The United States copyright law (Title 17 of the US Code) governs the making of copies of copyrighted material. A person making a copy in violation of the law is liable for any copyright infringement. Copying includes electronic distribution of the reserve materials by the user. The user should assume that any works in the reserve items are copyrighted.

Memory & Cognition
1983, 11 (3), 211-227

Ad hoc categories

LAWRENCE W. BARSALOU
Emory University, Atlanta, Georgia 30322

People construct ad hoc categories to achieve goals. For example, constructing the category of "things to sell at a garage sale" can be instrumental to achieving the goal of selling unwanted possessions. These categories differ from common categories (e.g., "fruit," "furniture") in that ad hoc categories violate the correlational structure of the environment and are not well established in memory. Regarding the latter property, the category concepts, concept-to-instance associations, and instance-to-concept associations structuring ad hoc categories are shown to be much less established in memory than those of common categories. Regardless of these differences, however, ad hoc categories possess graded structures (i.e., typicality gradients) as salient as those structuring common categories. This appears to be the result of a similarity comparison process that imposes graded structure on any category regardless of type.

The study of natural categories has been limited mostly to common categories such as "birds," "furniture," and "fruit." However, the use of highly specialized and unusual sets of items pervades everyday living. Some examples are "things to take on a camping trip," "possible costumes to wear to a Halloween party," and "places to look for antique desks." Since categories like these often appear to be created spontaneously for use in specialized contexts, I refer to them as ad hoc categories. Theories of natural categories primarily reflect what we have learned from common categories. By further considering ad hoc categories, we may discover a more general theory of categorization for which common and ad hoc categories are special cases.

This introduction first addresses two central properties of common categories: graded structure and well established category representation in memory. A comparison-network model is then proposed that accounts for these properties in common categories. The following section shows how this general model can also account for related predictions in ad hoc categories. Before going on to four experiments that address these predictions, an additional theoretical difference between ad hoc and common categories, the degree to which they reflect correlational structure, receives brief discussion.

Central Properties of Common Categories

Graded structure. As noted by Mervis and Rosch (1981) and Smith and Medin (1981), the discovery of graded structure has had a major impact on theories of categorization. Graded structure has three aspects. First, some instances are better examples of a category than are others; "chair" is a more typical example of "furniture" than is "bookcase." This aspect of graded structure has been found in all common categories investigated so far. Rosch (1973, 1975b) found typicality in color categories (e.g., red, green). Rips, Shoben, and Smith (1973) and Rosch (1973, 1975a) found typicality in common semantic categories (e.g., fruit, clothing). Rosch and Mervis (1975), Rosch, Simpson, and Miller (1976), Smith, Shoben, and Rips (1974), and Tversky (1977) have since argued that the typicality of a category member increases as it becomes more similar to other category members. The second aspect of graded structure is the presence of unclear cases, items whose category membership is uncertain (McCloskey & Glucksberg, 1978); people are not sure whether "radio" belongs to the category of "furniture." The third aspect of graded structure is that the nonmembers of a category (i.e., its complement) vary in how similar they are to the concept of the category; "typewriter" is more similar to the concept of "stereo equipment" than is "dog." This aspect has commonly been cited as the reason some false items take longer to reject in the category verification task than do others (McCloskey & Glucksberg, 1979; Smith et al., 1974); "bat" takes longer to reject as a member of "birds" than does "chair." No evidence bears on whether such similarity gradients cause subjects to reliably rate nonmembers of a category for typicality within the complement of the category. But, given the strong relation between judged typicality and category verification for category members, reliable judgments of typicality for nonmembers as well would not be surprising. In summary, graded structure is a continuum of category membership, ranging from prototypical members through

I am particularly grateful to Gordon Bower for supporting this work and to Herbert Clark for assistance in writing this paper. I am also grateful to colleagues at Stanford and Emory universities and to reviewers for helpful discussion and comments. The research was supported by Grant MH 13950 from the National Institutes of Mental Health to Gordon Bower and by a National Science Foundation graduate fellowship to the author. A brief summary of the work supported in this paper was presented at the American Psychological Association Convention in Toronto, 1980.

Copyright 1983 Psychonomic Society, Inc.

unclear cases to prototypical nonmembers (cf. Zadeh, 1965).

