Essentialism and graded membership in animal and artifact categories

CHARLES W. KALISH

University of Wisconsin, Madison, Wisconsin

A number of studies have argued that people view membership in animal and artifact categories as a matter of degree. These studies have generally failed to distinguish between the issues of typicality and category membership. Thus, data which have been taken to demonstrate that membership is a matter of degree may only demonstrate that typicality is graded. Partly on the basis of these findings, it has been argued that some categories are organized around an underlying essence. The essence determines membership absolutely. The present paper reports a series of studies that reexamine the question of graded membership. In the first study, subjects were asked to rate both typicality and category membership for the same stimuli as a way of distinguishing the two questions. A second method relied on the intuition that disagreements about membership in all-or-none and graded categories may have different qualities. Results from both studies suggest some support for claims that membership in animal and artifact categories is a matter of degree. A third study explored the possibility that graded responses were due to conflicting, or ambiguous, sets of criteria. A task focusing on biological features did not lead to more absolute categorization. These results contradict essentialist predictions.

The question of whether membership in categories is a matter of degree has a long history within cognitive psychology (see Lakoff, 1987, and Smith & Medin, 1981, for reviews). Interest in the question began with challenges to the view that concepts were composed of necessary and sufficient features. The rise of "probabilistic" views of conceptual structure implied that membership in categories was graded—a matter of degree rather than absolute. However, a series of papers, most notably those of Armstrong, Gleitman, and Gleitman (1983), Osherson and Smith (1981), and Rey (1983), raised conceptual and methodological problems with regard to existing arguments for graded category structure. These critiques challenged conclusions about the gradedness of category members but did not make positive claims. Existing data are suspect, but subsequent studies have not addressed the shortcomings of earlier work.

What is at issue in the question of whether membership in categories is all-or-none or a matter of degree? The concern is whether each element of a category is just as much a member as any other. The alternative is that there are gradations; some things are members to a lesser degree than others. Some categories clearly have all-or-none membership (e.g., *even number*; all even numbers are equally even). Others obviously admit degrees (e.g., *red*; things can be more or less red; Kay & McDaniel,

1978). For most concepts, though, intuitions are not so clear. Plant, animal, and substance categories (often called "natural kinds") and human artifacts (e.g., vehicles, furniture) have been of particular interest to researchers on concepts (Barr & Caplan, 1987; Keil, 1989; Rosch & Mervis, 1975). Because of their traditional importance, the nature of membership in animal and artifact categories is the focus of this paper. Recent essentialist accounts of concepts (Atran, 1987; Gelman & Wellman, 1991; Keil, 1987; Medin & Ortony, 1989) suggest that animal and artifact categories might be quite different.

At one time, there was wide consensus that membership in animal and artifact categories was a matter of degree. Data from a number of studies showed that some instances of a category were better members than others (e.g., Barr & Caplan, 1987; McCloskey & Glucksberg, 1978; Oden, 1977; Rips, Shoben, & Smith, 1973; Rosch & Mervis, 1975; Tversky & Gati, 1978). The classic paradigm (Rosch, 1973) was to have subjects judge how "good an example," typical, or representative an item was of a category. In these tasks subjects would rate some instances as better than others; a robin is a more typical, better, bird than a penguin is. A variety of wordings were used to pose questions (e.g., "how clear a member," Barr & Caplan, 1987) with similar effects. Other methods also yielded the same results; for example, subjects were more consistent in categorizing some items (e.g., robins) than others (e.g., penguins) across trials (McCloskey & Glucksberg, 1978). These typicality effects led to "similarity-based" views of categorization; membership was understood as a judgment of similarity between the potential instance and the cate-

Preparation of this paper was supported in part by a grant from the Graduate School of the University of Wisconsin. I would like to thank Susan Gelman for her helpful comments on this work. Requests for reprints should be sent to C. W. Kalish, Department of Educational Psychology, University of Wisconsin, 1025 W. Johnson St., Madison, WI 53706 (e-mail: ckalish@macc.wisc.edu).

gory. However, several questions were raised about the validity of typicality as a measure of category membership.

Past studies had operationalized category membership in terms of typicality or similarity. Rey (1983) has argued that this identification confuses two aspects of categories. On the one hand, there are epistemological issues concerned with how we tell that something is a member of a category. This would involve features used in identifying category members (Miller & Johnson-Laird, 1976; Smith & Medin, 1981). On the other hand are metaphysical (Rey, 1983) or core (Miller & Johnson-Laird, 1976) components of categories. This involves what it means to be a member of the category; what category membership consists of. These two elements may be quite distinct. For example, people may typically discover they are sick (belong in the *sick* category) by being told so by a doctor. But this is just a way of finding out, not a core part of what it means to be sick (being called sick by a doctor doesn't make one sick). The two aspects of categories contribute to fuzzy membership judgments in different ways. We may be unsure whether an item belongs in a category because it is difficult to observe the criterial features (e.g., it is hard to tell that a penguin has the requisite features of birds). Uncertainty might also arise because we don't know which features are criterial (e.g., whether the features penguins share with other birds are the important ones). The problem, argues Rey, is that studies of graded membership may have been measuring the first kind of fuzziness; how easy or hard it is to tell. The second—whether there are clear criteria or not—seems to be the central issue for category membership.

A significant body of empirical evidence suggests that typicality ratings do not measure subjects' intuitions about the meaning, or core, of category membership. That is, typicality and category membership are different things. Armstrong et al. (1983) gathered typicality ratings for examples of defined (all-or-none) categories such as "even number." Subjects gave graded typicality judgments; 8 was rated a better even number than was 18. However, subjects also indicated that they believed membership in the defined categories was all-or-none (8 is not more of an even number than is 18; but see note 1). Armstrong et al. argued that typicality ratings could not be measuring category membership. Later work by Rips (1989) and Keil (1989) demonstrated that changing some of the features of an animal (e.g., a bird exposed to chemical wastes coming to appear like an insect) may change its perceived typicality but not its category membership. Thus, as an index of category membership, typicality seemed to be misleading.

Traditional studies need to be reexamined in light of work suggesting that typicality and category judgments are distinct. Studies operationalizing membership in terms of typicality seem to have been measuring the wrong construct. Studies using alternative methods (e.g., McCloskey & Glucksberg, 1978; Oden, 1977) or wordings (Barr & Caplan, 1987) are at least suspect. Were these methods measuring typicality or category membership?

Graded responses on these tasks are ambiguous; do they mean that category membership is a matter of degree or that the tasks measure typicality? The goal of the present studies was to resolve this ambiguity.

The fact that gradedness of typicality does not imply gradedness of categorization is not, in itself, positive evidence about the nature of category membership. The studies discussed above merely argue that the two types of ratings are different. Although typicality clearly admits of degree, it is not clear whether category membership does. Models of graded membership in animal and artifact categories and models of absolute membership are equally consistent with existing data. As Armstrong et al. concluded in 1983, "we are back at square one in discovering the structure of everyday categories *experimentally*" (p. 291, emphasis in original).

One reaction to the distinction between typicality and category membership has been the development of essentialist conceptions of category structure (Gelman & Wellman, 1991; Keil, 1987; Medin, 1989; Medin & Ortony, 1989; Walker, 1992). According to this view, most of the features associated with a category are manifestations of some deeper, defining, set of essential properties. The essence is accorded this importance because it is thought to be the cause of all the other features of an individual. The features that cause or produce other features are candidates for the essence. This position differs from classical accounts of concepts in that these underlying properties may be unknown. What is important is the belief that there do exist some fundamental defining features—what Locke (1707/1961) called "real essences." Since these features may be unknown and are typically thought to be inaccessible, people must use some other means to identify members. This was the role Locke assigned to the "nominal essence" (or "prototype"; Putnam, 1975): a set of observable characteristics that enable one to pick out category members. According to this view, there are two kinds of features—features of the essence and features of the prototype. Features of the prototype are effects of the underlying essential features. Typicality effects are due to the distribution of prototype features, but category membership derives from the essential features that are present in every instance of the category.¹ Evidence for the essentialist view comes from data showing that inferences about underlying features, rather than appearances (features from the prototype) are the basis for category membership decisions (Keil, 1989; Gelman & Markman, 1986; Rips, 1989). Essentialist accounts may hold that not all categories have essences. Categories of plants and animals may have essentialist structure while artifacts do not (Atran, 1987). For example, although function seems central to our conceptions of artifacts (Keil, 1989), it is not always criterial (Malt & Johnson, 1989). This may be because function is a relational property, whereas the core features of animals are intrinsic (Keil, 1989).

