g., Medin & Schaffer, 1978). All of these models contrasted with the earlier "classical" view that categories could be represented by a set of necessary and sufficient features that described all and only category members (see Smith & Medin, 1981).

Theories designed to explain category "fuzziness," however, have both conceptual and empirical problems. Conceptually, probabilistic and prototype theories require more complex representations than does the classical theory. Prototype theories require a prototype and some rule that permits items that are discrepant from the prototype to be category members. Probabilistic models require a set of features, a set of weightings reflecting the importance of each feature to the category, and some "weighted feature combination rule" to determine category membership.

Empirically, these models have had difficulty explaining natural language category phenomena other than fuzziness. Various investigators have demonstrated that category structure is influenced by context (e.g., Anderson & Ortony, 1975; Anderson et al., 1976; Roth & Shoben, 1983). Probabilistic theories would have to explain these findings by positing changes in feature weightings and, thus, apparent changes in the meaning of the concept across different situations. Prototype theories would have to posit a changed prototype. In addition, such theories considerably complicate explanations of how concepts are combined. For example, Armstrong, Gleitman, and Gleitman (1983) pointed out that even for a classical theory, it is difficult to specify how the features of "good" and the features of "wife" combine to describe the concept "good wife," and how the features of "good" and "knife" combine to yield "good knife." For a probabilistic or prototype theory, the problem is compounded because there are no necessary features associated with the

Some of the ideas for this paper were developed while Robin A. Barr held an Albert J. Sloan Foundation Fellowship in Cognitive Science at the University of Pennsylvania. This research was partially supported by two research grants from Ball State University. The authors thank Paul Bertsch, Chris Ernestes, Denise Gurley, Debra Hricik, C. D. Kapadnis, Marina Koliaas, Karen Midkiff, Patricia Tolle, Charles Topp, and Elizabeth Wade for their help in data collection and analysis. We thank Thomas Foster for much helpful advice, and Alice Healy, Michael McCloskey, and three anonymous reviewers for very helpful comments on earlier versions of this manuscript.

Requests for reprints may be addressed to the first author at National Institute on Aging, National Institutes of Health, Building 31, Room 4C32, Bethesda, MD 20892.

component concepts. For the classical feature theory, the problem lies in choosing the relevant features from a known set of features; for the probabilistic or prototype theory, the problem lies in identifying the set of features to begin with, before proceeding to select the relevant ones.

Some authors have suggested that the difficulties associated with these models may be resolved by mixed models that combine elements of classical and of prototype/probabilistic theories (Armstrong et al., 1983; Landau, 1982; Medin & Smith, 1984). In these models, categories are represented by a classical "core" and a probabilistic or prototype "identification procedure." Such models do help to explain context effects and other phenomena that pose empirical problems for the pure probabilistic and prototype theories. However, they do so at the cost of further expanding and complicating the nature of the presumed representation (see Armstrong et al., 1983). Not only are there two aspects to representation, but such theories need to specify when each aspect (or combination of aspects) is invoked.

Another account of context effects in categorization has been offered by Barsalou (1985), who suggested that categories may not have invariant structures at all (whether probabilistic or classical). Instead, he argues that people's structuring of categories is dynamic and flexible, and thus changes with context. Although this explanation does, of course, account for the recent work on context effects, it is difficult to reconcile with the traditional role subsumed by categories, namely that they simplify the infinite variety of stimuli that we encounter by grouping together objects with properties in common (see Rey, 1983). In the model described below, we offer a way in which context effects can be explained, but in which the idea that categories describe common and invariant sets of properties among items is maintained.

In this paper, we demonstrate how this model of category representation (Barr & Caplan, 1985, 1986) shows that a category represented in the classical way may nevertheless give rise to a fuzzy extension (i.e., graded membership, disagreement among subjects over membership, etc.). If this model is correct, then the parsimonious aspects of classical models may be retained, and their explanatory power can be expanded.

A Model of Properties of Natural Language Representation

The model describes properties of an individual's representation of a category or concept. In particular, it explains how different kinds of features can yield a number of well-known empirical phenomena. In this paper, we focus on how the nature of the features that individuals use to represent a category can affect their extensions of that category (i.e., the set of entities that they consider to be category members).

The model distinguishes between two kinds of features: *intrinsic* features and *extrinsic* features. A feature is intrinsic when it is represented as being true of an entity

considered in isolation. For example, "has wings" might be considered an intrinsic feature of a bird. Intrinsic features are true of the object itself. Consequently, when a feature is represented as intrinsic, a change in the feature involves a change in the object itself. Thus, if a bird were to lose its wings, then a subject who represented "has wings" as intrinsic would believe that the bird itself had changed. A feature is extrinsic when it is represented as the relationship between two or more entities. For example, "used to work with" might be considered an extrinsic feature of a hammer, because it describes the relationship between the hammer and a worker. Extrinsic features are not true of the object itself. Therefore, when a feature that an individual holds to be extrinsic changes, the object itself does not change. If, for example, a hammer were no longer used to work with, the hammer itself would remain the same.

Whether a feature is held to be intrinsic or extrinsic is a judgment about an *individual's* representation of the feature and the entities involved. For example, "is purple" might be considered an intrinsic feature of a grape by an individual, in which case he/she represents "being purple" as being true of the grape in isolation. Another individual, however, may represent "purple" as the relationship between the grape and the ambient light; for this individual, "purple" is an extrinsic feature. It should be noted that being intrinsic or extrinsic is true of an individual's representation of a particular entity, not of the whole class. Thus, when an individual represents "purleness" as intrinsic of grapes, he/she does not necessarily think that all grapes are purple.

The criteria by which category membership is determined for some natural language categories seem to rely primarily on intrinsic features, whereas the criteria for other natural language categories seem to rely primarily on extrinsic features. According to one dictionary definition (*Webster's New World Dictionary*, 1980), "vehicles" are "devices or contrivances for carrying or conveying persons or objects"; for most people, this definition would involve extrinsic features. On the other hand, the definition of "bird" is "warm-blooded, two-legged, egg-laying, vertebrate with feathers and wings"; most individuals would judge most or all of these features to be intrinsic. (Later in the paper we examine the extent to which generated features of these and other categories are judged to be intrinsic or extrinsic.)

We have demonstrated elsewhere (Barr & Caplan, 1985) that extrinsic features give rise to extensions showing graded membership. Suppose that a particular object is being considered for membership in some category and that the category is represented by a single extrinsic feature (which is the relationship between the object and some other entity or entities). In that case, whether or not the object is judged to be a member of the category is going to depend not on characteristics of the object itself, but instead on the relationship it holds with the other entity or entities. In some situations the object may hold the appropriate relationship, and in other situations it may not

hold this relationship. For example, suppose that a given person represents the category Weapon by the extrinsic feature "used to hurt people." This individual considers to what extent "bomb" is a weapon. In most situations in which the individual can imagine bombs occurring, they are used to hurt people. Therefore, this person considers "bomb" to be a very good member of the Weapon category. On the other hand, when asked to judge to what extent "rubber band" is a weapon, the individual can think of only a few situations in which rubber bands are used to hurt people and many situations in which they are not used to hurt people. Therefore, the person considers "rubber band" to be a "borderline" example of Weapon.

Categories represented primarily by intrinsic features will have different structures. Suppose that membership in the category Vertebrate were determined only by the feature "has a segmented spinal column." An individual asked to judge whether an item is a member of this category needs to consider only whether it possesses this feature; the item's relationships with other entities are irrelevant. Items possessing this feature are clearly members of the category; items that do not possess this feature are not members. For this type of category, the item possesses the feature in almost all conceivable situations; the object must itself change to lose its category membership. Thus, this type of representation will generate few borderline members.

In this paper, we tested this model of category representation by (1) asking subjects to select the defining features of a set of categories (Experiment 1), (2) evaluating the extent to which these features were informative as to the nature of the category (Experiment 2), and (3) asking subjects to categorize the most useful features as intrinsic or extrinsic (Experiment 3). Then we tested the predictions of the model with respect to category fuzziness by (1) re-analyzing published data on graded category extensions, and (2) asking subjects to rate how clearly items from a selected subset of exemplars were members of categories (Experiment 4). Finally, because the model makes predictions about the extent to which typicality gradients reflect true membership gradients, we asked subjects to rate the typicality of the same exemplars used in Experiment 4 and compared the results of the two tasks (Experiment 5).

The categories we chose to work with were Birds, Clothing, Flowers, Fruit, Furniture, Mammals, Metals, Sports, Tools, Toys, Trees, Vegetables, Vehicles, and Weapons. These categories were chosen because they have been heavily investigated elsewhere (e.g., Battig & Montague, 1969; Hampton & Gardiner, 1983; McCloskey & Glucksberg, 1978; Rosch, 1975). Consequently, we were able to test predictions derived from our data using the published results of other investigators. In this way, it was possible to show that any effects that we found are not peculiar to our data or subject population. However, it should be noted that data from the category Metals had to be eliminated from our later analyses. Our subjects were unable to generate enough examples of the category to provide a sufficient data base to test our predictions.

PART 1: FEATURE COLLECTION

Experiment 1

In this experiment, we asked subjects to construct category definitions from a large set of features for 14 natural language categories: Birds, Clothing, Flowers, Fruit, Furniture, Mammals, Metals, Sports, Tools, Toys, Trees, Vegetables, Vehicles, and Weapons.

Method

Subjects. Fifty-two undergraduate volunteers served as subjects. Some of the students received course credit for participation.

Stimuli. In order to obtain large sets of attributes, we asked another group of 15 undergraduates to write down attributes that they considered to be generally true for each of the 14 categories. Stimulus booklets were then compiled, including (1) an instruction page, and (2) 14 pages, each headed with a category name and containing the list of attributes generated by the earlier 15 subjects. (The category Metals was later eliminated because subjects were unable to generate more than a few exemplars in later experiments.) The attribute lists had been modified slightly to eliminate redundancy and to reduce ambiguity to the best of our ability. The presentation order of both attributes and categories was randomized.

Procedure. Subjects were tested in small groups. They were each given a copy of the stimulus booklet. They were instructed to select and circle, for each category, those features that together constituted the definition of that category. Subjects proceeded through this task at their own paces.

Results

The results described below were obtained from 50 of the 52 subjects. Two subjects' data were eliminated because these subjects had circled at least 90% of all of the attributes listed for all 14 categories.

