A Re-examination of Graded Membership in Animal and Artifact Categories

Charles W. Kalish
Department of Psychology
3433 Mason Hall
University of Michigan
Ann Arbor, MI 48109
Charles_Kalish@um.cc.umich.edu

Abstract[1]

Previous studies of gradedness have 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. The present paper reports the results of two studies that attempt to overcome limitations of past methods. 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 study was based on the notion that there may be no definitive answer to questions about membership in graded categories. Thus, disagreements about membership in all-or-none and graded categories may have different qualities Stimuli included animal and artifact categories as well as animals that had undergone different kinds of transformations. Results from both studies suggest some support for claims that membership in animal and artifact categories is graded.

Graded Categories

Some categories clearly have all-or-none membership (e.g., "even number", "square"). Others obviously admit degrees (e.g., "red": things can be more or less red). 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). Whether or not membership in these categories is graded has important implications. For example, based in part on beliefs about the all-or-none nature of membership, it has been argued that mixed models or essentialist models are most appropriate for natural kinds (Keil, 1989). Yet what is the basis for these beliefs about category gradedness? Do we have good means for determining when membership is all-or-none and when it is a matter of degree?

At one time, there was wide consensus that membership in all 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, Rosch & Mervis, 1975). However, questions about the validity of existing measures of category membership suggest we should re-examine these conclusions.

Many of the studies arguing for graded membership have failed to distinguish between typicality and categorization (Rey, 1983). Studies which asked subjects for typicality ratings (e.g., "How representative/typical/characteristic is a penguin of a bird") were often taken as evidence about category membership (see Lakoff, 1987). Even when methods did not call for explicit typicality ratings (e.g., memory tasks or judgments of sentence appropriateness) there has been no way to determine whether results reflected the gradedness of typicality or the gradedness of category membership. We don't know which studies have been measuring typicality and which categorization.

The failure to distinguish between the two kinds of judgments is problematic because graded typicality may not indicate a gradedness of category membership. For example, 34 will receive a low typicality rating as an "even number" yet is a much a member of the category as is a high typicality even number (e.g, 2;

[1]This research was supported by an NSF Graduate Fellowship.

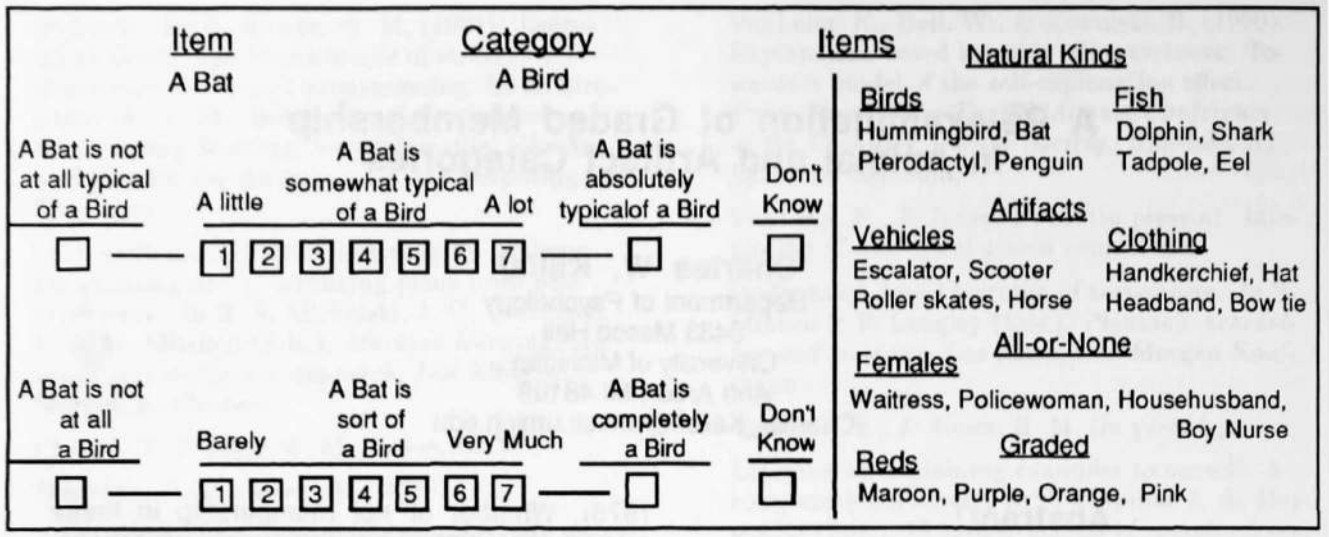

Figure 1: Stimuli for Study 1

Armstrong, Gleitman & Gleitman, 1983). Other studies (Keil, 1989; Rips, 1989) have demonstrated that certain kinds of transformations may affect an animal's typicality but not change the category the animal is assigned to. However, a crucial question not answered by these studies is whether or not the degree of category membership remains unchanged after the transformations. In other words, can category membership be reduced by degrees?