From the description above, it is possible to derive two claims regarding essentialist categories. In the first place, membership should be all-or-none. One of the

motivations for essentialist models was to account for intuitions that low-typicality instances are nonetheless full members of a category. For example, an ostrich is an atypical bird but it is not a partial bird. It has a "bird essence" equally as much as a robin does. Related to this is a distinction between essences and accidental features. Accidental features are the observable indicators of an essence; the stripes of a tiger, the long neck of a giraffe. While pragmatically important, their role is evidentiary rather than constitutive. Stripes provide evidence that the animal has a tiger essence but having stripes is not a criteria for being a tiger (Atran, 1987). If the presence of the accidental features fails to inform one about the underlying essence (e.g., because the feature was not actually caused by the essence), it should be irrelevant to judging category membership. For example, painted stripes on a lion are not evidence of a "tiger" essence (Keil, 1989). A major claim of an essentialist model is that some features are only indirectly relevant to category membership. Although accidental features may be important in computing similarity or typicality (Rips, 1989), they only figure in judgments of category membership by providing evidence for an underlying essence. Thus, judgments about instances with atypical features (ostriches, painted lions) are crucial evidence for essentialist views.

Although essentialist models may be tenable, they have not received direct support. The data supporting the essentialist view only demonstrate that apparent, surface, features are less important for category membership than are hidden or internal features, not that appearances are irrelevant (inessential). Category membership is usually assessed as a dichotomy (Gelman & Markman, 1986; Keil, 1989) which might mask a small, but still significant, role for "nonessential" features. Rips's (1989) study of animal transformations did use a scale to measure category membership. In this study, category membership ratings for "posttransformation" items (e.g., bird that changed to look like an insect) appear to have been significantly graded; they are intermediate between the endpoints of the scale. However, it is not possible to determine from the report whether this gradedness may have been due to conflicting absolute responses (e.g., some subjects' answering definitely "bird," others definitely "insect"). Furthermore, the provision of a scale may have biased subjects toward intermediate responses. Although typicality and category membership seem to be different judgments, we do not have clear evidence that they are qualitatively different. It is still possible that both judgments are matters of degree. However, improvements in method would seem to hold some promise for resolving the questions of graded membership.

Essentialist models are not the only responses to problems regarding typicality. It is possible to respond to the critiques without the need to posit two types of features. Perhaps typicality and category membership are different kinds of computations over the same set of features. This could be termed the "two-weight" model; there are not two kinds of features but rather two different con-

texts in which features are evaluated. It has been well demonstrated that judgments of graded structure vary by context (Barsalou, 1987). Different situations highlight different features of instances and categories. Membership judgments may be made in the same way as typicality judgments; they are just another context. For example, typicality judgments tend to weight perceptual features more heavily, whereas categorization judgments may consider structural or functional features as more important.² A related proposal is that there are two ways of categorizing, a "rule-based" and a "similarity-based" method (Smith & Sloman, 1994). In special cases, one process might operate to the exclusion of the other, but in most situations a categorization decision would involve input from both methods.

Several studies provide evidence for the "dual-weight" view. One source comprises findings of typicality effects in tasks that do not seem to be based on an identification procedure. For example, statements ascribing atypical members to a category ("a penguin is a bird") are rated as less true than statements about typical members ("a robin is a bird"; Oden, 1977). Smith and Sloman (1994) have found that characteristic (nonessential) features do seem to be considered along with essential features in categorization decisions. Furthermore, Rips's (1989) work may show that typicality, categorization, and similarity judgments are all different. The variety of different ratings observed suggests that context may be affecting salience of features, rather than that there are whole different sets of features. Thus, one response to recent developments is to accept that typicality and membership are different, but to deny that either is all-or-none.

There are two responses to critiques of typicality studies. Essentialist models posit a qualitative difference between typicality and category membership: the former may be a matter of degree; the latter is absolute. Furthermore, features that determine typicality may be only indirectly related to category membership. An alternative is to see the two judgments as results of the same process but with different parameters. Essentialist models could be seen as a special case of this "two-weight" view; they posit that the weights of a certain class of features are zero for categorization judgments. Thus the null hypothesis for these studies is that those features which lead to low typicality ratings will have nonzero weights and so lead to lowered category membership ratings.

One important source of evidence for essentialist structure consists of atypical instances. The theory predicts that instances with low typicality should be seen as members or not absolutely (e.g., penguins are low-typicality but absolutely birds). To the extent that this is not true (membership for these items is graded), the categories do not have essentialist structure. While graded membership is inconsistent with essentialist models, absolute membership is only negative evidence against the "two-weight" models. Two-weight models merely predict that it is possible to get graded responses—for example, if the measure is sensitive and the item is atypical. Thus the

focus of the present studies is to test the essentialist claim that membership will not be graded. The finding that an individual can be more or less of a member of a category argues that the category does not have essentialist structure.

The remainder of this paper describes three methods of measuring category membership. In each case, the goal was to separate judgments of category membership from typicality judgments with which they may often be confounded. Thus the first concern was whether the method would distinguish between typicality and membership judgments. To this end, revised methods and comparisons with items known to have graded or absolute category membership were used. After the validity of the measures has been assessed, the focus will be on which categories exhibit all-or-none or graded membership.

STUDY 1

The purpose of this study was to collect data that reflected judgments of category membership as distinct from typicality (Rey, 1983; Rips, 1989). In Study 1, subjects were asked to make both typicality and category membership judgments for the same stimuli. It was reasoned that this procedure would tend to distinguish the ratings of categorization from ratings of typicality. For this study, a modified scale was used which allowed both graded and absolute responses. In previous studies, judgments had been measured either with a multivalued (e.g., Likert) scale or with a dichotomous scale. Use of these scales might have forced subjects to give graded or absolute responses, respectively. The modified scale presents two absolute responses in addition to, but separate from, a multivalued scale. To check the effects of these manipulations, control categories were included. Items were picked which were thought to have an essentialist structure (e.g., even number, Armstrong et al., 1983). These items should show graded typicality and absolute category membership. It was predicted that another set of items would not show differences between typicality and categorization (e.g., red). Measures of categorization were considered valid to the extent that responses conformed with these predictions.

The stimuli were a set of low-typicality instances of categories (e.g., a penguin as an instance of *bird*; a piano as *furniture*). Low-typicality items were chosen because these represent the best chance for observing graded categorization responses. Absolute categorization responses for highly typical instances are not informative; membership may be graded but the instrument may not be sensitive enough to record this. However, it might also be the case that no familiar instances would yield graded responses. Subjects might have enough experience of examples such as ostriches and penguins to have adjusted the feature weights so that these items can be treated as full members (albeit with low typicality). Thus a set of unfamiliar atypical examples would also be useful.

A second set of stimuli depicted animals undergoing more or less radical alterations. These examples repre-

sented a different kind of low-typicality item. Some items depicted relatively minor changes—for example, a lion painted to appear like a tiger (see Keil, 1989). According to an essentialist view, changes to an animal's appearance or behavior are changes in surface features; they are relevant to typicality but not to category membership. Other items depicted more radical changes—for example, a bird exposed to radiation and grown insect-like (Rips, 1989). These items may also be seen as involving accidental features and leaving the essence of the animal unchanged. Alternatively, depending on one's conception of essence or changes involved in the transformation, they may be seen as changing the animal's identity; perhaps the bird really has become an insect. However, under either interpretation the question of interest was whether the change (or lack of change) would be absolute or a matter of degree. If the transformation has altered the animal's essence, it should be no longer a member of the original category; it is either a bird or not. Although previous work with these kinds of items has shown large typicality effects, it has not been possible to judge the effects of the transformations on category membership (Keil, 1989; Rips, 1989). The methodological improvements in Study 1 should allow interpretation of these data.

In Study 1, subjects were presented with a wide variety of atypical instances of categories. The question of interest was which (if any) atypical items would be rated as having all-or-none membership in the category. Which way the items were categorized (Is a penguin rated as a bird or a nonbird?) was not the focus of this study. As McCloskey and Glucksberg (1978) have demonstrated, there may be lack of consensus regarding how to categorize atypical items. The most important dependent measure used in the present study ignored this inter-subject variability and measured the extremity of responses. Regardless of which way an item is categorized, is it definitely (or absolutely) assigned to that category? The evidence which tells against essentialist models is intermediate categorization responses; the direction of the rating is not relevant. Clearly the direction is relevant for testing any particular claim regarding the nature of the essence of a category. Take, for example, the case of rating *penguin* as a *bird*. If one's hypothesis is that flight is the essence of the category *bird* (as opposed to, say, egg-laying), then whether or not a penguin is considered to be a bird is crucial. However, if the question is merely whether there is some essence or other to the category *bird*, then both positive and negative judgments about penguin could be consistent. What is crucial is whether there are only these two possibilities; either *penguin* is thought to have the right essence—in which case it is completely and absolutely a bird, or *penguin* has the wrong essence—in which case it is completely and absolutely not a bird.

Method

Subjects. The subjects were 31 undergraduates at a large mid-western university. Participation in the experiments was part of the requirements of an introductory psychology course.

Table 1
Categories of Items Used in Study 1

Category Type	Component Categories	Example
defined	unit of measure, U.S. currency	"a \$2 bill" as "U.S. currency"
graded	good grade	"a B+" as "a good grade"
animals	horse, dog, insect	"a donkey" as "a horse"
artifacts	furniture, weapon	"a piano" as "furniture"
control transformation	label changes	A seal is put in a cage labeled "dolphin."
surface transformation	appearance changes (from Keil, 1989)	Doctors make a tiger look like a lion.
deep transformation	physical changes (from Rips, 1989)	Chemicals makes a bird grow insect-like.