The features chosen by at least 50% of the subjects for each of the 13 categories (features for the category Metals have been excluded) are presented in Table 1. These features were used in Experiments 2 and 3, and to form predictions for Experiments 4 and 5.

Experiment 2

Many of the features chosen by subjects in Experiment 1 to describe category members did not appear to be particularly informative as to the nature of the category (e.g., "there are different kinds," "there are different varieties," "there are different styles"). In this experiment, therefore, we tried to discover those features that were maximally informative as to the nature of the category. Subjects were presented with the most frequently selected features from Experiment 1 and were asked to determine the category described by those features.

Method

Subjects. Forty-four undergraduate volunteers participated in this experiment, some of whom received course credit for their participation.

Stimuli. The stimuli were sets of 10 features (the 10 most frequently selected features from Experiment 1), written on pages of a small booklet, 1 feature to a page. Equal numbers of features were used per category in order to give subjects an equal number of opportunities to identify each category, and to equate, across categories, the number of times any particular feature occurred in

Table 1
Features Selected by at Least 50% of the Subjects for Each of 13 Categories (Experiment 1)

Feature	Proportion of Subjects Who Chose that Feature as Defining (Experiment 1)	Proportion of Subjects Who Identified the Category on that Feature (Experiment 2)	Proportion of Judges Who Marked the Features as Extrinsic (Experiment 3)
Birds			
Have feathers	.70	.90	0.00
There are various species	.66	.25	.67
Have beaks	.64	1.00	0.00
Lay eggs	.58	1.00	.29
Are hatched from eggs	.52	.72	.43
Build nests	.50	1.00	.86
Mean across features	.60	.81	.38
Clothes			
Made of different materials	.62	.40	.40
Cover people	.58	.91	1.00
There are different varieties	.56	0.00	.50
Have different textures	.54	0.00	.67
Made of cloth	.52	1.00	0.00
Mean across features	.56	.46	.51
Flowers			
Have petals	.64	.90	0.00
Need water	.60	0.00	.57
Have roots	.52	.50	0.00
Are colorful	.52	.50	.71
They grow	.52	.80	.29
Are beautiful	.50	.67	.83
They bloom	.50	1.00	.57
Mean across features	.54	.62	.42
Fruit			
Are nutritional	.62	.06	.57
Contain vitamins	.60	.69	0.00
Are seasonal	.54	.27	.71
Need sunshine	.54	0.00	.71
Have many uses	.54	0.00	1.00
Have a color	.50	0.00	0.00
Mean across features	.56	.17	.50
Furniture			
Has a function	.70	.13	.43
There are different styles	.70	0.00	.50
There are different kinds	.62	.13	.75
Has different sizes	.62	0.00	.67
Adds to house decor	.54	.80	.60
Used inside houses	.52	.54	.86
Mean across features	.62	.27	.64
Mammals			
Are warm blooded	.66	.38	0.00
Are air breathing	.58	.40	.43
Reproduce sexually	.58	.72	.50
Give birth to live young	.56	.50	.57
Have hair	.54	.25	0.00
Have eyes	.54	.86	0.00
Mean across features	.58	.52	.25
Sports			
Have rules	.78	.17	.14
Athletes participate	.74	.67	.60
Have many varieties	.70	.20	.33
Are competitive	.68	.50	.86
Involve physical activity	.64	.60	.83

Table 1 (Continued)

Feature	Proportion of Subjects Who Chose that Feature as Defining (Experiment 1)	Proportion of Subjects Who Identified the Category on that Feature (Experiment 2)	Proportion of Judges Who Marked the Features as Extrinsic (Experiment 3)
Are played in different seasons	.62	.88	.80
Are challenging	.62	.20	.43
They keep people fit	.56	.29	.86
Are suitable for all ages	.56	0.00	.67
They involve discipline	.54	.10	.57
Mean across features	.64	.36	.61
Tools			
Are used to fix things	.78	.91	.86
Are used to make tasks easier	.64	.86	1.00
Are used to create	.56	.25	.86
Are useful	.54	.37	1.00
Are used to build things	.50	1.00	1.00
Mean across features	.60	.68	.94
Toys			
There are many kinds	.70	0.00	.50
Are used to entertain and amuse	.66	.15	1.00
They give happiness	.62	0.00	1.00
Mean across features	.66	.05	.83
Trees			
There are different kinds	.70	0.00	.50
Have roots	.60	.50	0.00
Have leaves	.54	.72	0.00
Are alive	.52	.15	0.00
Grow outdoors	.50	.30	.71
Have a trunk	.50	.33	0.00
Mean across features	.56	.33	.20
Vegetables			
Have different shapes	.64	.03	.20
Contain many vitamins	.60	.25	0.00
Are a basic food group	.58	.71	.50
They grow	.54	0.00	.29
Are good for the body	.52	.59	.86
There are different kinds	.52	.02	.50
Need sun	.50	.93	.57
Mean across features	.56	.36	.42
Vehicles			
Are a means of transportation	.80	.88	.71
Can be driven	.68	1.00	.86
Become modified by technology	.50	.50	1.00
Mean across features	.66	.79	.86
Weapons			
Are unsafe in the wrong hands	.70	.89	1.00
Are used for protection	.58	.75	.86
There are different kinds	.56	.29	.50
Are used in crime	.50	.72	.86
Mean across features	.59	.66	.81

a particular position in the sequence of 10. Subjects received one booklet for each of the 14 categories used in Experiment 1; 10 orderings of the feature list, determined by a Latin square, were used for each category.

Procedure. Subjects were told that they would receive 14 booklets, each containing 10 features, and that the features in each booklet described one particular category. (Subjects were not given any information about the set of categories, other than the information gained from the features themselves.) They were instructed to go through the booklets, page by page. On each page, they were to write down their idea of what was being described by the features. Subjects were encouraged to guess, even if they had no particular idea as to the category, and to write down a response on each page before turning to the next page. They were permitted to look back over features on earlier pages, but they were not permitted to look forward beyond the current page. Finally, they were encouraged to think of the broadest, or most general, category that would fit the features. Each subject received the 14 booklets in a different random order.

Results

Column 2 of Table 1 shows the proportion of subjects who identified the category correctly for the first time, when they encountered that feature. For example, the proportion .86 for the feature "have eyes" in the category Mammals means that, of the subjects who had not previously identified the category, 86% successfully identified it on that feature, and 14% did not identify it on that feature. In the results shown in Table 1, we counted as a correct response either (1) the appropriate category label, or (2) the label of a subordinate of the category. For example, if a subject was working with the features of Furniture, his/her response would be scored as correct if he/she responded with a subordinate such as "bed," since the features listed were, in fact, true of subordinates as well as of the categories targeted by the experimenters. (Although such responses may at first appear to be incorrect, they are correct from the subject's perspective. Other literature on the "confirmation bias" in reasoning [Wason, 1960] suggests that subjects do, in fact, stop searching for alternative responses when the response they are currently considering is repeatedly confirmed. Such a bias could easily have led subjects in this experiment to stop searching after generating a subordinate.) Although 10 features were used per category in this task, the only features shown in Table 1 (and the only features upon which subsequent predictions were based) were those originally selected as defining by 50% or more of the subjects in Experiment 1.

Experiment 3

In the model of category representation discussed in the Introduction, we proposed that some of the features used to represent categories are intrinsic (i.e., true of an entity in isolation) and some are extrinsic (i.e., a relation between that entity and one or more other entities). In this experiment, we asked subjects to categorize the most frequently selected attributes in Experiment 1 as intrinsic or extrinsic.

Our goal for this experiment was to derive estimates of the degree to which categories' representations relied on extrinsic features. The estimates were used to make predictions for Experiments 4 and 5, as well as for re-analyses of published data from other investigators (presented in Part 2). To calculate those estimates, we used data from Experiments 1 and 2, as well as data from this study. Our estimates took into account both the extent to which individual features were judged to be extrinsic and the degree to which each feature is represented as criterial for category membership in the population as a whole. For example, when a feature is considered defining by only 50% of the subjects, then it is less criterial in the population as a whole than is a feature that is considered defining by 80% of the sample. Thus a feature that is considered criterial by more subjects is given more weight in our predictions.

Method

Subjects. Seven students enrolled in a graduate-level cognition class participated voluntarily. All of these students were naive to the hypothesis under investigation.

Stimuli. Stimulus booklets first included a page of instructions, which read as follows:

On the following pages you will see a list of characteristics which college students thought described the members of a category. The relevant category name is at the top of the page. Please try to decide whether each of the characteristics listed is INTRINSIC or EXTRINSIC.

INTRINSIC characteristics are true of individual category members when they are considered by themselves. For example, the characteristic "have a tail" is an intrinsic characteristic of dogs because it is true of an individual dog when it is considered by itself. Similarly the characteristic "are made of rubber" is an intrinsic characteristic of balloons since an individual balloon, considered by itself, is made of rubber.

EXTRINSIC characteristics only describe a category member as it relates to another thing or things. For example, "are used to cook with" is an extrinsic characteristic of stoves, since it describes how a stove relates to food and a person. Similarly "swim" is an extrinsic characteristic of fish because it describes how a fish relates to water.

Consider the following characteristics of the category "book." The characteristics are categorized as intrinsic or extrinsic below.

Made of paper—Intrinsic: books considered by themselves are made of paper

Are read—Extrinsic: Books are read by people, so this involves a relationship between books and people. A book by itself isn't read.

Has an author—This may be either intrinsic or extrinsic, depending on your point of view. If you think that the book has an author, even when considered by itself, you would classify this characteristic as intrinsic. If you consider "has an author" to describe the relation between a particular book and a particular person, you would classify this as extrinsic.

Next, consider the following practice category, "computer." Classify the following characteristics as intrinsic, extrinsic, or neither.

Has a memory

Is used to calculate

Runs on electricity

Is very powerful

Uses logical operations

Has silicon chips

You may find it helpful to look back at these instructions as you proceed through the task—please feel free to do so. As you con-

sider each characteristic please check the appropriate column (intrinsic or extrinsic). If you find a characteristic that seems to be neither intrinsic nor extrinsic, check the column labeled as "neither" for that feature.