The fact that gradedness of typicality does not imply gradedness of categorization is not, in itself, positive evidence about the nature of category membership. Such evidence must come from studies which specifically address the question of category membership. In the rest of the paper, we present the results of two studies which attempted to asses category membership rather than typicality.

Study I

The purpose of this study was to collect data that reflected judgments of category membership rather than typicality (Rey, 1983; Rips, 1989). In Study 1 subjects were asked to make both typicality and categorization judgments for the same stimuli. It was reasoned that this procedure, along with careful instructions and control items, would tend to distinguish the question of categorization from the question of typicality. These categorization ratings could then be examined for evidence of gradedness. Included in this study were items depicting animals which had undergone certain alterations. These items were included to see if transformation might alter degree of membership.

Methods

Subjects. Subjects were 19 University of Michigan undergraduates participating in experiments as part of the requirements of an introductory psychology course.

Procedure. Subjects were asked to make two judgments about the relationship of an instance (e.g., a robin) to a category (e.g., bird); first its typicality and then its degree of membership. Subjects first read brief instructions describing typicality and categorization, and the differences between the two. They were then presented with 32 computer "screens." One instance and one category were presented on each screen. 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. A modified Likert scale was used to collect both ratings. The scale included two absolute end points (0 and 8) and a graded scale in the middle (1-7) (see Figure 1). This was done to minimize demand characteristics for either absolute or graded responses. Subjects were also able to indicate "don't know" to either rating.

Stimuli. There were four instances to be rated for each of eight categories. All instances were chosen to have low typicality. Eight instances were to be rated as members of natural kinds. Eight instances were rated as artifacts. There was

one control category which was assumed to have all-or-none membership and graded typicality and one control category with graded membership and graded typicality (see Figure 1). Subjects were also asked to rate instances of natural kind categories that had undergone different kinds of transformation. Four items depicted surface transformations (following Keil, 1989; see Appendix 1, item 1). Four other items described more radical, deep transformations (following Rips, 1989; see Appendix 1, item 2). These transformed items were always the last eight screens. Items were otherwise presented in random order.

Results

Figure 2 presents the mean proportion of absolute (0 or 8) answers for a given kind of categorization and typicality judgment. Approximately 80% of the judgments about membership in the category "Female" were absolute. Categorization responses to natural kind and artifact items were compared with responses to "Female" items (the baseline) to test for gradedness. The difference between each subject's mean number of absolute responses to "Female" items and his/her mean response to the test category was computed (Female-target). If the only difference between the test items and "Female" was error variance, this difference should average zero; if the test items had fewer absolute ratings, this difference should be positive. Animal instances were judged as more graded than "Female" instances ($t(18) = 2.3$ $p<.05$). Graded responses were more frequent to transformation items than to "Female" items (surface vs Female, $t(18) = 2.5$ $p<.05$; deep vs Female, $t(18) = 8.5$ $p<.001$). Artifact items also showed high levels of graded responses ($t(18) = 6.6$ $p<.001$). Large differences between items within a kind were also found. Within "Bird," 40% of categorization responses were graded to "Penguin," while less than 10% were graded for "Hummingbird." Analyses of responses to typicality judgments, however, revealed pattern analogous to that found with categorization judgments (see Figure 2). Instances of "Female" were given more absolute typicality ratings than were instances of animals, for example. Only 46% of the typicality judgments for "Female" were absolute, though, compared with 80% of the categorization judgments.

Discussion

Categorization judgments were predominantly absolute for females and largely graded for reds. Comparisons show that membership judgments for artifact categories and natural kind categories showed a significant degree of gradedness compared with controls. However, a similar pattern of results would have been obtained by comparing typicality ratings. Typicality may have been affecting categorization judgments. It is unlikely that subjects were simply using typicality as the basis of both typicality and categorization