Stimuli. Seven types of items were included in the study. Each type of item was represented by several instances of one or more categories. Table 1 lists the categories used for each item type. A complete list of instances (judged for typicality and membership in the categories) is presented in Appendix B. Three item types (defined, graded, and incongruously labeled) were included as controls. Defined items were instances of categories predicted to have all-or-none membership. Graded items were from a category predicted to allow degrees of membership. Conclusions regarding the gradedness of animal and artifact categories were based on comparisons with these items. For example, were judgments of animals significantly more graded than judgments of defined items? The third set of controls described animals receiving incongruous labels. It was predicted that these alterations would not effect category membership; membership for these items should be largely absolute because they have no atypical features. These items were compared with sets describing surface (after Keil, 1989) and deep transformations (modeled after Rips, 1989). With the exception of control transformations, all instances were chosen to have low typicality within their respective category. Additionally, three items involved rating invented instances (e.g., a *Grinbla* as a piece of furniture). These items were included to check that subjects would use the "don't know" response.

Procedure. The procedure involved two judgments about the relationship of an instance (e.g., a *donkey*) to a category (e.g., *horse*); a judgment of typicality and a judgment of membership. The subjects were tested individually on microcomputers. Initially the subjects were presented instructions describing typicality and categorization, as well as the differences between the two (adapted from

Rosch, 1975; see Appendix A). They were then presented with 50 computer "screens." One instance and one category were presented on each screen (see Figure 1 for an example). The first task was to rate how typical the instance was of the category. Following this, subjects were to rate the degree to which the instance was (or was not) a member of the category. Half the transformation items of each kind involved rating the animal as an instance of the category it started in (base). Half involved rating the animal as the category it ended as (final). For example, the item describing the lion made to look like a tiger was rated as either a lion (base) or a tiger (final). A modified Likert scale was used to collect ratings. The scale included two absolute end points (0 and 8) and a graded scale in the middle (ranging from 1 to 7). This was provided to minimize demand characteristics for either absolute or graded responses. The subjects were also able to indicate "don't know" to both ratings. Transformed items were always the last 12 screens presented. Items were otherwise presented in random order.

Results

Table 2 presents the mean typicality and categorization ratings for the seven types of items rated in this study. Ratings for the three kinds of transformations are broken down according to whether the pretransformation (base) or posttransformation (final) identity was rated. Subjects' mean typicality ratings for the defined, graded, animal, and artifact items were entered into a repeated measures analysis of variance (ANOVA). The ef-

Item	Question Number: 4	Category
a Donkey		a Horse
-- How Characteristic, Representative, "Good an Instance" --		
1)	<div style="display: flex; justify-content: space-between; font-size: small;"> a Donkey is in no way typical of a Horse A Little a Donkey is somewhat typical of a Horse A Lot a Donkey is absolutely typical of a Horse </div> <div style="display: flex; align-items: center; margin-top: 5px;"> <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/> — <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>1 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>2 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>3 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>4 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>5 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>6 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>7 — <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/> </div> <div style="text-align: right; margin-top: 5px;">I don't know how typical <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></div>	
-- How Much of a Member --		
2)	<div style="display: flex; justify-content: space-between; font-size: small;"> a Donkey is not at all a Horse Barely a Donkey is sort of a Horse Very much a Donkey is completely a Horse </div> <div style="display: flex; align-items: center; margin-top: 5px;"> <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/> — <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>1 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>2 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>3 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>4 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>5 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>6 <input style="width: 30px; height: 20px; border: 1px solid black; text-align: center;"/>7 — <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/> </div> <div style="text-align: right; margin-top: 5px;">I don't know the degree <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></div>	

Figure 1. Example of the rating scale from Study 1.

Table 2
Mean Typicality and Categorization Ratings: Study 1

Item Type		Typicality	Categorization
Defined		3.9	6.7
Graded		5.8	6.1
Animals		3.7	3.9
Artifacts		3.0	3.9
Control transformations	base	5.4	6.8
	final	0.6	.81
Surface transformations	base	1.8	7.0
	final	5.1	1.6
Deep transformations	base	1.0	3.6
	final	5.3	4.1

fect of item type was significant [$F(3,90) = 35.2$, $MS_e = 1.3$, $p < .001$]. Pairwise comparisons (Tukey's HSD, $p < .05$) revealed that typicality ratings were significantly higher for graded items than for all other kinds of items. Ratings for defined items were higher than ratings for artifacts. The typicality of defined and animal items did not differ.

Defined and graded items served as controls to check the validity of the procedure. One prediction was that there would be a high proportion of absolute (0 or 8) category membership judgments for defined items and a low proportion for graded items. Seventy-eight percent of the responses (243) were 0 or 8 for defined items (8% were intermediate; 14% were "Don't know"). For graded items, 42% of the responses were absolute and 56% intermediate. Close to half (29 of 66) of the absolute responses came from one item (*an A* as an instance of *a Good Grade*³). A second prediction was that category membership and typicality ratings should differ for defined items but not for graded items. The mean difference between the two ratings for defined items was 2.8; for graded items, .3. The ratings differed more for defined items than for graded items [$t(30) = 7.4$, $p <$

.0001]. As a further indication that subjects were using the scales correctly, 93% of responses to invented items (e.g., *a Finaran as a Dog*) were "don't know."

Since the main question of interest is how absolute the ratings were, rather than whether they were high or low, raw scores were transformed into "extremity scores." Scores were computed by taking the absolute value of the difference between the scale midpoint and the raw score. Thus a response of 0 (*not at all*) and 8 (*completely*) both yielded an extremity score of 4. Transformed scores ranged from 0 to 4. Extremity scores were averaged for each of the types of items. These means are presented in Figure 2. Transformation items were analyzed separately from the other category and typicality judgments.

Nontransformed items. Data were analyzed in a 2 (judgment type: categorization/typicality) $\times$ 4 (item type: defined, graded, animal, artifact) repeated measures ANOVA. The two main effects and their interaction were significant [judgment type, $F(1,30) = 35.0$, $MS_e = .41$, $p < .0001$; item type, $F(3,90) = 11.9$, $MS_e = .48$, $p < .0001$; interaction, $F(3,90) = 35.6$, $MS_e = .13$, $p < .0001$]. Analysis of simple effects revealed no differences between item types for typicality judgments [$F(3,90) = 2.4$, $MS_e = 3.7$, n.s.]. Only categorization judgments differed between types of items [$F(3,90) = 39.8$, $MS_e = .24$, $p < .0001$]. Pairwise comparisons within categorization judgments showed that judgments for defined items were significantly more extreme than for other items ($p < .01$, Tukey's HSD). Graded, animals, and artifacts did not differ. This is also reflected in the percentage of absolute (0 or 8) responses; 38% for animals, 41% for artifacts, but 78% for defined items. Categorization judgments were more extreme than typicality judgments for defined and animal items [$F(1,30) = 85.8$, $MS_e = .29$, $p < .0001$, and $F(1,30) = 15.2$, $MS_e = .20$, $p < .001$, respectively] but not for graded or artifact items (both F s < 2).

Figure 2. Mean extremity scores, Study 1. Error bars represent one standard deviation (computed across subjects and all items of a type).

Transformed items. Extremity scores for the transformed items were analyzed in a 2 (judgment type: categorization/typicality) $\times$ 3 (item type: control/surface/deep) repeated measures ANOVA. Only the main effect of item type was significant [$F(2,60) = 17.1$, $MS_e = .27$, $p < .001$]. Typicality and category membership judgments for control items were significantly more extreme than for deep transformations (Tukey's HSD, $p < .01$). This pattern also held for surface transformations, however only ratings of final states showed more graded categorization than did ratings of control items ($p < .01$; see Figure 2). The two kinds of transformations did increase the absolute difference between actual categorization and typicality ratings. For control items, the mean difference between the two ratings was 1.4. For surface items, the difference was significantly greater [$M = 4.8$, $t(30) = 6.3$, $p < .001$]. This was also the case for deep items [$M = 3.3$, $t(30) = 3.8$, $p < .001$].

Discussion

The procedure used in Study 1 seemed effective for distinguishing typicality from category membership judgments. As predicted, there was a significant difference between the two kinds of judgments for defined items. This was not a general response strategy, as can be seen by the much smaller difference between the two judgments for graded items. Defined items, but not graded, were generally rated as having all-or-none category membership. Categorization ratings were not due to simple "generalizations" of differences in typicality ratings. Defined and animal items were rated as equally typical, yet categorization ratings differed significantly. Artifacts were significantly less typical than defined items (though the ratings did not differ in extremity); thus differences in membership ratings could be explained just by differential typicality. However, because subjects were making distinct categorization and typicality judgments in other cases, it is unlikely that they conflated the two judgments in the artifact cases.