These instructions were followed by 13 pages, each of which was headed with a category name and contained the features for that category, which are shown in Table 1. The 13 categories were all those used in Experiment 1, with the exception of Metals. Each page also included three columns in which subjects recorded their judgments of the features; these columns were labeled "intrinsic," "extrinsic," and "neither."

Procedure. Subjects were given the stimulus booklets in class by the experimenter, who then reviewed the instructions and practice category with them. Subjects were encouraged to justify their classifications of the features given for the practice category. They then completed the task individually on their own time, and returned the booklets 1 week later.

Results

The proportion of judges who marked each feature as extrinsic is shown in the third column of Table 1. This proportion, referred to as *E* below, was calculated by dividing the number of extrinsic judgments by the number of judgments that were either intrinsic or extrinsic. (Judgments that a feature was neither intrinsic nor extrinsic were not included in these proportions.)

It is clear from Table 1 that some features are rated as extrinsic by most or all subjects, whereas others are rated as extrinsic by no one. It is also clear from the table that extrinsic features include such obviously functional features as "used to make tasks easier." However, other extrinsic features are more difficult to classify as functional features (e.g., "are unsafe in the wrong hands," "become modified by technology," "build nests," and "are beautiful"). In order to test the assumption in our model that extrinsic features lead to graded extensions, we needed to derive some measure of the extent to which a category is represented by extrinsic features. Accordingly, we twice calculated measures of the importance of each feature to the category, once using the data from Experiment 1 (the feature selection task) and the second time using the data from Experiment 2 (the informative feature task). In each case, the formula for calculating the importance of a particular feature to the category is:

$$I = \frac{n_i}{\sum_{i=1}^k n_i} \tag{1}$$

For the measure based on the feature selection task, *n_i* represents the proportion of subjects who selected that feature. For the measure based on the informative feature task, *n_i* represents the proportion of subjects who identified the category correctly for the first time when presented that particular feature. In both measures, *k* represents the total number of features chosen by 50% or more of the subjects in Experiment 1 (i.e., the number of features listed for each of the categories in Table 1). For example, for the category Toys, the value of the feature selection measure of the importance of the feature "used to entertain and amuse" is .66/(.66 + .62 + .70) = .33.

In order to estimate the degree (*W*) to which a category is represented by extrinsic features, we used both the extrinsicity measure and the importance measure described above. The following formula was used;

$$W = \sum_{i=1}^k (E_i \times I_i), \tag{2}$$

where *E_i* represents the degree to which a particular feature was judged to be extrinsic, *I_i* represents the importance of a particular feature to the category (obtained from Formula 1), and *k* represents the number of features chosen by at least 50% of the subjects for a given category.

The value of *W* for a category represents the estimate of the extent to which the representation of that category is extrinsic. *W_s* is the estimate derived from the data generated by the feature selection task; *W_I* is the estimate derived from the informative feature task. These values of *W* are presented in Table 2. For example, for the category Toys, the three features "give happiness," "used to entertain and amuse," and "there are many kinds" had importance weightings of .31, .33, and .35, respectively, based on the feature selection task. The *E* values of "give happiness" and of "used to entertain and amuse" were both 1.00, and the *E* value for "there are many kinds" was .50. Therefore, the *W_s* for Toys is (.31 × 1.00) + (.33 × 1.00) + (.35 × .50) = .82.

As might be expected, the two estimates, based on the two different tasks, are very close to each other for most of the categories. In addition, some categories (e.g., Weapons, Tools) seem to be represented primarily by extrinsic features, whereas other categories (e.g., Trees, Mammals) seem to rely much less heavily on extrinsic features.

*** PART 2: TESTING THE MODEL**

We hypothesized in the Introduction that extrinsically represented categories would yield fuzzy extensions (i.e.,

Table 2
Estimates of the Degree to Which the Representations of 13 Categories Depend on Extrinsic Features

Category	<i>W_s</i>	<i>W_I</i>
Birds	.35	.34
Clothing	.53	.46
Flowers	.40	.35
Fruit	.49	.22
Furniture	.63	.69
Mammals	.25	.26
Sports	.61	.70
Tools	.95	.94
Toys	.82	1.00
Trees	.21	.11
Vegetables	.40	.56
Vehicles	.83	.83
Weapons	.82	.86

Note—*W_s* is the estimate derived from the data generated by the feature selection task. *W_I* is the estimate derived from the informative feature task. See text for more details.

systematic variations in exemplar production and disagreement over membership and gradients of membership). In Experiments 1-3, we derived estimates of the extent to which 13 natural categories are extrinsically represented. These estimates form the basis of our predictions in the following reanalyses and studies. In particular, we predicted that the extent to which a category's representation is extrinsic would determine: (1) the relative number of exemplars generated across subjects, (2) the extent to which subjects agreed over membership decisions, (3) the degree to which membership judgments are graded, and (4) the degree to which typicality distributions differ from true membership distributions.

Reanalyses of Earlier Published Data

Membership in an extrinsically represented category is decided by the relations that putative members of the category hold with other objects. On the other hand, membership in an intrinsically represented category is decided by the properties of the putative member considered by itself. When subjects are asked to generate examples of extrinsically represented categories, they must consider the relation or relations that are the criteria for membership in the category. Because these relations are between potential category members and *some other entity*, then the subjects must also think of some class of entities with which the category members interact. For example, if subjects are asked to generate examples of Vehicles and they consider the criterion for membership in the category to be "used for transportation," then they must consider a class of entities that are transported. Different subjects are likely to think of different entities when performing this task. Some subjects may restrict the class of entities to "members of Western societies in the late 20th century" and thus miss such exemplars as "horse" and "rickshaw." Others may restrict the class to "all people" and thus miss such potential examples as "furniture dolly," "mail carrier," and "freight train." Because different individuals consider different entities that are transported, then (1) the total number of exemplars generated across subjects will be large, and (2) subjects will show relatively poor agreement with each other in the examples that they individually generate.

On the other hand, when subjects are asked to generate examples of an intrinsically represented category, they need consider no other class of entities than the putative members themselves. Therefore, (1) the total number of exemplars generated across subjects will be smaller for intrinsically represented categories than for extrinsically represented categories, and (2) subjects will show more agreement in the exemplars of intrinsically represented categories that they generate than in the exemplars of extrinsically represented categories that they generate.

Number of Exemplars Generated Across Subjects

Battig and Montague (1969) asked subjects to generate examples of a large number of naturally occurring cate-

gories. The 11 categories for which they collected data and for which we have predictions are: Birds, Clothing, Flowers, Fruit, Furniture, Sports, Toys, Trees, Vegetables, Vehicles, and Weapons. We tested the prediction that the more extrinsically represented a category is, the greater would be the number of exemplars generated. Obviously, our predictions were derived from data generated by Midwestern college students in the mid-1980s. The Battig and Montague norms were collected from Midwestern and East Coast students of the mid-1960s. Therefore, one would expect lower correlations between our predictors and their data than between our predictors and our data (see Experiments 4 and 5). Nevertheless, the obtained correlations were still significant ($r = .522$, $p < .05$ for W_s ; $r = .577$, $p < .05$ for W_t).

Intersubject Agreement

We next tested the hypothesis that the degree to which a category's representation relies on extrinsic features is related to intersubject agreement on category membership. As stated earlier, different people will consider different classes of entity when generating category exemplars of extrinsically represented categories. Therefore, the number of exemplars generated by a substantial proportion of subjects should be smaller for extrinsically represented than for intrinsically represented categories. In the following analyses of Battig and Montague's (1969) data, we tested this prediction by counting the number of category exemplars generated by 10 or more subjects. Clearly, this is not as fair a reflection of intersubject agreement as would be obtained from individual subject data, but it was not possible to extract individual subject data from the Battig and Montague norms. The obtained correlation was in the predicted direction for both predictors. The correlation between the number of exemplars generated and our measures of "extrinsicity" was significant for the W_t predictor ($r = -.527$, $p < .05$), but failed to reach significance for the W_s predictor ($r = -.420$, $p > .05$).

McCloskey and Glucksberg (1978) also reported data using a measure of intersubject agreement. They asked subjects to judge whether items (preselected to be typical members, borderline members, and nonmembers) were category members. They expressed intersubject agreement by calculating the proportion of nonmodal responses for each category member. Thus scores close to .50 indicate little agreement among subjects, since about 50% of the subjects identified the item as a member and about 50% did not so identify it. Scores close to 0, conversely, indicate high intersubject agreement. Accordingly, we predicted that the more the representation of a category is extrinsic, the higher the proportion of nonmodal responses would be. Categories for which we had obtained predictors and for which data were reported in the McCloskey and Glucksberg study were Birds, Clothing, Fruit, Furniture, Sports, Vegetables, and Vehicles. The obtained correlations for both the feature selection predictors (W_s) and the informative feature predictors (W_t) were both positive and significant, as predicted ($r = .824$, $p < .05$ for W_s ; and $r = .737$, $p < .05$ for W_t).

Measure of "Goodness-of-Example"

In the above reanalyses, we used intersubject agreement as our major measure of category fuzziness. However, this is not the measure most commonly associated with the term. The data that formed the empirical base for probabilistic and prototype theories came from studies of goodness-of-example, or typicality, ratings. In these studies, individual subjects rated some category exemplars as "better" category members than others.

Based on our model, we predict that gradients of membership become more apparent as category representations become more extrinsic. Extrinsically represented categories will have many members that are of moderate and low typicality. Borderline members will be those items that sometimes hold the appropriate relationship with another entity, and sometimes do not. In particular, the degree of membership in the category will be determined by how often an individual thinks that the potential category member exists in the appropriate relation with some other entity (i.e., how often it possesses the critical extrinsic feature). If the object occurs frequently in the appropriate relationship, it will be considered a clear, or typical, category member. If it rarely exists in that relationship with some other entity, then it will be a borderline, or atypical, member. For example, if an individual were rating the degree to which "shoe" is an exemplar of the category Tool, he/she may be able to think of an occasion on which a shoe has been used to fix things (e.g., by pounding in a nail). However, shoes occur rarely in this context, and would therefore be considered borderline exemplars at best. On the other hand, hammers occur frequently in this context; they would be considered typical tools.