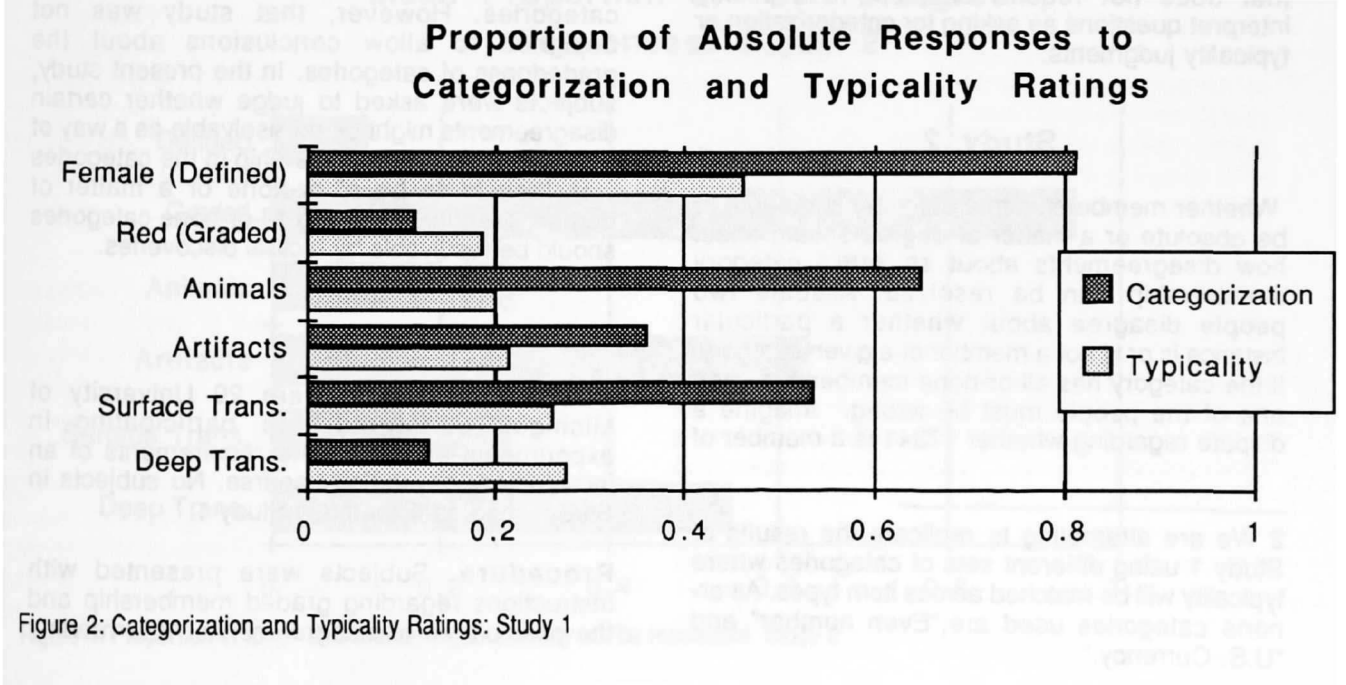

Figure 2: Categorization and Typicality Ratings: Study 1

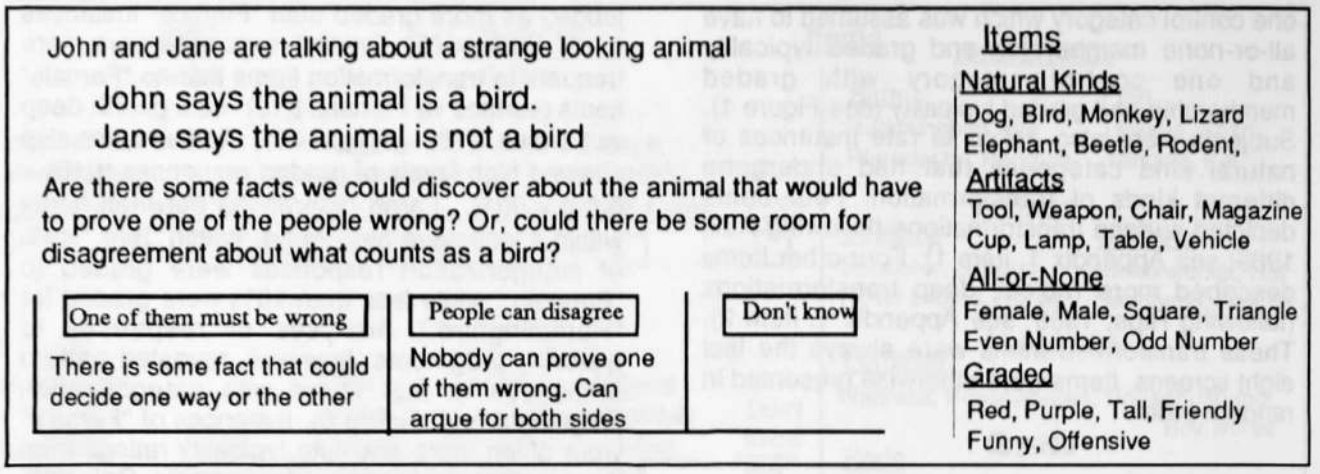

Figure 3: Stimuli for Study 2

judgments because there were significant differences in the percentage of absolute responses to the two kinds of judgments (for some items). However, we cannot conclude with certainty that the categorization ratings truly reflected subjects' beliefs about whether category membership is a matter of degree or not; we might still be measuring the gradedness of typicality[2].