The results of this study suggest that membership is seen as a matter of degree for at least some animal and artifact categories. Categorization judgments for these items were significantly more graded than were judgments for defined items. This was true even though the gradedness of typicality was about the same for the three kinds of items. Subjects made more absolute judgments about the category membership of atypical defined items (units of measure and currency) than they did about atypical animals and artifacts. Graded responses were more pronounced for artifacts than for animals in this study. However, this difference may not signify a general difference between the two kinds of categories; the effects may be due to the specific stimuli chosen (e.g., a different set of items might show the reverse effect). This same caution holds for other universal claims—for example, that all animal categories admit degree. What can be argued from the data is that some familiar, common animal and artifact categories seem to have graded membership.

Results from the transformation items suggest that minor (surface items) and more radical (deep items) changes in an animal affect its category membership as well as its typicality. In the first place, this study replicated previous findings that the effects of the transformations are more pronounced for typicality than for category membership (Keil, 1989; Rips, 1989); the transformations led to increased differences between the two types of ratings. Control transformations showed high levels of absolute categorization and typicality ratings. Presumably these items were not atypical and were comparable to rating prototypical instances (e.g., rating a robin as a bird). As expected on the basis of either model, typicality rating of surface and deep transformations were more graded than ratings for control items. However, categorization judgments were also more graded for the transformations. Contrary to an essentialist prediction, transformations could incrementally alter category membership. Changing the appearance and behavior of an animal seems to affect category membership. Painting stripes on a lion may not make it a tiger, but the action does seem to reduce the animal's lion-hood to a small degree. Although appearances may not be the only features relevant to category membership, or even very important ones, they are not irrelevant.

These data tend to support the "dual-weight" hypothesis. While the essentialist prediction tested was that items with intermediate typicality would have all-or-none membership, both types of judgments seemed to admit degrees. This held both for familiar items (animals and artifacts) and for unfamiliar (transformations) atypical items. According to essentialist claims, people may be unsure about whether or not an anomalous item belongs in a category, but they should see the choice as a definitive yes or no. For example, both Keil (1989) and Rips (1989) found that transformations such as those presented in Study 1 did not change subjects' ratings of category membership. The essentialist interpretation of these results is that the altered features were not relevant to category membership; the essence remains unchanged across the transformations. However, an alternative interpretation is that the altered features do figure in membership judgments but are not weighted as strongly as those features which remain constant. This is consistent with Smith and Sloman's (1994) suggestion that rule-based and similarity-based categorization may operate together to determine membership decisions. Data from Study 1 support claims that the same features figure in both typicality and categorization judgments. The difference in the two kinds of ratings would be attributed to the different contexts determined by typicality and categorization questions.

One limitation of the procedure in Study 1 is the way in which uncertainty might have been reflected in the ratings. An intermediate categorization rating may have meant that a subject was unsure about the category membership, not that the subject was sure that the item was an intermediate case. As Rey (1983) points out, there are two kinds of uncertainty; epistemological (*I don't know*

whether a coyote is a dog.) and metaphysical (*A coyote is kind of a dog but kind of not.*). These could also be described as an "identification" uncertainty and a "core" uncertainty. It is the latter sort of uncertainty that Study 1 was designed to measure. Subjects were able to answer "don't know" for any particular judgment and did so overwhelmingly for invented controls.⁴ Nonetheless, they may have confused the two kinds of uncertainty.

The procedure in Study 1 depended on subjects' making an explicit distinction between typicality and categorization. Different subjects may have understood this distinction differently and/or it may have struck them as unnatural and artificial. Since this procedure placed some fairly heavy demands on subjects, it seemed desirable to try to replicate the results with a different method.

STUDY 2

Claims about membership in all-or-none and graded categories are different kinds of judgments. To place an instance in an all-or-none category is to decide a matter of fact; the object either is or is not a member. Making a membership judgment for a graded category may be more like offering an opinion or preference; we judge that the object seems "close enough" to count as an instance.

One consequence of this distinction is how disagreements about category membership are treated. What happens when people disagree about whether or not an instance is a member of a category? If the category has all-or-none membership, there is a right and a wrong answer. Imagine a dispute regarding whether 1731 is a member of the category "prime number." We may not know if the number is prime, but presumably there is some fact to the matter. Having all the necessary information available would prove the issue one way or the other. Put bluntly, categorizing the object as a prime number is either correct or incorrect.

When a category admits degrees of membership, disputes may not have a clear resolution. There may be no way to demonstrate that one position is incorrect, no facts to present. Since an object may be partly a member of the category (have some degree of similarity to the category), the question is whether it is "enough" of a member to be so identified. If the boundaries are truly fuzzy, we should expect some disagreements. Consider a debate about categorizing somebody as friendly or not. We could agree on all the facts of the matter yet not agree on how friendly someone has to be to be called friendly. The boundaries of the category are fuzzy so we cannot enforce claims about membership.

Asking people to resolve disputes over category membership is a method that has been used in some previous research on categories. Malt (1991) had people judge who would be able to determine how to categorize a series of odd objects. Her interest was in people's beliefs about the completeness of their category representations: Did they rely on experts to make decisions or not? However, whether or not people rely on experts to decide category membership is a different question than whether

they think that membership is a matter of degree or not.⁵ In another part of this research, Malt's subjects judged whether an object described as "halfway between" two categories (e.g., a lemon tree and an orange tree) (1) was "probably one or the other, but I don't know which," (2) could be called either one, or (3) couldn't be either. She found that most subjects chose the first option (probably one or the other) for most natural-kind (plant and animal) items. Responses for artifacts tended to be "either one" or a mix of all responses.

Malt's (1990) work suggests that animal categories are thought to have clear and unambiguous membership, whereas artifact categories may be more graded. However, there was a significant amount of variability in responses. Malt reports the response that subjects gave to the majority of items; even then, only 26 of 40 subjects gave the "one or the other" response to most natural kinds. In addition, fewer than half the subjects (19/40) chose the "one or the other" for the majority of nominal-kind items (e.g., triangle or grandmother). As Armstrong et al. (1983) point out, membership in these categories is all-or-none. Thus there is reason to question whether the task is measuring beliefs about gradedness of membership.

The method used in Study 2 offers several advantages as a measure of category membership. Subjects were asked whether a dispute over category membership had definite resolution. Malt's (1990) subjects had to, additionally, make predictions about which way the dispute would be resolved. A simple belief in absolute membership is consistent with two of the response options offered: "must be one or the other" and "can't be either one." Subjects were not allowed to remain agnostic. Second, the wording of Malt's items explicitly instructed subjects that items were "halfway" between two categories. One explanation for the variability in responses is that subjects had different ideas about what features were involved in making the object seem halfway between two categories: Were they thinking of a relatively surface confusion (e.g., coloration) or a deeper ambiguity (e.g., genetic structure)? An alternative method would be to present items as instances of a superordinate kind and have subjects judge disputes regarding the subordinate level classification of the instance. For example, an item might present a dispute over whether a particular instance of fruit was an orange or not. This form of question would remove some extraneous demands (e.g., visualizing a fruit halfway between an orange and lemon). It also would allow one to avoid questions associated with picking contrasting categories (e.g., contrasting *orange* with *lemon*; would the answers have differed had the contrast been *orange* and *banana*?). As with Study 1, a set of control items was included to check the validity of the procedure. The superiority of the new procedure remained an empirical question. Would subjects' responses consistently distinguish between defined and graded categories? If so, responses might be taken as good evidence regarding the structure of animal and artifact items.

In the present study, subjects were asked to judge whether various disagreements had a single outcome or

John and Jane are discussing an odd looking animal.

John says the animal is an elephant
Jane says the animal is not an elephant

Are there some facts we could discover about the animal that would have to prove one of the people wrong? Or, could there be some room for disagreement about what counts as an elephant?

Question

39

One of them must be wrong	People can disagree	I don't know
There is some fact that could decide one way or the other.	Nobody can prove one of them wrong. Can argue for both sides	

Figure 3. Sample item from Study 2, asking subjects to rule whether a dispute has a definite resolution or not.

might be irresolvable. Must one of the people be wrong in an argument about category membership? This task was a way of assessing whether membership in the categories was thought to be all-or-none or a matter of degree. Disputes involving all-or-none categories should have resolution one way or the other. In contrast, when the category in question admits of degree there need not be a definite answer about membership; opinions may vary about how much membership is enough.

Method

Subjects. Subjects were 20 undergraduates from a large mid-western university. Participation in experiments was part of the requirements of an introductory psychology course. No subjects in Study 2 had participated in Study 1.

Procedure. Data were collected on microcomputers. Subjects were presented with instructions regarding graded membership and the possibilities of disagreements. After reading these instructions, they were presented with scenarios describing an encounter with an unusual object and a disagreement about how that object should be categorized (see Figure 3). The instances being argued about were described as unusual members of a superordinate class (e.g., *John and Jane have come across a strange animal*).⁶ A disagreement about categorizing the object was presented (e.g., *John says the animal is a lizard; Jane says the animal is not*

a lizard). For each scenario, subjects were asked to indicate whether one of the people must be wrong or whether it was possible for people to legitimately disagree.