Intrinsically represented categories, on the other hand, should demonstrate less graded membership. Intrinsic features are true of entities in isolation. Therefore, the contexts in which the entities occur are not relevant to whether or not they possess the feature. Accordingly, items either possess the features or do not possess them. Therefore, items should either be category members or not. For example, suppose someone were considering whether a dog is a member of the category Vertebrate. It is difficult to imagine an instance in which a dog does not possess a backbone. Similarly, it is difficult to imagine a situation in which a worm might have a backbone. Therefore, items are either vertebrates or not.

By reanalyzing data from Rosch (1975) and from Hampton and Gardiner (1983), we tested the prediction that the degree to which a category's representation relies on extrinsic features should correlate with the degree to which its membership is graded. In both studies, subjects were asked to rate items' goodness-of-example for a particular category. In Rosch's (1975) scale, a 1 indicated that an item was a very good member, and a 7 indicated that an item was either a very poor member or not a member at all. In Hampton and Gardiner's (1983) scale, a 1 also indicated a very good, or typical, member of the category. However, their lowest value, 6, indicated an

item that was not a member of the category at all. Our model specifically predicts that extrinsic features lead to *graded* membership decisions and that intrinsic features lead subjects either to mark items as very clear members or to mark items as clearly not members. A 7 on Rosch's scale is ambiguous (it may mean either that the item is a very poor member or that the item is not a member of the category at all). Consequently, we eliminated from our reanalysis of Rosch's data all items whose mean typicalities were 6 or greater because a large proportion of the individual responses to these items must have been 7s. Because Hampton and Gardiner did distinguish between very poor members and nonmembers, we were able to use all of their data in our reanalysis.

For both of our reanalyses, we used the mean rated typicality across items as the measure of the degree to which a category's membership is graded. Categories that include many atypical items should have higher means than categories whose membership is less graded. We therefore expected that the higher the mean rated typicality was, the more extrinsically represented the category would be. These means across items were provided in Hampton and Gardiner's (1983) Table 1, and were easily calculated from Rosch's (1975) data.

Our prediction that mean typicality would be positively correlated with W_s and W_t was confirmed. In the reanalysis of Rosch's data, we used only those categories for which we had predictors: Birds, Clothing, Fruit, Furniture, Sports, Toys, Vegetables, Vehicles, and Weapons. For both the predictor based on the feature selection task (W_s) and the predictor based on the informative feature task (W_t), the correlations were positive and significant ($r = .770, p < .01$ for W_s ; $r = .737, p < .025$ for W_t). In the reanalysis of Hampton and Gardiner's data, the categories for which we had predictors were Birds, Clothing, Flowers, Fruit, Furniture, Sports, Vegetables, Vehicles, and Weapons. Again, the correlations were positive and significant for both predictors ($r = .682, p < .025$ for W_s ; $r = .614, p < .05$ for W_t). Thus, our prediction that extrinsic features lead to more graded membership was confirmed by these reanalyses.

In sum, the reanalyses described above provide support for the model outlined in this paper. The more extrinsic a category's representation, the more likely the category is to show a larger (across-subject) set size, intersubject disagreement, and gradients of goodness-of-example.

Experiment 4

The results of studies such as those of Rosch (1975) and of Hampton and Gardiner (1983) have been interpreted as support for the hypothesis that membership in natural categories is graded. However, this interpretation of goodness-of-example ratings has recently been questioned. Typicality ratings do not necessarily reflect true gradients of membership. Armstrong et al. (1983) found typicality gradients in categories whose exemplars were all apparently fully members of the category. For exam-

ple, they found that the numbers 1 and 3 are more typical odd numbers than such numbers as 467 and 681. Yet all these numbers are equal members of the odd number category. Similarly, Barsalou (1983) pointed out that typicality gradients can be derived from categories that are clearly defined. Also, Barr and Caplan (1985) demonstrated that typicality has a stronger effect on category membership judgments for some categories (Furniture and Clothing) than for others (Fruit and Birds). In summary, it is apparent that typicality effects are, at best, only indirectly connected to true membership gradients.

In addition to these difficulties, published data on typicality effects are ambiguous in another sense. Rey (1983) distinguished between fuzziness in the sense of partial membership (e.g., an individual may be partly bald) versus fuzziness in the sense of uncertainty (e.g., the probability that a given individual is bald). Fuzziness in the first sense seems to be what researchers mean when they refer to "gradients of membership." Yet as Rey (1983) pointed out, the measures used do not distinguish between the partial membership sense and the uncertainty meaning. For example, "duck" received a mean rating of 3.24 in the category Bird in Rosch's (1975) ratings. This score might have been obtained by about 40% of the subjects assigning a rating of 7 and about 60% assigning a rating of 1. In other words, the obtained mean rating might reflect disagreement in the population about the membership status of ducks (i.e., the uncertainty sense of fuzziness). Alternatively, 75% of the subjects might have assigned the rating 3, and 25% of the subjects might have assigned the rating 4. This would indicate that all subjects agreed that ducks are not very typical birds (i.e., the partial membership sense of fuzziness). Unfortunately, when data are reported as means, it is impossible to tell which kind of fuzziness, or what mix of the two kinds, is being measured.

In summary, there are two major problems with interpreting published findings on typicality. First, it is by no means clear that typicality effects do reflect true membership gradients, and, second, the published data do not distinguish between the two senses of fuzziness discussed by Rey (1983). In the present study, therefore, we dealt with the first issue by designing questions for subjects that probed *not* category goodness-of-example, but gradients of membership. We asked subjects to rate potential category items on a 7-point scale, on which the end points were labeled *clearly not a member* and *clearly a member*. Therefore, subjects were forced to make judgments about the degree to which an item was a member of a particular category, rather than judgments about its typicality. To tackle the second problem, we used, as our major measure of gradients of membership, the extent to which subjects used ratings 2 through 6 on the 7-point scale. A 1 on this scale indicated that the subject believed that the item was clearly not a member of a given category, and a 7 indicated that the subject believed that the item was clearly a member of the category. A rating between these two numbers, therefore, reflects a subject's belief that the

item is neither clearly a member nor clearly not a member. In other words, such ratings indicate partial membership.

Method

Subjects. Subjects were 88 undergraduates, most of whom received course credit for their participation, depending on the class in which they were enrolled.

Stimuli. Stimulus booklets consisted of (1) a cover sheet, (2) an instruction page, and (3) lists of 50 possible exemplars, each accompanied by a 7-point rating scale, from each of the 13 categories used in previous experiments (Metals was again excluded). The lists of potential exemplars were derived by first asking a different group of 13 college students to think of and to write down as many instances of each category as they could. The generated exemplars were then divided into quartiles, based on the frequency with which they had been generated. Ten items were randomly selected from each quartile. Of the remaining 10 items presented for each category, 5 were chosen to be related nonmembers and 5 to be unrelated nonmembers.

Subjects were instructed to consider how clearly each item listed was or was not a member of the category designated at the top of the page. The 7-point rating scale, which was presented with each item, designated 1 as *clearly not a member* and 7 as *clearly a member*.

Procedure. Subjects were tested in small groups of 2-10. The experimenter read the instructions aloud as subjects followed the instructions in their booklets. Subjects then proceeded through the task at their own paces. All subjects received all 13 categories. Each subject received a different randomized order of the categories.

Results and Discussion

The 50 exemplars from each of the 13 categories are listed in the Appendix. In addition to each exemplar, its mean rating (higher ratings refer to higher degrees of membership) and the percentage of subjects who gave it a rating from 2 through 6 are listed.

As discussed in the introduction to this experiment, there are two possible interpretations of mean ratings: the means may be in the fuzzy range either because many subjects agreed that the item is a partial member of the category or because they disagreed over whether it is a category member. For example, in our data, in the Mammals category, "bluejay" had a mean rating of 3.05; however, only 25% of the subjects assigned it a rating from 2 through 6. In the Sports category, "tap dancing" received a mean rating of 3.15; however, 62% of the subjects had given it a rating from 2 through 6. In short, the mean ratings alone are ambiguous. "Bluejay" is apparently fuzzy in the sense of disagreement: different subjects gave it markedly different ratings. "Tap dancing," on the other hand, appears to be a partial member of the category Sports. Therefore, our decision to use the proportion of subjects who gave ratings between 2 and 6 as our major measure of the extent to which a category's membership is graded seemed justified.

For our first analysis, we calculated for each category, across exemplars, the mean proportion of subjects who gave partial membership ratings, that is, ratings 2-6 (see Table 3). Because our model predicts that extrinsic features should yield partial members, we expected that cate-

Table 3
Mean Proportion of Subjects Who Assigned Partial Membership Ratings* to Each of the 13 Categories Examined in Experiment 4

Category	Mean Proportion of Subjects
Birds	.194
Clothing	.223
Flowers	.243
Fruit	.198
Furniture	.331
Mammals	.184
Sports	.261
Tools	.287
Toys	.368
Trees	.205
Vegetables	.238
Vehicles	.307
Weapons	.341

*Partial membership ratings range from 2 to 6 on a 7-point scale.

gories whose representations were more extrinsic should have higher proportions of subjects giving partial membership judgments.

The estimates of the degree to which a category's representation is extrinsic, which were presented in Table 2, actually combine two different kinds of information. On the one hand, they reflect in part the extent to which the features used to describe the category are extrinsic (see Formula 2). On the other hand, they also reflect the proportion of subjects who found the feature useful or important in describing the category (see Formula 1). It is possible that either of these two factors may, on their own, explain intercategory differences in the distribution of partial membership judgments. Therefore, in addition to the two predictors shown in Table 2 (W_s and W_t), we derived (1) a measure of extrinsicity that was not weighted by importance (this measure was the mean extrinsicity of the features for each category, as shown in Table 1), (2) two measures that reflected the mean importance of the features used to describe a category (these measures were the means, for each category, of the proportions of subjects who found the features to be defining or who identified the category on that feature; see Column 1 and Column 2 of Table 1).

Accordingly, in order to determine the best predictors of partial membership judgments, we conducted a stepwise multiple regression analysis in which the criterion variable was mean proportion of subjects who assigned ratings 2-6 for a category. The predictor variables were: (1) W_s ; (2) W_t ; (3) unweighted extrinsicity (mean, across subjects, of the proportion of subjects who judged the features to be extrinsic); (4) mean, across features, of the proportion of subjects who judged the features to be defining of the category in Experiment 1; and (5) mean, across features, of the proportion of subjects who identified a category on each feature in Experiment 2.