Well established category representation. The second property of common categories this paper focuses on is their possession of well established category representations. Organization of information to be remembered is clearly central to recall performance (Bousfield, 1953; Mandler, 1967; Tulving, 1962) and is also important to recognition performance (Mandler, Pearlstone, & Koopmans, 1969). Taxonomic organization, almost exclusively employing common categories, has been one of the primary types of organization to receive attention. That taxonomically organized lists are better recalled than lists of unrelated words has been demonstrated on numerous occasions (e.g., Bower, Clark, Lesgold, & Winzenz, 1969; Cofer, 1967; Puff, 1970). Such organization also results in clustering during recall (e.g., Bousfield, 1953; Bousfield & Cohen, 1953). Most investigators have viewed the effects of taxonomic organization as reflecting the existence of preexperimental structure in memory. Words from common categories are easily organized during encoding because memory structures for these categories assimilate presented information. Retrieval is facilitated during testing because these structures provide a network for locating presented information. When such organization does not exist (e.g., for a list of unrelated words), subjects have more difficulty organizing and retrieving a list because (1) subjects have less relevant structure to begin with, and (2) the structures used may be created during learning and therefore not be well established. Further evidence of well established category representations for common categories comes from the free association literature. The highest associates of many words are often the names of common superordinate categories and the names of common contrast categories (e.g., "chair" as a cue often produces "furniture" and "table," respectively). Superordinates and contrast categories are high associates presumably because well established memory structure interrelates these categories.

A Comparison-Network Model

This model, which accounts for graded structures and well established category representations in common categories, contains two interrelated components: a similarity comparison process and a spreading activation network.

The similarity comparison process. This process computes the similarity of two concepts in working memory. Generally, I will assume that similarity is some function of the concepts' properties, and specifically, I will assume that this function is along the lines of Tversky's (1977) contrast model. Tversky's account states that two concepts become more similar as the number of properties shared by them increases and the number of distinctive (i.e., nonshared) properties decreases. His model also allows for weighting the importance of properties and the importance of common vs. distinctive property sets (cf. Ortony, 1979).

Graded structure in a category results from computing how similar the concepts for instances, unclear cases, and noninstances are to the concept for the category. The properties in a category concept are those occurring most often for category instances and least often for noninstances (see probabilistic concepts in Smith & Medin, 1981). The central assumptions are: (1) as instances become more similar to a category concept, they become more typical of the category; (2) as noninstances become less similar to a category concept, they become more typical of the category's complement; and (3) the similarity of unclear cases to a category concept is close to the minimum amount necessary for category membership.

Rosch and Mervis (1975) report data for common categories consistent with this model. They found that the more similar an instance is to all other category members (i.e., its family resemblance), the more typical it is of the category. Assuming that a category concept is the average (in some sense) of all category members, how similar an instance is to the category concept should be at least highly correlated with (if not the same as) how similar the instance is to all other instances. So, the finding that typicality correlates highly with family resemblance is consistent with typicality depending on how similar category instances are to their category concept.

The spreading activation network. This network represents concepts and properties as nodes and represents associations between concepts and properties as pathways that carry spreading activation. A more complex but very similar network model can be found in Collins and Loftus (1975). Loftus (1975) and Rosch (1975c) provide additional comments of interest to this discussion. In the model I am proposing, each concept is associated to properties characteristic of the concept's referents in the environment. In addition, properties can be associated to each other and concepts can be associated to each other. Associations have labels such as "has" (e.g., a robin has wings), "cooccurs" (e.g., "gills" cooccurs with "swims"), and "is an instance of" (e.g., "robin" is an instance of "bird"). Strength of association is free to vary continuously and increases as a function of how frequently and recently an association has been active in working memory. Associations can also be asymmetrical; that is, the strength of association from Node X to Node Y may not be the same as the strength from Node Y to Node X. Although activation arrives at the terminal nodes of both weak and strong associations equally fast (Lorch, 1982; Ratcliff & McKoon, 1981), greater activation accumulates at the terminal nodes of strong associations than of weak ones in a fixed time period. A node is active in working memory when the total amount of activation arriving at it is above some threshold value. When a node becomes active, there is

competition for activation among associations leaving it (i.e., the fan effect; Anderson & Bower, 1973) only for weak associations. As found by Hayes-Roth (1977), well learned associations do not compete with each other for activation.

This network model accounts for the well established category representations of common categories in three ways: Common categories have well established concept-to-instance associations, well established instance-to-concept associations, and well established category concepts. All are well established because of frequent and recent processing in working memory. The performance implications of each type of structure are discussed in turn.