Stimuli. The stimuli were 42 disagreements (see Appendix B for a complete list of the items). Six disagreements involved membership in defined categories; 6 involved graded categories. Eight disagreements revolved around animals and 8 concerned artifacts. Twelve scenarios involved a transformed animal (4 surface transformations, 4 deep transformations, and 4 controls). To give the subjects an idea of the diversity they should consider, the 12 transformation items were presented first. Order of presentation was otherwise random.

Results

Figure 4 presents the mean proportions of disagreements allowed for each kind of item. Responses for individual items are shown in Appendix C. Very few disagreements were allowed for defined items (95% confidence interval for median: 0.0–.12, Wilcoxon test). Very few of the graded items were seen as having definite resolutions (95% confidence interval for median: .90–1.0, Wilcoxon test). Animal and artifact items were intermediate between these two extremes. Only 6 subjects allowed any disagreements for defined items, 12 accepted

Proportion Of Disagreements Judged Not To Be Resolvable
Study 2

Figure 4. Results from Study 2. Mean proportion of judgments that disagreements about assigning instances may be irresolvable for a particular kind of category. Error bars represent one standard deviation (computed across subjects and all items of a type).

some disagreements for animals, and 18 did for artifacts. Subjects allowed more disagreements for animals than for defined items (Wilcoxon test, $N = 12$, $p < .01$). Fewer disagreements were accepted than for graded items ($N = 19$, $p < .001$). The same held true for artifacts; more disagreements than defined ($N = 18$, $p < .001$), but fewer than graded ($N = 18$, $p < .005$). Fewer disagreements were allowed for animals than for artifacts [$t(19) = 4.5$, $p < .001$]

Transformation items also differed in the proportion of disagreements that were allowed. Very few disagreements were allowed for the control transformations (95% confidence interval for median: 0.0–.12, Wilcoxon test). Nine subjects accepted at least one disagreement for the surface transformations. Fifteen accepted some disagreements for deep items; only 5 did for control items. Compared with control items, significantly more disputes were allowed for deep transformations (Wilcoxon test, $N = 13$, $p < .001$) and for surface transformations (Wilcoxon test, $U = 8$, $p < .05$). Deep transformations were seen as less clear-cut than surface transformations (Wilcoxon test, $U = 11$, $p < .01$).

Discussion

As with Study 1, results from Study 2 suggest that category membership in animal and artifact categories is seen as a matter of degree. Asking subjects to judge whether disagreements about categorization were (in principle) resolvable seemed to be a valid index of beliefs about the gradedness of category membership. For items expected to have “all-or-none” membership (defined items), subjects overwhelmingly thought there was a right answer to a dispute. If one person says a number is odd and another says it is not, one of them must be wrong. Debates over items with graded membership were seen more as differences of opinion, not matters of fact. It was thought that people could legitimately disagree over whether someone was friendly or not, for example. Categorizing animals and artifacts fell in between these two extremes. Most subjects thought, at least for some items, that categorizing an animal or artifact was not an absolute matter of fact; some fuzziness was allowed.

Study 2 revealed a difference between animal and artifact categories. Although disagreements were allowed for both, significantly more were allowed for artifacts. This might suggest that animal categories are “less graded” than artifact categories. Barr and Caplan (1987) make just this argument, on the basis of the finding that typicality judgments for animals tend to be more extreme than typicality judgments for artifacts. The claim that categories of animals have a “tighter” structure (probability of membership drops off faster with distance) than artifacts must be based on carefully matched samples. Claims that more artifact categories than animal categories admit of degree would, additionally, require the ability to tell when a category is absolute. However, as mentioned above, it is difficult to draw conclusions from absolute judgments. Even according to probabilistic mod-

els some items may be so close to the prototype that categorization decisions are virtually certain (e.g., there is no question that a robin is a bird, because it is so close to the prototype). The conclusion to be drawn from these studies, though, is that at least some animal (and artifact) categories allow significant degrees of graded membership.

Results from the transformation items in Study 2 also supported the results from Study 1. The identities of control transformations were thought to be resolvable one way or the other. These items were not atypical, so there should be no basis for disputes. Surface transformations introduced some discrepant features. These changes did increase the number of subjects who accepted uncertainty regarding how the animals should be categorized. Subjects allowed even more disagreements about category membership for deep transformations. The alterations in both surface and deep transformations did seem to affect the category membership of the animals involved. Transforming the properties of an animal makes its category membership less clear-cut. In contrast, from an essentialist perspective, a transformation may be either effective at changing category membership (if it alters essence) or ineffective (if it only alters accidental features). One might be unsure whether the essence was changed or not. However, a belief in essence would seem to commit one to believe that there is some fact to the matter; either the essence changed or it didn't. Some sort of test or inspection of the essence would settle the matter for certain.

Study 2 was an attempt to measure a particular kind of uncertainty or fuzziness. Subjects had to judge whether there were right and wrong answers to questions about category membership, without actually having to judge which answer was right and which was wrong. Posing this question rules out uncertainty of identification. People are not uncertain of the category membership of an object because it is hard to tell or because they don't possess certain facts. The results suggest that animal and artifact categories show a core uncertainty. People do not think that there is a clear list of criteria against which instances can be compared in order to make definite category judgments. There is room for interpretation as to how well an instance meets the criteria.

Membership in animal and artifact categories may (at least sometimes) be a matter of degree. However, there remain different ways to account for this gradedness. One possibility is that categorization is essentially a probabilistic computation. Membership of an instance is based on its degree of difference from some prototype, for example. A second possibility is that membership decisions are based on definitive criteria but that there may be several, conflicting, sets of criteria for any particular category. Lakoff (1987) has suggested that multiple criteria, or multiple models, underlie the apparent gradedness of membership in categories. For example, he argues, there are distinct sets of biological, social, and psychological criteria (which he terms “models”) for the category *mother*. Each set of criteria might be clear-cut; what it takes for someone to be biologically a mother is

not fuzzy. Any particular instance, though, might meet some criteria but not others. Someone may be a psychological mother, but not meet the biological criteria (e.g., adoptive mothers). Thus, overlapping sets of criteria can give rise to unclear cases.

The "multiple-model" interpretation of graded responses was suggested by some pilot work for Study 2. A group of subjects was asked to justify their answers regarding the categorization disputes. Several subjects indicated that they had allowed disagreements because they thought the participants might be using different standards. For example, one subject responded that one of the disputants might have been thinking about how the animal looked while the other was thinking about its biological kind. Also from Study 2, the item which seemed the least a matter of degree (fewest disagreements accepted) was *mammal*. Although this is a superordinate (and might be expected to allow the most leeway), it is also a scientific term, and so it may demand a biological interpretation. A final study was designed to test the possibility that subjects were giving graded responses because they were uncertain which criteria to use for judging membership.

STUDY 3

This study addressed the question of whether some ambiguity in the context of the categorizations judgments in Studies 1 and 2 might have been responsible for the graded responses. In Study 3 the methods from these prior studies were used, but the stimuli were limited to animal items and the subjects were instructed to base their judgments on biological criteria. Animals were chosen for this study because animal categories seem to be among the best candidates for essentialist structure (Atran, 1987; Keil, 1989). Furthermore, Malt's (1990) work suggests that people may accept the authority of scientific experts over natural-kind categories (including the kinds of animals). Biology seems to be a very important context in which to view animal categories.

Both dual-weight and essentialist models would be consistent with some reduction in graded responding when the focus is restricted to biological features. An essentialist model would predict that explicit limitation of judgments to biology should eliminate competing essences. Graded responses will disappear as subjects focus on the presence or absence of biological essence. According to a two-weight model, restriction to biological features may reduce the number of distinctive features present in atypical instances. Some of the features that make an animal an atypical instance of a category are nonbiological in nature (e.g., typical birds are associated with warm weather, whereas penguins are associated with cold). The purely biological differences between penguins and other birds might be smaller than the overall difference. Nonetheless, items used in the study do exhibit atypical features that are plausibly identified as biological (e.g., penguins cannot fly). Thus, typicality and membership ratings were predicted to show some

gradedness. Although limiting the features considered to biological ones may reduce graded responding overall, the question remains whether gradedness is reduced to zero or continues to be significant.

Method

Subjects. The subjects were 41 undergraduates at a large mid-western university. Participation in experiments was part of the requirements of an introductory psychology course. Twenty-one of these subjects participated in a procedure similar to that used in Study 1. Twenty participated in a procedure similar to that of Study 2. No subject had participated in either of the other studies.