The results of that analysis indicated that the higher the value of W_t , the greater the proportion of subjects who responded with numbers 2-6 [$R = .893$, $R^2 = .797$,

$F(1,11) = 43.08$, $p < .001$]. Once the effects of W_t were removed, no other significant predictors were obtained.

The correlation matrix for the predictors and criterion variable is presented in Table 4. Inspection of that matrix reveals that much of the predictive power of W_t lies in the unweighted extrinsicity values. (The zero-order correlation between the criterion variable and unweighted extrinsicity is high, $r = .826$, $p < .001$.) Neither of the two measures derived from importance ratings alone predicted the criterion variable well. The importance measure based on the results of Experiment 1 (S in Table 4) is correlated reliably with proportion of partial membership judgments ($r = .648$, $p < .02$); however, the correlation is the reverse of what would be predicted by most probabilistic theories of categorization. In fact, the more subjects believed the features to be defining, the more likely they were to give partial membership ratings. The importance measure based on the results of Experiment 2 (I in Table 4) is not significantly correlated with the proportion of partial membership judgments ($r = -.154$, $p > .10$).

Although the mean proportion of partial membership judgments is one measure of the extent to which a category's membership is graded, it ignores the *distribution* of such judgments across items, which provides a more complete measure of *gradients* of membership. Any category that reflects true gradients of membership should have some clear members, some not-so-clear members, some borderline members, and so on. If all or most members of the category are clear members, then membership is not very graded. Similarly, if all or most members are borderline members, then membership is not very graded. (In this latter case all, or most, members of the

Table 4
Correlations in Experiments 4 and 5

	Unweighted Extrinsicity	S	I	W_s	W_t
Part A: Correlations Among Predictor Variables					
S	.657				
I	.094	-.053			
W_s	.998	.635	.090		
W_t	.930	.718	.035	.927	
Part B: Correlations Among Predictor and Criterion Variables					
P (Expt. 4)	.826	.648	-.154	.826	.893
D (Expt. 4)	-.832	-.455	.079	-.835	-.852
C-T (Expts. 4 & 5)	-.741	-.249	.228	-.745	-.629

Note—The table shows correlations among predictor variables (Part A), and between predictor variables and (1) the mean proportion of subjects who assigned partial membership ratings (P) in Experiment 4, (2) chi square reflecting the degree of divergence (D) from an ideally graded distribution from Experiment 4, and (3) chi square reflecting the difference (C-T) between category clearness distributions from Experiment 4 and typicality distributions from Experiment 5 (Part B). S is the mean proportion of subjects who selected the features of a category to be defining, and I is the mean proportion of subjects who identified the category given each of the features of the category (see Table 1).

category would be alike in being only partial members.) Thus, maximal gradients of membership are obtained when category members are distributed evenly across all grades of membership. Any departure from this function reflects some bunching of members around some particular degree of membership and, therefore, reflects a less than ideally graded category.

Therefore, we obtained a distribution for each category in which the x -axis represented the proportion of subjects who assigned fuzzy ratings (i.e., ratings 2-6) and the y -axis represented the proportion of exemplars so marked by that proportion of subjects. An ideally graded category should show a rectilinear distribution (i.e., there should be equal numbers of items receiving ratings 2-6, across all proportions of subjects). For example, in an ideally graded category, 10 items would be marked as 2-6 by fewer than 20% of the subjects, another 10 items would be so marked by 20%-40% of the subjects, another 10 would be so marked by 40%-60%, and so forth. As discussed above, any clustering of exemplars within any of these ranges would indicate a less than ideally graded category. Therefore, comparisons of the obtained distributions with this rectilinear distribution were made by calculating chi square. The expected values of chi square were calculated by assuming an even distribution of exemplars across the x -axis divided into quintiles (0-.199, .200-.399, .400-.599, .600-.799, .800-1.00), where the values represent the proportion of subjects who marked a given exemplar as fuzzy. In calculating these chi squares, we eliminated the .800-1.00 quintile, since there were no items that were marked as fuzzy by 80% or more of the subjects. Our dependent measure, then, was the chi square calculated by comparing the obtained distribution with a hypothetical fuzzy distribution over the first four quintiles (i.e., a distribution in which all exemplars judged to be fuzzy are distributed equally across the first four quintiles). The higher the score, the less graded the membership.

The resulting chi squares are shown in Table 5. Next to each chi square is the obtained distribution of proportions across the first four quintiles. For example, for the category Mammals, the number .62 for the first quintile can be interpreted to mean that 62% of the Mammal exemplars were assigned ratings 2-6 by fewer than 20% of the subjects. Similarly, for the category Weapons, the number .18 for the fourth quintile can be interpreted to mean that 18% of the exemplars in that category were judged to be fuzzy by between 60% and 79.9% of the subjects.

A stepwise multiple regression analysis was performed, with chi square as the criterion variable, using the same predictor variables that were used in the analysis of proportion of partial membership judgments described above. The strongest predictor of chi square was again W_i [$R = -.852$, $R^2 = .725$, $F(1,11) = 29.00$, $p < .001$]. Once the effects of W_i were removed, none of the remaining variables were significant predictors. In addition, the zero-order correlation between unweighted ex-

Table 5
Obtained Distributions of Proportions of Exemplars Given Ratings 2-6 in the Membership Judgment Task, and Chi Square Reflecting the Degree of Difference Between these Distributions and an Ideally Graded Distribution in Which All Cell Frequencies Are Equal (Experiment 4)

Category	χ^2	First Four Quintiles			
		0-.199	.200-.399	.400-.599	.600-.799
Birds	42.64	.50	.46	.04	0.00
Flowers	21.20	.44	.32	.22	.02
Fruit	37.20	.56	.34	.08	.02
Mammals	55.76	.62	.38	0.00	0.00
Trees	38.00	.60	.20	.20	0.00
Vegetables	24.23	.44	.38	.18	0.00
Clothing	27.28	.54	.26	.16	.04
Furniture	12.88	.30	.28	.38	.04
Sports	26.32	.54	.22	.14	.10
Tools	12.88	.38	.30	.28	.04
Toys	5.20	.22	.36	.28	.14
Vehicles	14.96	.38	.18	.38	.06
Weapons	8.08	.40	.14	.28	.18

Note—See text for more details.

trinsicity and chi square was high ($r = -.832$, $p < .001$), although the correlations between chi square and the importance weights derived from Experiments 1 and 2 were not reliable. Clearly, most of the predictive power of W_i lies in its extrinsic component, not in the importance weightings from which it is derived. The correlations among the predictors are shown in Table 4.

Therefore, the predictions from our model that the degree to which a category's representation is extrinsic will be correlated with the degree to which its membership is graded were supported, using two different dependent measures. These findings lend further support to the results of the reanalyses reported earlier.

Experiment 5

In the introduction to Experiment 4, we argued that typicality effects are, at best, only indirect measures of gradients of category membership. Indeed, two different interpretations of such goodness-of-example ratings have been proposed. The first interpretation, offered originally by Rosch (1975), is that such ratings reveal evidence of gradients of membership in categories. The second interpretation, offered most forcibly by Armstrong et al. (1983), is that typicality effects are just that: they reveal that some members of a category are more typical than others. However, they do not indicate that some exemplars of a category are only members to a degree.

Do typicality distributions ever reflect gradients of membership? Our model implies that the answer to this question will depend on the extent to which a category's representation relies on extrinsic or on intrinsic features. As argued earlier, categories whose representations consist primarily of intrinsic features will tend to demonstrate relatively clear-cut membership gradients, that is, members will tend to be judged as either clear members or clear nonmembers. However, as Armstrong et al. (1983) have suggested, there may still be differences in typical-

ity among items that are clear category members (e.g., a "penguin" is truly a bird, but an atypical one). Accordingly, for intrinsically represented categories, distributions of typicality ratings will be more graded than will distributions of membership ratings.

Categories whose representations are primarily extrinsic, however, will show a different relationship between typicality and membership judgments. In these categories, membership itself is graded. For example, a Midwesterner will consider a "sled" to be only a partial member of the category Vehicle, because he/she rarely encounters sleds being used for transportation. Unlike intrinsically represented categories, then, extrinsically represented categories have true partial members. Accordingly, exemplars such as "sleds" are considered atypical when rated for goodness-of-example, and are considered partial members when rated for degree of membership. Thus, for extrinsically represented categories, typicality and membership gradients will be similar.

Therefore, in Experiment 5 we investigated the hypothesis that the degree to which a category is represented by extrinsic features would predict the extent to which typicality and membership distributions are similar. In particular, the more intrinsic the representation of a category, the greater the difference between the two distributions should be. Conversely, the more extrinsic the representation of a category, the more similar the two distributions should be.

One obvious problem in comparing typicality and clearness scales is that the typicality scales normally reported (e.g., Barsalou, 1983, 1985; Hampton & Gardiner, 1983; Rosch, 1973) are unipolar, ranging from *very good* to *very poor example*. However, the clearness scale may be thought of as bipolar, ranging from *clearly a member* to *clearly not a member*. One possible solution is to try to make the typicality scale bipolar; however, there are two problems with this approach. First, the data would be impossible to compare with the other unipolar typicality scales that have been reported. Second, the labeling of the endpoint opposite to *very good* is problematic. Possible labels, analogous to the label used in Experiment 4, are either *very good example of the class of things that are not members of the category* or *very good nonmember of the category*. These labels are more complex conceptually than the *very good member* endpoint, and could easily confuse subjects' understanding of the task. For these reasons, we adopted the same typicality scale that was used by Rosch (1975). The data collected, therefore, are comparable to Rosch's data.

Our decision to compare the unipolar typicality ratings with the bipolar category membership ratings was also influenced by the fact that any difficulties that might arise from comparisons of the two types of scale should be constant across categories. However, our model's predictions involve differences among categories in the degree to which typicality and membership judgment distributions resemble each other. Such intercategory differences could

not be explained by differences between the scales used for typicality and for membership judgments.

Method

Subjects. Subjects were 41 undergraduates. Depending on the course in which they were enrolled, some students received extra course credit for their participation, and others participated to fulfill a course requirement.