(1) Strong concept-to-instance associations in common categories enable category concepts to easily activate category instances. Such top-down associations are useful when one is trying to generate category members during category production tasks or trying to recall information from a categorized list. For example, activating "furniture" might activate "chair," "table," "desk," and so on. In addition, Barsalou (1981) and Mervis, Catlin, and Rosch (1976) report that typical instances generally have stronger concept-to-instance associations than atypical instances.

(2) Strong instance-to-concept associations in common categories enable instances to activate their category concepts. Such bottom-up associations are useful for categorizing single instances and for organizing multiple instances of the same category. For example, perceiving the words "oak," "maple," and "pine" all activate "tree," which can be used for the purposes of categorization and organization.

(3) The category concepts for common categories are well established in memory because the associations between a concept and its properties and between the properties themselves are well established. For example, "bird" is highly associated to "wings," "flies," "feathers," and so on, which are highly associated among themselves. To the extent a category concept is well established, it should be easier to locate in memory. This follows from the assumption that well established concepts are more "visible" to a memory scanning mechanism or from the assumption that well established concepts have more pathways into them from other information in memory.

The similarity comparison process interfaces with the network in that concepts entering the comparison process are concept-property node sets activated above threshold. Following Barsalou (1982), only a subset of a concept's properties is usually active. This active subset may contain (1) context-independent properties that are active on all occasions the concept is processed, and (2) context-dependent properties that are activated only by relevant contexts. For example, "basketball" may activate "round" on all occasions, but it may activate "floats" only in contexts involving bodies of water.

Ad Hoc Categories and the Comparison-Network Model

The comparison-network model is sufficiently general to make predictions for ad hoc categories as well as to explain the previous findings for common categories. Predictions regarding graded structure and category representation in ad hoc categories are addressed in turn.

Graded structure. The comparison-network model predicts that ad hoc categories should exhibit graded structure. If graded structure results from concepts bearing different amounts of similarity to a category concept and if concepts vary in similarity from the category concepts of ad hoc categories (a safe assumption), then ad hoc categories should exhibit graded structure. In contrast, it is possible that ad hoc categories may not be processed in the same way as common categories. Instead, they may be processed as true equivalence classes upon which the similarity comparison process does not operate. Ad hoc categories may be represented more as lists without internal structure than as categories possessing typicality gradients. Experiment 1 addresses whether ad hoc categories can possess graded structure.

As proposed in the next section, an important difference between ad hoc and common categories is that ad hoc categories do not have well established category representations in memory. However, it is hard to see how this difference would affect the similarity comparison process, assuming this process is found to generate graded structure in ad hoc categories. Experiments 2a and 2b address whether the graded structures of poorly established categories (e.g., ad hoc categories) differ from those of well established categories (e.g., common categories).

Category representation. A central difference between common and ad hoc categories appears to be that common categories have well established category representations in memory, whereas ad hoc categories do not. Ad hoc categories are not well established simply because people rarely, if ever, think of them. This precludes the development and strengthening of associations between the nodes representing them. Besides observing whether lack of established category representation eliminates graded structure in ad hoc categories (as just discussed), the experiments to follow explore three ways that this lack of structure may cause the processing of ad hoc categories to differ from the processing of common categories.

The first way lack of category representation may affect the processing of ad hoc categories centers on concept-to-instance associations. These associations enable a category concept to act as a cue to activate category instances during category production and recall. Unlike common categories, ad hoc categories may not have direct associations from their category concepts to their instances. If so, retrieving instances from

ad hoc categories should be more difficult than retrieving instances from common categories. Experiments 2a and 3 explore this prediction.

The second way lack of category representation may affect the processing of ad hoc categories centers on instance-to-concept associations. These associations enable people to activate categories to which an instance belongs. Unlike common categories, ad hoc categories may not have direct associations from their instances to their category concepts. If so, categorizing instances into ad hoc categories should be more difficult than categorizing them into common categories. Experiment 4 explores this prediction.

The third way lack of category representation may affect the processing of ad hoc categories centers on how well established their category concepts are. Unlike common categories, the properties composing ad hoc category concepts may not be well associated, since these properties have rarely, if ever, been processed simultaneously. If so, the category concepts of ad hoc categories should not be as accessible as those of common categories. Experiment 3 explores this prediction.