Procedure. Subjects participated in one of the two procedures from the previous studies. In both cases, the procedures were slightly modified. Instructions informed the subjects to base their judgments on biological criteria. In the typicality and categorization task (Study 1), the hedge "biologically speaking" was added to each scale item (see Lakoff, 1987; Malt, 1990). Thus, subjects were asked to choose between responses such as "biologically speaking a donkey is not at all typical of a horse," or "biologically speaking a donkey is completely a horse." For the dispute task (Study 2), the two positions were described as explicitly concerned with biological category membership, as in the example "John says the animal is biologically a monkey." In presenting the item, subjects were asked whether or not there were any "biological facts" that could settle the dispute one way or the other. Aside from these changes, the procedures were kept the same.

Stimuli. The stimuli for the two components of Study 3 were the animal and transformation items from the appropriate previous study. Four items were added to typicality and categorization task (*penguin* and *pterodactyl* as birds, *eel* and *shark* as fish).

Results

Typicality and category membership judgments.

Table 3 presents the mean typicality and categorization judgments for animal items as well as for the control, surface, and deep transformation items. As in Study 1, items were of moderate typicality (the mean typicality for animals from Study 1 = 3.7). For animal items, 33% of categorization responses were absolute. For transformation items (excluding controls), 32% were absolute.

Extremity scores (along with confidence intervals) are presented in Table 4. These scores were used as the basis for the following analyses (see Study 1 for computation of scores). Categorization judgments for animal items were significantly less extreme than the maximum [$t(20) = -10.7, p < .0001$]. For these items, categorization scores were more extreme than typicality scores [$t(20) = 2.2, p < .05$]. Surface and deep transformation items received less extreme categorization ratings than did control items [$t(20) = 6.5, p < .001, t(20) = 8.0, p < .001$, respectively]. Membership ratings for deep items

Table 3
Mean Typicality and Categorization Ratings: Study 3

Rating		Typicality	Categorization
Animals		4.2	4.3
Control transformations	base	6.4	6.6
	final	3.6	3.6
Surface transformations	base	4.3	5.1
	final	3.2	1.9
Deep transformations	base	2.7	3.7
	final	4.7	3.7

Table 4
Extremity Ratings and 95% Confidence Intervals: Study 3

Rating	Typicality		Categorization	
	Extremity Score	Confidence Interval	Extremity Score	Confidence Interval
Animals	2.0	1.7–2.3	2.4	2.0–2.7
Control transformation	3.7	3.3–4.0	3.9	3.5–4.0
Surface transformation	2.3	2.0–2.7	2.7	2.3–3.1
Deep transformation	2.3	1.8–2.7	2.0	1.5–2.5

were less extreme than for surface items [$t(20) = 3.2, p < .005$]. In each case, typicality and membership judgments were equally extreme for transformation items.

Dispute resolution. Table 5 presents the mean proportion of disagreements allowed for animal and transformation items. Confidence intervals are also indicated. Six (of 20) subjects allowed one or fewer disagreements for animal items [binomial probability of 1 or fewer out of 8 = .03, $p(\text{success}) = .5$]. However, 6 subjects accepted seven or more of the disagreements for these items (also $p < .05$). Of the remaining subjects, 3 answered “don’t know” to all animal items. This rate of acceptance is roughly equal to that observed in Study 2 (Study 2 mean = .29; 5 of 20 subjects accepted 50% or more of the disagreements). Significantly more disagreements were accepted for deep transformations than for controls ($N = 20, p < .001$, Wilcoxon test), but surface transformations did not differ from controls (Wilcoxon test).

Discussion

Category membership judgments remained significantly graded even with the addition of the biological hedges. Subjects frequently accepted partial membership in animal categories. Surface and deep transformations were thought to affect category membership as well as typicality. The focus on biology did lead subjects to treat the surface transformations in the opinion task as definitely resolvable. The biological effects of surface transformations may not have been enough to legitimate disagreements but were, perhaps, enough to reduce category membership by some degree. In other cases, subjects did recognize the possibility of irresolvable disputes. Nontransformed items were often seen as lacking resolution (about 1/3 of the time). Results for dispute items were strikingly bimodal. People either believed that all the animal disputes could be resolved definitely or believed that all disagreements could not be resolved. The addition of the hedges seemed to clarify responses to disputes, but not reduce gradedness. Overall, the extent to which subjects gave graded membership deci-

sions and accepted disagreements seemed largely unchanged from Studies 1 and 2.

Focusing category membership questions on biological classification was not effective at reducing graded responses. Subjects in this study did not treat categories as especially “all-or-none” when explicitly asked to consider biological criteria. This suggests that the graded responses in Studies 1 and 2 were not simply due to subjects’ focusing on extraneous or idiosyncratic conceptions of the categories. Graded responses did not result just because subjects were making (absolute) biological judgments somehow mitigated by social or psychological concerns. Of course the results of Study 3 are essentially negative results; adding the hedge did not make a difference. These data suggest, though, that a sense of crisp, absolute biological categories is at least difficult to elicit.

GENERAL DISCUSSION

The studies presented in this paper applied new methods to answer the old question of whether membership in animal and artifact categories is a matter of degree. In Study 1, subjects were asked to make both typicality and category membership ratings on the same set of stimuli as a way of distinguishing these judgments. This study also included control items (items expected to show absolute and graded membership) and a scale designed to allow graded or absolute responses. Study 2 involved judging whether disagreements regarding category membership could be definitely decided. It was argued that these judgments should reflect beliefs about the gradedness of the categories involved. In a third set of studies the same methods were used, but subjects were asked to focus on biological criteria for deciding membership in animal categories.

Results from the three sets of studies are generally similar. Subjects recognized degrees of membership in animal and artifact categories. This conclusion is based on absolute measures and on comparisons with control items. Animal and artifact categories were more graded than categories such as *U.S. currency* or *unit of measure*. However, there were significant proportions of absolute judgments for animal categories; more so than for artifacts or graded controls. This seems to support arguments that the probabilistic structure of natural kinds is “tighter” than that of artifacts (Barr & Caplan, 1987), though samples chosen for this study may not be repre-

Table 5
Mean Proportion of Disagreements Allowed: Study 3

	Mean	95% Confidence Interval
Animals	.33	.15–.46
Control transformation	.00	.00–.00
Surface transformations	.04	.00–.13
Deep transformation	.47	.25–.67

sentative. More generally, these results are consistent with previous findings of graded membership (e.g., McCloskey & Glucksberg, 1978; Oden, 1977; Rosch, 1975). Existing claims had been suspect because they failed to make the case that their conclusions were based on category membership judgments rather than typicality judgments. The studies presented in this paper represent an important advance, because the distinction between typicality and category membership was specifically addressed. Thus the finding of graded membership ratings in Studies 1–3 cannot be due to a simple conflation of typicality and category membership. Graded membership in animal and artifact categories is a robust phenomenon.

Data from the studies above contradict essentialist predictions. Categories of living things have been offered as the best example of essentialism (Atran, 1987). The implication of these studies is that not all animal (or artifact) categories have essentialist structure. This is a modest conclusion, but it is sufficient to rule out strong claims. A response might be that only some living thing categories have essences. Although the items in this report were chosen only for low typicality (not to be an accurate sample of the population), they do seem to be representative of common living thing categories. If categories such as *monkey*, *dog*, *horse*, and *bird* do not have essentialist structure, what reason is there to believe other categories will? Although some categories may have essentialist structure, many common living thing categories do not.

The studies presented here also inform recent work on the distinction between typicality and category membership (Keil, 1989; Rips, 1989). The two kind of judgments do seem to be different, at least for some categories. Categorization ratings, for animals and defined controls, tended to be more absolute than typicality ratings. This did not hold for artifacts or graded controls. The results also suggest that it is difficult to manipulate typicality independently of category membership. Transformations of animals seemed to reduce category membership (to a degree) as well as typicality. Features that make up an animal's appearance do play some role in determining its category membership as well as its typicality. Category membership does not seem to be solely determined by an invisible essence (e.g., DNA), though this may be the most important feature. The same features involved in typicality also affect categorization. This suggests that the difference between the two kinds of ratings occurs in the way features are evaluated or weighted. Dichotomous membership judgments on other tasks (Gelman & Markman, 1986; Keil, 1989) may have masked the effects of "nonessential features" on category membership.

One explanation for the findings is that ambiguous questions led to the graded responses. Lakoff (1987) has suggested that multiple models may underly our familiar categories. For example, there may be several different contexts in which someone has to judge what species an animal belongs to. Within any one context, the decision would be clear and absolute. When subjects are asked "in

general" to judge category membership, the criteria from the various contexts compete and result in a graded response. People might give a graded response when asked whether a mule is a horse but give an absolute response when asked if a mule could run in a horse race or pull a horse-drawn carriage. Along the same lines, people may differ in their criteria for categories. Person A holds that birds must have feathers, while person B does not. Each individual's concept of bird is absolute, but they are unwilling to impose their criteria on others. According to both these lines of argument, further specifications of context should reduce graded responding.