Stimuli. Stimulus booklets were similar to those used in Experiment 4. Each booklet consisted of a cover sheet, followed by lists of 50 possible exemplars, each of which was accompanied by a 7-point rating scale. The scales' endpoints were labeled in the same way that Rosch's (1975) scales were labeled, except that the endpoints were reversed (i.e., lower numbers indicated lower typicality). Therefore, a 1 on this scale was labeled *very poor example (or not a category member)*, and a 7 was labeled *very good example*. The 13 categories used were the same as those used in previous experiments (i.e., all of the categories used in Experiment 1 with the exception of Metals). The exemplars listed were the same exemplars listed in the stimulus booklets used in the category membership judgment task in Experiment 4.

Procedure. Subjects were tested in small groups of 2-10. After the experimenter explained the instructions, subjects proceeded through the task at their own paces. All subjects received all 13 categories. Each subject received a different randomized order of the categories.

Results and Discussion

As might be expected, subjects' ratings of typicality from this experiment correlated highly with ratings of membership given by subjects in Experiment 4. Separate correlations were computed for each of the 13 categories; all were greater than .95.

Our original hypothesis, however, was that clearness and typicality ratings would reveal different distributions. In particular, we expected that when subjects made ratings for intrinsically represented categories, they would be more likely to mark numbers 2-6 when making typicality judgments than when making membership judgments, since atypical members are, nevertheless, category members. On the other hand, we expected that when subjects rated potential exemplars of extrinsically represented categories, they would be just as likely to use ratings 2-6 for membership judgments as for typicality, since these categories do, unlike intrinsically represented categories, show true gradients of membership. Correlations, like those reported above, do not reflect such characteristics of distributions. Therefore, we investigated whether the differences in distributions between typicality and membership judgments were correlated with the degree to which a category's representation relies on extrinsic features.

We first calculated, for each exemplar, the proportion of subjects who assigned ratings 2-6 (see also Results, Experiment 4). We then divided these 50 values into four intervals: 0-.199, .200-.399, .400-.599, .600-.799. Each interval represents the proportion of subjects who assigned a rating from 2 through 6 to that item. As was the case for membership judgments in Experiment 4, there was no exemplar which received ratings 2-6 by 80% or more

of the subjects. These distributions for typicality ratings are presented in Table 6.

Next, we calculated chi squares comparing the typicality distributions shown in Table 6 with the distributions for membership judgments shown in Table 5, for each category. The chi squares shown in Table 7, then, reflect the degree to which the two types of distributions differ.

A stepwise multiple regression analysis was performed on these chi squares, using the same predictor variables that were used in the analyses of Experiment 4. The correlation matrix resulting from this analysis is shown in Table 4. The strongest predictor was W_s [$R = -.745$, $R^2 = .555$, $F(1,11) = 13.74$, $p < .005$]. Once the effects of W_s were removed, none of the remaining variables were significant predictors. Most of W_s 's predictive power lies in its unweighted extrinsicity component. The zero-order correlation between chi square and the unweighted extrinsicity measure was again high ($r = -.741$). The correlations between chi square and the two importance weightings derived from Experiments 1 and 2 were not significant.

GENERAL DISCUSSION

In the five experiments presented here, as well as in a number of reanalyses of data reported by other investigators (Battig & Montague, 1969; Hampton & Gardiner, 1983; McCloskey & Glucksberg, 1978; Rosch, 1975), we have demonstrated that the estimated degree to which a category is represented by extrinsic features predicts the relative extent to which a category's membership is graded. The predictions hold across a number of different tasks and subject populations. In particular, estimated degree of extrinsicity predicts: (1) variations in total number of category exemplars generated across subjects (reanalysis of Battig & Montague, 1969); (2) the extent to which subjects agree when generating category members (reanalysis of Battig & Montague, 1969); (3) the proportion of nonmodal responses when deciding category membership (reanalysis of McCloskey & Glucksberg,

Table 6
Obtained Distributions of Proportions of Exemplars Given Typicality Ratings 2-6 by the Proportion of Subjects Indicated (Experiment 5)

Category	First Four Quintiles			
	0-.199	.200-.399	.400-.599	.600-.799
Birds	.20	.36	.44	0
Clothing	.22	.38	.36	.04
Flowers	.10	.58	.30	.02
Fruit	.38	.32	.30	0
Furniture	.18	.22	.48	.12
Mammals	.10	.56	.34	0
Sports	.10	.46	.30	.14
Tools	.16	.52	.30	.02
Toys	.14	.40	.38	.08
Trees	.16	.56	.24	.04
Vegetables	.10	.46	.36	.08
Vehicles	.10	.26	.44	.20
Weapons	.22	.28	.28	.22

Table 7
Obtained Chi Squares Comparing Distributions of Ratings 2-6 Given for the Membership Rating Task from Experiment 4 and the Typicality Rating Task from Experiment 5

Category	Chi Square
Birds	23.7
Clothing	11.7
Flowers	15.1
Fruit	9.1
Furniture	4.4
Mammals	37.5
Sports	22.6
Tools	7.8
Toys	2.6
Trees	23.4
Vegetables	18.1
Vehicles	12.9
Weapons	5.1

1978); (4) mean typicality ratings (reanalyses of Hampton & Gardiner, 1983; Rosch, 1975); (5) distributions of membership judgment ratings (Experiment 4); and (6) the degree to which category membership judgment distributions differ from typicality judgment distributions (Experiment 5). Our results also clarify the interpretation of gradients of membership in natural language categories. One of the dependent variables we used to assess category fuzziness was the proportion of subjects who gave a "partial member" response (i.e., any number between 2 and 6). This measure assesses fuzziness in the sense of partial membership, without confounding it with population disagreement.

One major implication of these results is that graded category structures are indeed consistent with clear category representations. For example, our subjects gave us a clear definition of Weapons: they are unsafe in the wrong hands, used in crime, and used for protection. On the other hand, the extension of Weapons is fuzzy; over 50% of our subjects agreed that "rocks," "ropes," and "rubber bands" are neither clearly weapons nor clearly non-weapons. Traditionally, category fuzziness has been viewed as being inconsistent with clearly represented categories (i.e., categories represented by sets of necessary and sufficient features). However, when at least some of these features are extrinsic, then even a clearly represented category can be characterized by a fuzzy extension. As mentioned in the paper's introduction, whether an object is judged to be a member of an extrinsically represented category will depend not on the characteristics of the object itself, but instead on the relationship(s) it holds with other things. If it holds the appropriate relationship frequently, it will be considered a clear member; if it holds it infrequently, it will be a fuzzy or borderline member.

Our original goal was to offer a more parsimonious explanation of category fuzziness than that offered by current mixed models of category representation (e.g., Armstrong et al., 1983; Landau, 1982; Medin & Smith, 1984). Mixed models could, indeed, account for the fact that membership and typicality distributions differ. They could, for example, propose that identification functions

are used in typicality judgments, whereas core criteria are used in membership judgments. To the extent that the two sets of criteria differ, so there will be differences in the two kinds of distribution. However, only by adding distinctions, such as the one proposed here, to a model that is already acknowledged to be cumbersome (Armstrong et al., 1983) could one predict which categories have similar identification and core criteria and which categories have diverging sets of criteria. In addition, mixed models, by themselves, cannot account for the high correlations we repeatedly obtained between measures of the degree to which a category's representation is extrinsic and many aspects of category extensions.

Clearly, these results also have implications for the interpretation of goodness-of-example ratings. Correlations between goodness-of-example ratings (Experiment 5) and category membership ratings (Experiment 4) were uniformly high. However, for a number of the categories, there were differences in the distribution of responses. Such differences between distributions are relevant to recent controversy over the meaning of typicality effects. Two major interpretations of goodness-of-example ratings have been proposed. The first interpretation, offered by Rosch (1975), is that such ratings reveal evidence of gradients of membership in categories. The second interpretation, offered by Armstrong et al. (1983), is that typicality judgments reflect identification functions, not category membership judgments. For a number of the categories that we investigated, there was little difference between the distributions of typicality and membership judgments (Experiments 4 and 5). To the extent that these distributions are similar, then typicality effects do reflect degrees of membership and there do appear to be true partial members, as Rosch (1975) claimed. On the other hand, for those categories for which the two types of distribution differ (e.g., Mammals, Birds, Trees), then the best that can be said of typicality ratings is that they reflect differences in typicality, but not differences in category membership.

Our model can also account for how context changes the membership status of items in a category, without positing changes in the criteria for membership in the category. In the paper's introduction, we discussed recent findings (e.g., Roth & Shoben, 1983) that demonstrate that variation in context influences category structure. These findings appear to contradict the usual assumption that a major function of categories is to provide stability across contexts (Rey, 1983). As a consequence, the results appear to contradict the widely held assumption that categories have an invariant internal structure. On the other hand, we have distinguished between the criteria for membership in a category, which are relatively stable across contexts, and the set of items that are members of this category. This latter set may be variable when the membership criteria for the category are extrinsic. For example, stable criteria for the category Tools might be "used to fix things" and "used to work with." In most everyday contexts, the extension of this category

would not include such items as sticks and rocks. However, when camping in the woods, such items may indeed become tools, because, in this context, they are used to fix things and to work with. Thus, within our model, it is possible for category membership to be dynamic and changing, even though the conditions for membership in the category remain stable. Obviously, it follows from this argument that an item's rated typicality in a category may be sensitive to changes in the contexts considered. Yet, at the same time, the subject's membership criteria for the category itself can remain stable across the same changes in context.

We should note that we have repeatedly used the term "representation" to refer to an individual's understanding of the meaning of a category term. We do not, however, mean to imply that true, classical definitions of such terms exist. An individual may use a set of necessary and sufficient features in making category membership judgments and in his/her own internal category representation. However, these features may be neither the societal definition of the term nor a scientifically acceptable definition of the term. In addition, unlike a definition in the classical sense, an individual may change his/her criteria for category membership. Thus, for example, as a novice becomes an expert, so his/her representation of categories in the subject area may change (Murphy & Wright, 1984); the criteria need not remain fixed.