The results of study 1 can be taken as suggestive that subjects view membership in natural kind and artifact categories as a matter of degree. However, such results depend on subjects explicitly distinguishing typicality from categorization. Study 2 reports an attempt to replicate the results of Study 1 using a method that does not require subjects to explicitly interpret questions as asking for categorization or typicality judgments.

Study 2

Whether membership in a category is thought to be absolute or a matter of degree should affect how disagreements about an item's category membership can be resolved. Assume two people disagree about whether a particular instance is or is not a member of a given category. If the category has all-or-none membership, then one of the people must be wrong. Imagine a dispute regarding whether 17341 is a member of the category "prime number." We may not know the answer, but presumably there are some facts we could uncover that would prove the issue one way or the other. If the category admits degrees of membership, on the other hand, then it is possible for there to be no way to demonstrate that one position is incorrect. Since an object may be partly a member of the category, there may be unresolvable disputes about whether it is "enough" of a member to be so identified. 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. A similar method was used by Malt (1991; studies 4&5) in a study of people's beliefs about the completeness of their representations of categories. However, that study was not designed to allow conclusions about the gradedness of categories. In the present study, subjects were asked to judge whether certain disagreements might be unresolvable as 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 be resolvable by factual discoveries.

Methods

Subjects. Subjects were 20 University of Michigan undergraduates participating in experiments as part of the requirements of an introductory psychology course. No subjects in Study 2 had participated in Study 1.

Procedure. Subjects were presented with instructions regarding graded membership and the possibilities of disagreements. After reading

2 We are attempting to replicate the results of Study 1 using different sets of categories where typicality will be matched across item types. All-or-none categories used are,"Even number" and "U.S. Currency."

these instructions they were presented with computer screens describing an encounter with an unusual object and a disagreement about how that object should be categorized. 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). The disagreement was then presented as to how to categorize the object (e.g., John says the animal is a lizard; Jane says the animal is not a lizard). For each screen, subjects were asked to indicate whether one of the people must be wrong or whether it was possible for people to legitimately disagree (see Figure 3).

Stimuli. Stimuli were 32 disagreements. Four disagreements involved categories with absolute membership; four involved graded categories . Eight disagreements revolved around animals and eight concerned artifacts (see Figure 3). Eight scenes involved a transformed animal (four "surface" transformations and four "deep" transformations). To give subjects an idea of the diversity they should consider, the eight transformation items were presented first. Order of presentation was otherwise random.

Results

Figure 4 presents the proportion of subjects who indicated that the disputants could legitimately disagree. To assess gradedness, responses to target categories were compared with responses to "defined" items. As in Study 1, the difference in each subject's mean number of "can disagree" responses to "defined" and target items was calculated (target-"defined"). If these differences were significantly positive, then this means that more disagreements were accepted for the target category. Subjects were significantly more likely to answer "can disagree" to animal items than they were to "defined" items ($t(19) = 3.3$, $p<.005$). Transformation items also differed from "defined" items ($t(19) = 1.8$, $p<.05$ and $t(19) = 4.4$, $p<.001$ respectively). Disagreements were more often accepted for artifact items than for "defined" items ($t(19) = 6.7$, $p<.001$) . Again there were some differences between items of the same kind; 42% of subjects accepted disagreements for categorizing lizards, while only 20% allowed disagreements for dogs.

Discussion

Results from Study 2 largely support the findings of Study 1. Again, the validity of the measure can be confirmed by the predicted performance on control items. Very few disagreements were accepted for defined items, while disagreements were largely accepted for graded items. Subjects allowed a significant number of disagreements for both natural kind and artifact categories. This suggests that membership in both natural kind and artifact categories may be a matter of degree.