In the final set of studies, subjects were asked to make categorization judgments within a specified context. If people believe in an underlying essence for animal categories, that essence would seem likely to be biological in nature. Specifying that judgments be based on biological criteria should have removed any ambiguity in the tasks and focused subjects on the underlying (biological) essences. This did not seem to be the case. Membership judgments remained graded; they were largely the same as judgments made without a specified context. Of course, the manipulation may have been insufficient. Subjects may have ignored the instructions or seen the term *biological* as itself ambiguous (does it refer to genetics, evolution, ecology?). If categorization decisions are dependent on very specific contexts, then methods for studying concepts which rely on property inferences (e.g., Gelman, Collman, & Maccoby, 1986) may be preferable to those which ask for membership judgments.⁷ The multiple-models approach is useful as long as there are a few, repeatable, contexts. As the number of different contexts increases, the utility of treating categories as composed of stable components decreases and alternate approaches become more plausible (Barsalou, 1987).

One concern with regard to these studies is the effect of expertise. The subjects were college students, who may have constituted a special sample. Experts in a field often disagree about the definitions of terms. Categories must be continually updated to reflect new discoveries in a field (Enç, 1976). This is especially true in the field of biology, where experts deny there are essences for species (Dupré, 1986) and debate the correct basis for defining species (i.e., the debate between phenetics and cladistics). The college students might have had enough expertise to be aware of this uncertainty. More generally, some social commentators have argued that college students tend to be relativists and deny any absolute forms of knowledge (Bloom, 1987). Replicating these studies with laypeople (noncollege adults, children) might produce more absolute responses.

The question of whether category membership is all-or-none or a matter of degree cannot receive a single answer. Different categories may have different structures (Keil, 1989; Lakoff, 1987; Malt, 1990). There does seem to be a set of categories that have an essence (or intension) distinct from a set of superficial features. Some of these are nominal kinds such as *even number* or *U.S. currency*. There is also a set where there seems to be

little or no difference between the features used for categorizing and those used for judging typicality; perhaps artifact categories fit this model. Animal categories may fit somewhere in between. Clearly some features are very important for category membership, but many other features seem to have some smaller role to play.

It may be precisely because the categories of animals are natural kinds, discovered rather than invented, that people are less willing to dismiss features as totally irrelevant to category membership. Features that we think of as being a part of the prototype (as opposed to the core) of animal categories (such as appearances, vocalizations, and behaviors) are part of what we discover about animals (Locke, 1707/1961). Only under particular scientific theories are these features ruled inessential to category membership. Although biological theories may have criteria for assigning animals to species, it is not clear that laypeople do (or should) attempt to follow scientific usage (Dupré, 1981). A commitment to a taxonomic system is based on beliefs about the importance of particular properties and causal relationships (Dupré, 1981; Enç, 1976). A theory that gives primary importance to coloration will develop a different set of categories than will one that stresses interbreeding. Thus, to accept that some experts have the criteria for assigning category membership is to accept that those experts can determine what is most important about the category. Perhaps people are unwilling to cede so much authority and prefer to remain agnostic (or, as Dupré would say, "promiscuous") in their judgments of the fundamental essences of categories.

One important result of essentialist models has been to focus attention on the underlying causal connections between the features associated with categories. For a long time, categories had been represented as simple lists of features with no internal organization. Essentialism posits that the "root" cause of features is attributed unique importance as the determinant of category membership. The properties (known or otherwise) that cause other ones will be the basis of category membership. The essentialist hypothesis is that we categorize things on the basis of the source of their observable features.

However, causal role may be only one of a number of roles that determine the importance (weight) of a feature. It is well accepted that a host of factors can affect the role a feature plays in determining typicality (e.g., agents' goals; Barsalou, 1985). It seems that a similar combination of considerations can give features a role in determining category membership. Focusing on the single category of *water*, Malt (1994) argues that at least four dimensions play a role in classification. In addition to chemical composition, the source, location, and function affect people's decision to call something *water*. Similarly, the atypical stimuli used in the present studies differed in a number of ways (perceptually, functionally, etc.) from typical instances. These dimensions seem to affect not only an object's typicality but also its category membership.

Rather than being based on a single metric, category membership may be an emergent measure reflecting various roles and purposes of component features. There are various reasons for categorizing objects together. Capturing identities in underlying causes may be only one among many.

REFERENCES

- ARMSTRONG, S. L., GLEITMAN, L., & GLEITMAN, H. (1983). What some concepts might not be. *Cognition*, **13**, 263-308.
- ATLAN, S. (1987). Folkbiological universals as common sense. In S. Modgil & C. Modgil (Eds.), *Noam Chomsky: Consensus and controversy* (pp. 247-268). Philadelphia: Falmer Press.
- BARR, R. A., & CAPLAN, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition*, **15**, 397-418.
- BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instantiation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 629-654.
- BARSALOU, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge: Cambridge University Press.
- BLOOM, A. (1987). *The closing of the American mind*. New York: Simon & Schuster.
- DUPRÉ, J. (1981). Biological taxa as natural kinds. *Philosophical Review*, **90**, 66-90.
- DUPRÉ, J. (1986). Sex, gender, and essence. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy: Vol. 11, Studies in essentialism* (pp. 441-457). Minneapolis: University of Minnesota Press.
- ENÇ, B. (1976). Spiral dependence between theories and taxonomy. *Inquiry*, **19**, 41-71.
- GELMAN, S. A., COLLMAN, P., & MACCOBY, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, **57**, 396-404.
- GELMAN, S. A., & MARKMAN, E. M. (1986). Categories and inductions in young children. *Cognition*, **23**, 183-209.
- GELMAN, S. A., & WELLMAN, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, **38**, 213-244.
- KAY, P., & MCDANIEL, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, **54**, 610-646.
- KEIL, F. C. (1987). Conceptual development and category structure. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 175-200). Cambridge: Cambridge University Press.
- KEIL, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- LAKOFF, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago press.
- LOCKE, J. (1961). *An essay concerning human understanding* (Vol. 2, 5th ed.). New York: E. P. Dutton. (Original work published 1707)
- MALT, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory & Language*, **29**, 289-315.
- MALT, B. C. (1994). Water is not H₂O. *Cognitive Psychology*, **27**, 41-70.
- MALT, B. C., & JOHNSON, E. C. (1989). Does function provide a core for artifact concepts? In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 742-748). Hillsdale, NJ: Erlbaum.
- MCCLOSKEY, M. E., & GLUCKSBERG, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, **6**, 462-472.
- MEDIN, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, **44**, 1469-1481.
- MEDIN, D. L., & ORTONY, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). New York: Cambridge University Press.
- MILLER, G., & JOHNSON-LAIRD, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

- ODEN, G. C. (1977). Fuzziness in semantic memory: Choosing exemplars of subjective categories. *Memory & Cognition*, 5, 198-204.
- OSHERSON, D., & SMITH, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35-58.
- PUTNAM, H. (1975). The meaning of meaning. In H. Putnam, *Mind, language and reality: Philosophical papers* (Vol. 2, pp. 215-271). Cambridge: Cambridge University Press.
- REY, G. (1983). Concepts and stereotypes. *Cognition*, 15, 237-262.
- RIPS, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York: Cambridge University Press.
- RIPS, L. J., & COLLINS, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General*, 122, 468-486.
- RIPS, L. J., SHOEN, 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, 573-605.
- SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- SMITH, E. E., & SLOMAN, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition*, 22, 377-386.
- TVERSKY, A., & GATI, I. (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 79-98). Hillsdale, NJ: Erlbaum.
- WALKER, S. J. (1992). Supernatural beliefs, natural kinds, and conceptual structure. *Memory & Cognition*, 20, 655-662.

NOTES

1. Because essential features are typically unobservable and people may not even know what the essential features are, membership decisions are based on inferences about the presence or absence of an

essence. This position also differs from philosophical versions, because no claims are advanced that categories really have essences, merely that people believe they do. This caveat also applies to the discussion of "metaphysical" aspects of categorization. Again, the issue is not what a category really is (what are the "true" criteria for *cat*, for example) but what people think (e.g., what they think it is to be a cat).

2. Alternatively, typicality judgments may involve a measure of central tendency, whereas membership decisions may be based on frequency of the same features. Rips and Collins (1993) suggest that similarity judgments are based on centrality, whereas membership judgments are based on frequency. Typicality judgments are seen to involve both computations. It is important to note that membership judgments in Rips and Collins's study were likelihood estimates: assessments of the probability that some instance is a member of the category. This is not necessarily the same as asking for degree of membership judgments.

3. This item and one other (*sofa as furniture*) had high typicality scores. An additional set of analyses excluding these items was conducted; the results did not differ significantly. In the body of the paper, results including these two items are reported.

4. Overall, 6% of the responses to categorization questions were "don't know."

5. For example, we may rely on experts to judge what is a good wine yet still believe that "good wine" admits of degrees.

6. Instances were not identified with a subordinate level label as in Study 1 (is a bat a bird?). This was done so that subjects would not be affected by their own beliefs about the way in which the object should be classified to answer the questions. For example, a subject who did not know whether or not a bat was a bird might not indicate that there was a correct answer that he/she didn't know. The procedure used in Study 2 ensures that subjects have no opinions about which way the item should be classified.