It is evident that the explanation we are proposing for why extrinsically represented categories reveal gradients of membership is related to Barsalou's (1985) suggestion that "frequency of instantiation" (people's subjective estimates of how often an item is experienced as a category member) accounts, in part, for graded category structure. Our account differs from Barsalou's in two ways. First, "frequency of instantiation" is not synonymous with our concept of the proportion of times an extrinsic feature is true of an item. For example, for most Americans, chicken hawks are not very often encountered as birds; in fact, they are rarely encountered at all. Therefore, according to Barsalou, chicken hawks rarely instantiate birds, and would consequently be considered atypical. However, nearly every time a chicken hawk is encountered, it possesses the features that are criterial for the category Bird. Thus, according to our model, chicken hawks are clear members of the Bird category. Second, Barsalou uses frequency of instantiation to predict category typicality structure. We, on the other hand, are interested in predictions of *category membership*. It is possible, therefore, that frequency of instantiation may predict typicality, but not actual membership. On the other hand, the proportion of times an item possesses the critical feature(s) predicts category membership, not typicality.

The data reported in this paper are correlational. Therefore, the differences we observed among categories could be due to the effects of other, unknown, variables. Indeed, a number of distinctions made by other investigators are similar to the intrinsic/extrinsic distinction. For example, Tversky and Hemenway (1984) distinguished

between parts and nonparts, Miller and Johnson-Laird (1976) distinguished between perceptual and functional features, and Paivio (1971) discussed concreteness. Intrinsic features are more likely to be perceptual, concrete parts than are extrinsic features, and some might argue that such a confound is responsible for the results presented here. We argue, however, that the predictive power of such distinctions comes from the fact that functional-abstract-nonpart features are often relational (i.e., extrinsic). We have shown earlier in the paper the mechanism by which relational features can produce graded membership in a category: namely, where the criteria for membership are relational, an object can slip in and out of membership without itself changing. The results of the present studies show that extrinsicity is very highly correlated with a number of measures of category structure; if some other variable is responsible for these results, it must in turn be very highly correlated with extrinsicity. However, our case would obviously be strengthened by studies using artificial categories, where the different types of features could, in theory, be controlled more rigorously. Our theory would also be strengthened by deriving new predictions for other categorization phenomena, and by confirming those predictions. To this end, we have demonstrated that intrinsically represented categories exhibit "stronger" class inclusion properties than do extrinsically represented categories (Barr & Caplan, 1986). Obviously, further use of both of these research strategies would strengthen the evidence for our model.

Finally, we are arguing that the concept of extrinsic features is an elegant and parsimonious way to account for much of the data on gradients of category membership and context effects. However, we make no claim that this particular approach will explain all kinds of fuzziness. It is patently the case that there are some terms that are fuzzy for other reasons. "Euglena" are "fuzzy" plants/animals studied both by botanists and by zoologists. They are fuzzy because they possess some, but not all, of the features of plants (e.g., they have chlorophyll) and some, but not all, of the features of animals (e.g., they move and take food like animals). Other terms will be fuzzy because the category is represented by a continuous scale. If a mountain is a hill over 1,000 ft, then borderline mountains are those hills that are close to 1,000 ft. Another source of fuzziness arises from lexical ambiguity. For example, the term "ash" may refer to the tree, or to the residue of a fire. The term itself is fuzzy because it is sometimes used to refer to one and sometimes used to refer to the other; however, the referents, in either case, are not themselves fuzzy members. Still other terms may be fuzzy because people do not fully understand their meanings. For example, someone may think that a frog is a "fuzzy" mammal because he/she is unsure of what a mammal is. Any of these types of fuzziness may lead to intersubject disagreement or partial membership judgments.

In sum, it is apparent that our model does offer a parsimonious account of why there are gradients of membership in a number of common taxonomic categories.

At the same time, the predictions of the model account for a large proportion of the between-category variance among the categories that we and other investigators have studied. The model accounts for these results without distinguishing between "core" and "identification function" components of a category's representation. As such, it provides a simpler account of gradients of membership and related results than do many other current models of categorization.

REFERENCES

- ANDERSON, R. C., & ORTONY, A. (1975). On putting apples into bottles—A problem of polysemy. *Cognitive Psychology*, *7*, 167-180.
- ANDERSON, R. C., PICHERT, J. W., GOETZ, E. T., SCHALLERT, D. L., STEVENS, K. V., & TROLLIP, S. R. (1976). Instantiation of general terms. *Journal of Verbal Learning & Verbal Behavior*, *15*, 667-679.
- ARMSTRONG, J. L., GLEITMAN, L. R., & GLEITMAN, H. (1983). What some concepts might not be. *Cognition*, *13*, 263-308.
- BARR, R. A., & CAPLAN, L. J. (1985). Two kinds of feature? A test of two theories of typicality effects in natural language categories. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society* (pp. 181-189). Hillsdale, NJ: Erlbaum.
- BARR, R. A., & CAPLAN, L. J. (1986, May). *Class inclusion dos and don'ts: "Some" implications for "all" hierarchies*. Presented at the 16th Annual Symposium of the Jean Piaget Society, Philadelphia.
- BARSALOU, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211-227.
- BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 629-654.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80*(3, Part 2).
- GLASS, A. L., HOLYOAK, K. J., & O'DELL, C. (1974). Production frequency and the verification of quantified statements. *Journal of Verbal Learning & Verbal Behavior*, *13*, 237-254.
- HAMPTON, J. A., & GARDINER, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, *74*, 491-516.
- LANDAU, B. (1982). Will the real grandmother please stand up? *Journal of Psycholinguistic Research*, *11*, 47-62.
- MCCLOSKEY, M. E., & GLUCKSBERG, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition*, *6*, 462-472.
- MCCLOSKEY, M. E., & GLUCKSBERG, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, *11*, 1-37.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.
- MEDIN, D. L., & SMITH, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*, 113-138.
- MILLER, G. A., & JOHNSON-LAIRD, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- MURPHY, G. L., & WRIGHT, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 144-155.
- PAIVIO, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- REY, G. (1983). Concepts and stereotypes. *Cognition*, *15*, 237-262.
- RIPS, L. J., SHOBE, E. J., & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, *12*, 1-20.
- ROSCH, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.

ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 533-605.

ROTH, E. M., & SHOEN, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, **15**, 346-378.

SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

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

WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, **12**, 129-140.

Webster's New World Dictionary of the American Language. (Second College Edition) (1980). New York: Simon & Schuster.

WILKINS, A. T. (1971). Conjoint frequency, category size, and categorization time. *Journal of Verbal Learning & Verbal Behavior*, **10**, 382-385.

APPENDIX

Exemplars Used in Experiment 4, the Membership Rating Study, and in Experiment 5, The Goodness-of-Example Rating Study

A 7 indicated a *clear or typical member*, and a 1 indicated the item was *clearly not a member* (Experiment 4) or a *very poor example or not a category member* (Experiment 5). Each exemplar is accompanied by the mean rating in Experiment 4 (Membership), the mean rating in Experiment 5 (Typicality), and the percentage of subjects who gave each exemplar a rating of 2-6 in Experiment 4.

Exemplar	Mean Rating		Percentage of Subjects Giving Fuzzy Ratings
	Membership	Typicality	
Birds			
Bluebird	7.00	6.68	0.00
Blackbird	7.00	6.49	0.00
Canary	7.00	6.80	0.00
Cardinal	7.00	6.73	0.00
Bluejay	6.99	6.68	1.14
Redheaded woodpecker	6.97	6.29	2.27
Eagle	6.94	6.80	4.55
Mockingbird	6.92	6.65	4.55
Hummingbird	6.92	6.56	4.55
Golden eagle	6.88	6.63	5.68
Wild canary	6.85	6.24	4.55
Seagull	6.85	6.27	11.36
Pigeon	6.85	6.51	9.09
Quail	6.84	6.22	7.95
Barn swallow	6.82	5.63	7.95
Vulture	6.82	6.00	11.36
Mynah bird	6.72	5.90	14.77
Turtledove	6.68	6.12	17.05
Snowbird	6.66	6.12	13.63
Peacock	6.64	6.22	23.86
Flamingo	6.58	5.83	21.59
Wren	6.51	5.63	17.05
Turkey	6.44	5.78	28.41
Pelican	6.44	5.37	20.45
Bobwhite	6.39	5.61	13.64
House wren	6.34	5.71	20.45
Goose	6.34	5.76	34.09

Chicken hawk	6.32	5.05	21.59
Duck	6.28	5.71	30.68
Mallard duck	6.23	5.54	32.95
Hen	6.22	5.29	35.23
Rooster	6.18	5.39	36.36
Chicken	6.16	5.54	34.09
Blue crane	6.06	5.49	28.41
Condor	5.89	5.29	33.33
Chickadee	5.72	5.68	26.14
Phoenix	5.52	5.00	35.63
Cowbird	5.50	5.05	32.95
Martin	5.42	4.83	19.32
Penguin	5.40	5.20	39.77
Sandcrane	4.97	4.71	30.68
Duck-billed platypus	4.39	4.37	45.45
Big Bird	4.33	4.32	36.36
Thunderbird	3.72	4.34	21.59
Bat	3.26	2.88	45.98
Flying squirrel	1.85	2.15	31.82
Butterfly	1.83	1.71	29.55
Housefly	1.47	1.46	11.36
Helicopter	1.23	1.29	4.55
Airplane	1.20	1.44	5.68

Flowers

Rose	7.00	6.59	0.00
Tulip	6.98	6.68	1.14
Carnation	6.97	6.83	2.27
Marigold	6.94	6.54	3.41
Daisy	6.94	6.41	3.41
Violet	6.94	6.51	3.41
Chrysanthemum	6.87	6.20	6.90
Lilac	6.86	6.66	9.20
Iris	6.84	6.05	8.05
Petunia	6.83	6.51	3.41
Mum	6.82	6.15	6.90
Lily	6.80	6.66	11.36
Blue violet	6.80	6.32	9.20
Wildflower	6.73	6.34	5.75
Poinsettia	6.72	6.37	6.98
Orchid	6.72	6.29	10.23
Magnolia	6.68	6.29	10.23
Gladiola	6.63	6.00	13.64
Gardenia	6.48	6.34	21.84
Sunflower	6.48	5.73	21.84
Pansy	6.47	6.32	14.94
Cherry blossom	6.43	6.10	23.86
Meadow flower	6.38	6.00	26.14
Zinnia	6.36	5.98	24.14
Poppy	6.25	6.00	27.27
Azalea	6.22	5.05	17.24
Geranium	6.20	6.00	16.09
Forget-me-not	6.09	6.05	21.59
Snapdragon	6.02	5.80	34.88
Dried flower	5.99	4.59	24.14
Peony	5.79	6.02	31.40
Goldenrod	5.66	4.32	40.23
Baby's breath	5.58	5.34	34.09
Hibiscus	5.39	4.95	42.53
Crocus	5.33	5.22	46.59
Black-eyed Susan	5.30	5.66	40.23
Bridal wreath	5.03	5.02	54.02
Hyacinth	4.83	5.10	39.78