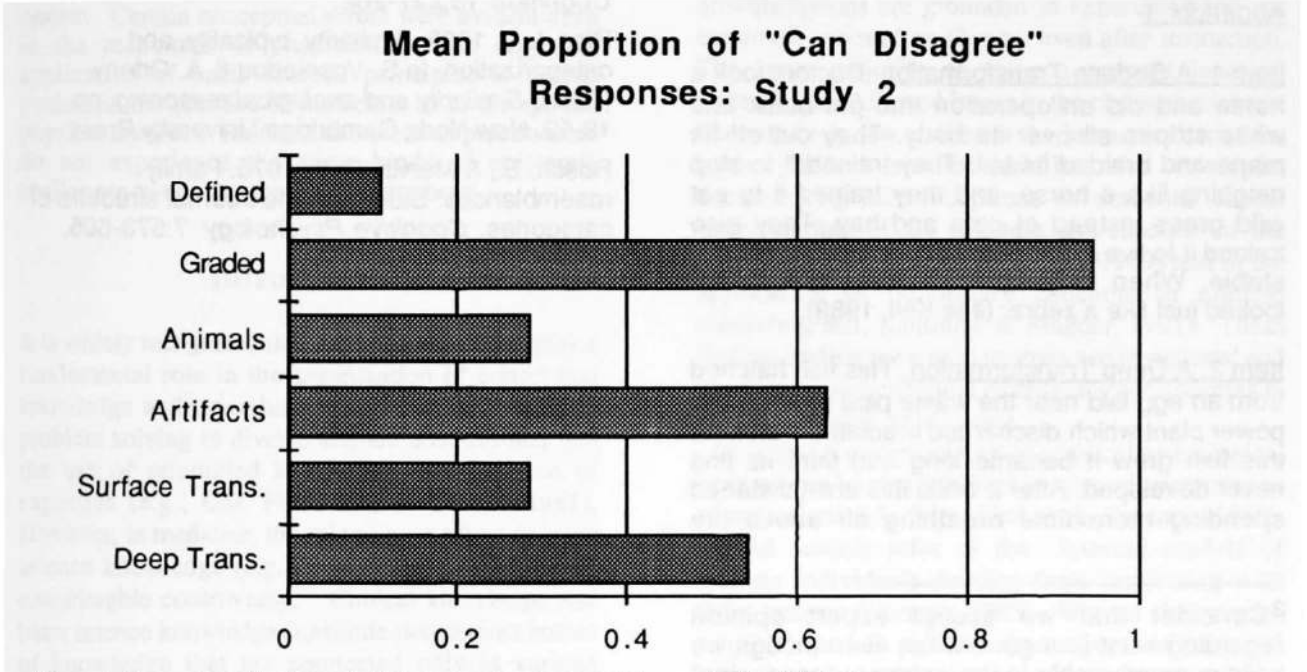

Figure 4: Proportion of judgments that an argument may not be resolvable, Study 2

General Discussion

Our findings generally support earlier conclusions drawn from studies of typicality effects (e.g., Rosch & Mervis, 1975). However, in these studies we have been very careful to distinguish questions of categorization from questions of typicality. In apparent contrast to our findings, Malt (1991) has argued that subjects' believe experts are able to make some definitive categorization decisions. One possibility is that people might be inclined to accept expert opinion even though they believe categories to be graded.[3] On the other hand, gradedness of membership may be the result of having multiple sets of criteria for membership (Lakoff, 1987). Pilot work on Study 2 revealed that many subjects who accepted disagreements about the way animals could be categorized reasoned that the disputants might be using different criteria for category membership. For instance, one might have perceptual, functional and/or biological criteria for categorizing something as a bird. Any one set of criteria, used by one type of expert, may provide an absolute categorization, but multiple sets would allow graded responses. Gradedness could arise because of uncertainty or disagreement about which criteria/expert were relevant. We are currently conducting a study to test this possibility using linguistic hedges (e.g., "biologically speaking") to focus subjects on particular criteria.

Appendix 1

Item 1: A Surface Transformation. Doctors took a horse and did an operation that put black and white stripes all over its body. They cut off its mane and braided its tail. They trained it to stop neighing like a horse, and they trained it to eat wild grass instead of oats and hay. They also trained it to live in the wilds of Africa instead of in a stable. When they were all done, the animal looked just like a zebra. (see Keil, 1989)

Item 2: A Deep Transformation. 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. (see Rips, 1989)

[3]Consider that we accept expert opinion regarding what is a good wine, even though we believe membership in the category "good wine" to be a matter of degree.

References

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. 1983. What some concepts might not be. *Cognition*, 13:263-308.

Barr, R. A., & Caplan, L. J.1987. Category representations and their implications for category structure. *Memory & Cognition*, 15:397-418.

Keil, F. C. 1989. *Concepts, kinds, and cognitive development.* Cambridge: MIT Press.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Malt, B. C. 1991. Features and beliefs in the mental representation of categories. *Memory & Cognition* 29:289-315.

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

Oden, G. C. 1977. Fuzziness in semantic memory: Choosing exemplars of subjective categories. *Memory & Cognition* 5:198-204.

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. 19-59. New York: Cambridge University Press.

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