7. One limitation of the present studies is the reliance on linguistic measures. Membership and typicality were assessed by having subjects apply labels (e.g., is a penguin a "bird"?). While this method is standard and is used exclusively in the literature for and against graded membership, it is important to recognize that this is just one way of measuring category membership. For example, it might be argued that applications of labels are particularly subject to conversational or contextual factors.

APPENDIX A

Instructions from Study 1

In the experiment today you will be asked to make judgments about how some items are categorized. You will be presented with pairs of items, for instance, "Rock Musician" and "Artist," and asked to make two kinds of judgments.

In the first case, we would like you to rate how typical the first item ("Rock Musician") is of the category ("Artist"). This part of the study has to do with what people have in mind when they use words which refer to categories.

For example, think of artists. You have some notion of what a "real artist," an "artistic artist" is. To me a painter or a sculptor is a typical artist while a Rock musician is a less typical artist.

Notice that this kind of judgment has nothing to do with how well you like the thing. You may prefer Rock musicians without thinking they most represent what people mean by artist.

In the first question on each page, you are given an instance and asked how good an example it is of a category.

Don't worry about why you feel that something is or isn't a good example of the category. And don't worry about whether it's just you or people in general who feel that way. Just mark it the way you see it. If you don't understand one of the words, or are really unsure of your answer, put a check where it says "don't know."

After you make the typicality rating, you will be asked to judge whether or not, and to what degree, the instance belongs to the category. In other words, is a Rock musician really and completely an Artist, only partly an Artist, or not at all an Artist? Sometimes, an instance of a category is just as much a member of that category as anything else. You might think that a Rock musician deserves to be called an artist just as much any other artist. Other times, things are more or less members. For instance, someone might think that a Rock musician is an artist, but not as much so as a painter or sculptor. This person believes that being an artist is a matter of degree. When answering the questions about category membership, think about whether membership in the category is all-or-none or a matter of degree.

In this rating we are not asking you how certain you are about membership. Thus we do not want you to rate how easy (or hard) it might be to tell that a Rock musician was an Artist. Rather we want you to tell us whether you think a Rock musician is as much an Artist as any other Artist, is only an Artist to some degree, or is not an Artist at all.

Again, don't worry about why you feel that something is or isn't a total member of the category. And don't worry about whether it's just you or people in general who feel that way. Just mark it the way you see it. If you don't understand one of the words, or are really unsure of your answer, put a check where it says "don't know."

APPENDIX B

Study 1 Items	Mean Extremity Scores		Study 1 Items	Mean Extremity Scores	
	Typicality	Category Membership		Typicality	Category Membership
U. S. Currency			a Pillow	2.68	2.45
a \$15 Bill	3.45	3.97	a Sofa	3.87	3.87
a \$2 Bill	2.61	3.83	a Wheelchair	3.06	3.45
a \$500 Bill	2.77	3.59	Weapon		
a Bicentennial Quarter	2.03	3.77	a Car	2.68	2.29
a Silver Dollar	2.03	4.00	a Fork	2.16	1.71
Unit of Measure			a Rubber Band	2.68	2.55
a Cubit	2.52	3.57	a Satellite	2.53	2.41
a Decimeter	2.10	3.77	Dog		
a Fathom	1.96	3.67	a Coyote	1.87	2.48
a Light-Year	2.60	3.83	a Hyena	2.21	2.93
a Micron	2.47	3.88	a Wolf	1.90	2.23
Good Grade			Horse		
a B	1.77	1.73	a Donkey	2.19	1.97
a B+	2.32	2.57	a Mule	1.80	1.69
an A	3.58	3.87	a Zebra	2.39	2.56
an A–	3.16	3.42	Insect		
Satisfactory	1.97	1.97	a Caterpillar	2.43	3.07
Furniture			a Praying Mantis	2.11	3.30
a Clock	2.39	2.48	a Scorpion	2.30	2.82
a Piano	2.16	2.61	a Spider	3.00	3.60
a Picnic Table	2.06	3.03			

Transformation Item	Mean Extremity Scores	
	Typicality	Category Membership
Control1. A pet store owner decided to sell a porcupine in her store. She didn't have much room, so she kept the porcupine in a cage with her mice. The animal didn't change but just lived in a cage labeled "MICE" and was surrounded by mice.		
Rated as Mouse	3.61	3.77
Control2. A farmer bought a baby sheep. He didn't have much room on his farm so he kept the animal in the same pen with his goats. This animal didn't change, but lived in a pen surrounded by goats.		
Rated as Sheep	3.03	3.58
Control3. The Sea World park was making some repairs to the seal tank. The biologists put a seal into the same tank as some dolphins. When they were done, the animal didn't change, but lived in a big tank marked "DOLPHIN" and was surrounded by dolphins.		
Rated as Seal	3.31	3.61
Control4. A small zoo bought a new frog. The zoo didn't have any cages ready for the frog, so they put it into a cage with a bunch of turtles. The animal didn't change, but lived in a cage marked "TURTLES" and was surrounded by turtles.		
Rated as Turtle	3.69	3.83
Surface1. Doctors took a Camel and did an operation that removed its hump. They dyed its fur black and white. Then they trained it to live in the mountains of South America instead of in a desert. When they were all done, the animal looked just like a llama.		
Rated as Llama	2.54	2.93

APPENDIX B (Continued)

Transformation Item	Mean Extremity Scores	
	Typicality	Category Membership
<p>Surface2. Some doctors found a very large duck. They bleached all the animal's feathers white and lengthened its neck and feet. The doctors changed the vocal cords so that the animal honked and didn't quack anymore. When they were done, the animal looked just like a goose.</p> <p>Rated as Goose</p>	2.84	2.81
<p>Surface3. The doctors took a raccoon and shaved away some of its fur. They dyed what was left all black. Then they bleached a single stripe all white down the center of its back. Then, with surgery, they put in its body a sac of super smelly odor, just like a skunk has. When they were done, the animal looked just like a skunk.</p> <p>Rated as Raccoon</p>	3.31	3.14
<p>Surface4. Some doctors took a big tiger and used a special fur bleach to take away its stripes. Then they sewed big mane onto its neck. When they were done, the animal looked just like a lion.</p> <p>Rated as Tiger</p>	2.86	3.63
<p>Deep1. Hazardous chemicals contaminated the vegetation that a bird ate. As time went by it began to change. The bird shed its feathers and sprouted a new set of wings composed of a transparent membrane. The animal abandoned its nest, developed a brittle iridescent outer shell, and grew two more pairs of legs. At the tip of each of the animal's six legs an adhesive pad was formed so that it was able to hold onto smooth surfaces; for example, it learned to take shelter during rainstorms by clinging upside down to the undersides of tree leaves. The animal eventually sustained itself entirely on the nectar of flowers.</p> <p>Rated as Insect</p>	1.89	2.32
<p>Deep2. This fish hatched from an egg laid near the waste pipe of a nuclear power plant which discharged irradiated water. As this fish grew it became long and thin. Its fins never developed. After a while this animal started spending more time breathing air above the surface of the water. Eventually it came to live on land where it slithered around catching and eating small bugs. By the time the animal matured, its head had flattened and it had developed a forked tongue.</p> <p>Rated as Snake</p>	2.30	2.69
<p>Deep3. This rabbit lived in a burrow near a pharmaceutical plant. One day the rabbit fell into a pool of chemicals which had spilled from the factory. Soon the animal's hair fell out and its skin became green and bumpy. Its ears fell off and the animal started to use its tongue to catch flying insects to eat. The animal still hopped, but now it spent most of its time resting in shallow ponds.</p> <p>Rated as Rabbit</p>	3.13	3.17
<p>Deep4. When it was very young, this kitten caught a strange disease. This disease caused the kitten to stay very small and develop a large, bushy tail. The animal spent most of its time climbing in the trees searching for nuts, which it cracked and ate. The animal never purred or meowed and eventually came to live in a hollow in a tall tree.</p> <p>Rated as Cat</p>	3.13	2.38

APPENDIX C

Study 2 Items	Proportions of Disagreements	Study 2 Items	Proportions of Disagreements
Defined		Artifacts	
Female	.16	Tool	.85
Male	.11	Weapon	.90
Square	.10	Chair	.60
Triangle	.10	Magazine	.47
Even No.	.15	Cup	.75
Odd No.	.15	Lamp	.35
		Table	.58
Graded		Vehicle	.63
Red	.75		
Purple	.85	Transformed Items	
Tall	.90	Surface1	.35
Friendly	1.00	Surface2	.35
Funny	1.00	Surface3	.30
Offensive	1.00	Surface4	.15
		Deep1	.60
Animals		Deep2	.50
Dog	.20	Deep3	.56
Bird	.25	Deep4	.55
Monkey	.39	Control1	.05
Lizard	.42	Control2	.15
Elephant	.37	Control3	.15
Beetle	.33	Control4	.05
Rodent	.42		
Mammal	.05		

(Manuscript received April 1, 1994;
revision accepted for publication June 28, 1994.)