APPENDIX (Continued)				Parsley	1.74	2.22	32.95
Exemplar	Mean Rating		Percentage of Subjects Giving Fuzzy Ratings	Onion	1.68	1.88	22.73
	Membership	Typicality		Potato	1.49	1.32	12.50
				Spinach	1.38	1.44	10.23
				Ice cream	1.27	1.12	12.50
				Rose	1.15	1.24	3.41
				Cake	1.11	1.15	4.55
Silk flower	4.83	5.00	48.28	Mammals			
Philodendron	4.81	4.49	44.32	Ape	6.77	6.24	9.09
Heather	4.64	4.49	52.87	Polar bear	6.69	6.10	11.36
Schefflera	4.44	4.24	63.53	Grizzly bear	6.67	5.90	9.09
Bachelor button	4.44	4.12	45.35	Baboon	6.64	6.17	15.91
Boston fern	3.69	3.61	51.14	Fox	6.61	6.02	13.64
Ground ivy	3.41	3.80	56.32	Chimpanzee	6.59	6.17	13.64
Leaf	1.89	2.00	39.08	Panther	6.57	6.12	14.77
Seaweed	1.79	1.61	31.03	Tiger	6.55	6.05	11.36
Pine cone	1.72	1.98	30.68	Horse	6.55	6.20	10.23
Root	1.41	2.10	22.73	Goat	6.55	6.15	13.64
Nut	1.30	1.78	13.79	Bear	6.55	6.37	6.82
Fruit				Lion	6.52	6.22	12.50
Pear	7.00	6.80	0.00	Cheetah	6.51	5.98	13.64
Pineapple	6.99	6.70	1.14	Donkey	6.50	6.00	13.64
Apple	6.99	6.85	1.14	Deer	6.49	6.17	9.09
Plum	6.98	6.61	2.27	Koala bear	6.47	6.22	12.50
Banana	6.95	6.80	4.55	Wolf	6.46	6.02	11.49
Yellow apple	6.94	6.66	2.27	Water buffalo	6.45	5.54	15.91
Peach	6.93	6.76	3.41	Cow	6.43	5.88	13.64
Apricot	6.90	6.37	6.82	Cat	6.43	6.22	10.23
Grapefruit	6.90	6.68	3.41	Leopard	6.40	6.05	12.50
Grape	6.88	6.78	5.68	Kangaroo	6.39	5.51	18.39
Tangerine	6.88	6.68	3.41	Rabbit	6.36	5.85	13.64
White grape	6.88	6.39	6.82	Giraffe	6.32	5.85	18.18
Navel orange	6.88	6.61	3.41	Antelope	6.27	5.90	21.59
Mandarin orange	6.81	6.20	11.36	Mouse	6.27	5.37	20.45
Watermelon	6.78	6.61	9.20	Groundhog	6.20	5.44	18.18
Cantaloupe	6.76	6.61	4.55	Chipmunk	6.15	5.20	19.32
Blueberry	6.72	6.63	12.50	Boar	6.10	5.34	22.99
Red cherry	6.72	6.56	9.09	Orangutan	6.05	5.44	21.84
Blackberry	6.65	6.34	12.50	Squirrel	6.02	5.22	21.59
Raspberry	6.60	6.29	15.91	Hippopotamus	5.90	5.60	28.74
Black cherry	6.48	6.37	15.91	Guinea pig	5.89	5.00	26.14
Prune	6.30	6.02	27.27	Gazelle	5.86	5.44	30.68
Mushmelon	6.24	5.78	18.39	Gerbil	5.83	5.12	25.00
Boysenberry	6.07	5.51	36.78	Mongoose	5.65	4.59	31.82
King David apple	6.06	5.83	21.59	Whale	5.63	5.05	20.45
Crabapple	5.99	5.17	20.45	Rat	5.58	4.63	22.73
Date	5.86	4.95	32.95	Muskrat	5.56	5.10	30.68
Papaya	5.82	5.37	29.55	Sea otter	5.11	4.46	32.95
Mulberry	5.74	5.54	34.09	Goose	3.82	3.02	37.50
Mango*	5.47	4.93	32.95	Bluejay	3.05	2.88	25.00
Santa Claus melon	5.44	5.02	35.23	Frog	2.23	2.46	29.55
Pumpkin	5.01	4.66	35.23	Lizard	2.09	2.83	18.18
Pomegranate	4.93	4.51	44.32	Worm	2.00	1.80	25.00
Guava	4.85	4.34	60.47	Snake	1.86	2.41	18.18
Kumquat	4.84	4.61	50.00	Crab	1.76	2.22	29.55
Coconut	4.70	4.83	38.64	Spider	1.70	1.93	18.18
Tomato	4.52	4.41	34.09	Germ	1.34	1.22	7.95
Avocado	4.44	3.98	37.50	Venus flytrap	1.33	1.20	10.23
Rhubarb	4.07	3.61	45.98	Trees			
Cucumber	3.16	3.07	23.86	Apple tree	7.00	6.54	0.00
Currant	3.03	2.63	39.78	Oak	7.00	6.71	0.00
Jelly	2.50	2.24	42.05	Walnut tree	6.99	6.68	1.14
Carrot	1.76	1.90	12.50				

APPENDIX (Continued)							
Exemplar	Mean Rating		Percentage of Subjects Giving Fuzzy Ratings	Bus driver	Carton	Newspaper	
	Membership	Typicality					
Door knob	1.35	1.55	13.64	1.38	1.88	4.55	
Rainbow	1.31	1.39	14.77	1.25	1.49	12.50	
Rose	1.31	1.00	12.50	1.20	1.12	4.60	
Photocopying machine	1.24	1.68	12.50	Weapons			
Vehicles				Neutron bomb	6.98	6.88	1.14
Truck	6.97	6.68	2.27	Hydrogen bomb	6.97	6.83	2.27
Bus	6.97	6.54	3.41	Rifle	6.94	6.83	5.68
Family car	6.93	6.85	0.00	Torpedo	6.92	6.83	5.68
Car	6.93	6.83	0.00	Fire bomb	6.88	6.68	7.95
Van	6.92	6.73	1.14	Hand gun	6.86	6.80	10.23
Pickup truck	6.92	6.56	1.14	410 shotgun	6.85	6.46	9.09
Sports car	6.90	6.39	6.82	Missile	6.78	6.56	7.95
Taxi	6.85	6.59	5.68	Anti-aircraft gun	6.77	6.34	9.09
"Semi" truck	6.84	6.51	10.23	Dynamite	6.76	6.85	12.50
Limousine	6.82	6.66	5.68	Sawed-off shotgun	6.75	6.78	9.09
Refrigerated truck	6.73	5.98	6.82	Dagger	6.72	6.44	19.32
Garbage truck	6.72	5.78	15.91	M-16	6.70	6.68	4.55
Recreational vehicle	6.65	6.07	17.05	Billy club	6.69	6.17	13.64
Dune buggy	6.62	5.71	21.84	Knife	6.68	6.88	18.18
Sprint car	6.43	5.93	20.45	Bow & arrow	6.55	6.22	25.00
Stock car	6.39	6.02	20.45	Fire thrower	6.35	6.00	25.00
Train	6.26	6.07	22.73	Destroyer	6.30	5.71	15.91
Motor home	6.26	5.78	17.05	Fist	6.08	5.95	44.32
Semi-tractor	6.11	5.32	30.68	Ice pick	5.88	6.05	56.82
Jet plane	6.10	5.90	28.41	Machete	5.86	5.56	29.55
Camper	5.92	5.56	31.03	Chain	5.68	5.61	55.68
Space shuttle	5.63	5.15	45.45	Dart gun	5.61	5.51	44.32
Bulldozer	5.58	4.61	47.73	Mace	5.57	5.20	50.00
Dragster	5.57	4.90	39.77	Scissors	5.57	5.78	57.95
Three-wheeler	5.56	5.05	48.86	Mustard gas	5.48	4.93	46.59
Spaceship	5.55	5.32	42.05	Fire	5.48	5.88	53.41
Submarine	5.42	4.78	45.45	AK 47	5.31	5.41	51.19
Street cleaner	5.38	3.83	43.18	Rocks	5.25	5.29	62.50
Motorboat	5.36	4.85	45.45	Derringer	5.13	4.73	32.18
Scooter	5.25	4.51	57.95	Gas	4.90	4.85	64.77
Triplane	5.11	4.53	56.82	Rocket launcher	4.59	3.88	34.09
Bicycle	5.08	4.83	56.82	Fork	4.43	4.83	68.18
Rocket	5.06	4.80	45.45	Rope	4.40	4.51	73.86
Tricycle	4.92	4.61	64.77	Fingernails	4.23	4.27	76.14
Sailboat	4.65	4.24	57.95	Satellite	3.76	3.59	62.50
Wheelchair	4.24	3.76	70.45	Drugs	3.44	3.66	50.00
Rickshaw	3.74	2.63	59.77	Car	3.41	3.90	60.23
Gondola	3.58	3.44	50.00	Chair	3.27	3.49	62.50
Roller skates	3.15	3.29	65.91	Rubber band	2.84	2.88	70.45
Horse	3.01	2.88	54.55	Computer	2.73	2.80	40.91
Roller coaster	2.98	2.78	46.59	Book	2.44	2.68	54.55
Skis	2.73	3.34	53.41	Hairspray	2.42	2.73	56.82
Escalator	2.59	2.49	51.14	Pillow	2.28	2.61	48.86
Lawnmower	2.09	2.66	43.18	Paper bag	1.72	1.76	26.14
Husky (dog)	1.77	1.93	27.27	Tree	1.48	1.76	23.86
Indianapolis 500	1.70	2.34	14.77	Puppy	1.47	1.66	17.05
Interstate highway	1.42	1.61	10.23	Rain	1.32	1.90	15.91
				Marshmallow	1.20	1.22	5.68
				Bread	1.06	1.24	4.55

*In local dialect, "mango" refers to bell peppers as well as to the tropical fruit.