The Relation of Ad Hoc and Common Categories to Correlational Structure

An important difference between ad hoc and common categories not addressed empirically in this paper, but certainly worthy of future attention, centers on the correlational structure of the environment. As Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) have noted, properties of entities in the environment are not independent but, instead, form clusters of correlated properties. For example, if an entity has feathers, there is a much higher probability that it flies and builds nests than that it swims and has gills. Entities instantiating a set of correlated properties are very similar to each other and are very different from entities instantiating other sets of correlated properties. For example, different kinds of birds are very similar to each other and are very different from members of fish and vehicles. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) show that people are sensitive to the correlational structure of the environment and that they prefer to use categories that take maximal advantage of it.

What I have been referring to as common categories appear to reflect correlatonal structure. They circumscribe sets of entities that share many correlated properties and that do not share many properties with members of other categories (Rosch & Mervis, 1975). We readily perceive these sets as categories because they have so much in common (although, as Rosch and Mervis show, no property or correlation of properties need be true of all a category's members).

In contrast, ad hoc categories appear to violate the correlational structure of the environment. Consider the category of "things to take from one's home during a fire," which contains members as diverse as "children," "dog," "stereo," and "blanket." This category's instances do not appear to share correlated properties. Moreover, its instances share many correlated properties with entities in the complement of the category. For example, "dog" also belongs to the common category of "mammals," many of which belong to the complement of "things to take from one's home during a fire."

If ad hoc categories cut across the correlational structure of the environment, then why do people perceive them as categories? Most likely, this is because ad hoc categories are instrumental to achieving goals. For example, someone trying to escape a burning home and minimize loss might try to construct the category of "things to take from one's home during a fire" before heading for safety. Similarly, someone interested in taking a trip would need to consider "things to pack in a suitcase." To the extent that someone is achieving either of these goals for the first time, the corresponding categories should not have well established category representations in memory. As noted in the general discussion, however, frequently used ad hoc categories may develop well established category representations much like those of common categories.

Before proceeding to experiments that address graded structure and category representation in ad hoc categories, it is necessary to comment on the sampling of ad hoc categories. Ad hoc categories will be defined for the purpose of this paper as sets that (1) violate correlational structure and (2) are usually not thought of by most people. Clearly, there are an indefinitely large number of such categories, and it would probably be impossible to enumerate them all. At this point, it is even difficult to imagine what kinds of ad hoc categories exist. For this reason, these experiments do not attempt to draw conclusions about all ad hoc categories but concentrate instead on the nature of a few of them. Although these experiments can primarily be interpreted with respect to existential as opposed to universal claims, they provide constraints on a general theory of categorization. Finding that some uncommon categories exhibit graded structure forces a general theory to include a mechanism that can generate such structure in a wider range of categories. Finding that categories vary in how well established they are forces a general theory to include mechanisms that account for this difference. And finding that graded structures occur in poorly established categories further constrains the mechanism responsible for generating graded structure.

EXPERIMENT 1

Rather than possessing graded structure, ad hoc categories may be represented as unordered lists in which all instances are equally good members. Alternatively, the same similarity comparison process that generates graded structure for common categories may also operate during the processing of ad hoc categories.

This experiment and the next investigate whether ad hoc categories possess graded structure. Do subjects reliably perceive typicality gradients in these categories? Are subjects uncertain regarding the membership of certain instances, namely, unclear cases? And do subjects reliably perceive some nonmembers to be more typical nonmembers than others?

Method

Design. Subjects performed two categorizing operations on eight sets of six items. For each set, they were asked first to separate out those items that belonged to some category and, second, to rank order all six items for how good a member each was of that category.

Materials. Each subject received a booklet containing instructions followed by eight pages of category materials, one for each category. Each page contained a context vignette, the name of an ad hoc category, and six randomly ordered labels in a column, some of these denoting instances of the category. The categories were "things to inventory at a department store," "ways to make friends," "things that conquerors take as plunder," "nouns," "ways to escape being killed by the Mafia," "things that babies do," "times to write a term paper," and "things that could fall on your head."

Each context vignette described a person engaged in an activity. The category label that followed denoted a category relevant to the person's goals. The vignettes were used to establish the ad hoc categories in goal contexts. For each of four item sets, three items were obviously from the a priori category and the other three were not. For each of the other four sets, two items were obviously from the a priori category, two were not, and the remaining two were, intuitively, unclear cases. Table 1 contains examples of the contexts and item sets.

The instruction sheet directed subjects to read both the vignette and the category label, to look through the six items that followed, and to circle those belonging to the category; there was no constraint on the number they could circle. Next, they were to rank all six items from the best example of the category to the worst, with no ties.

Subjects and Procedure. Twelve undergraduates participated for either course credit or pay. Half received one randomized version of the list, and the other half received a different randomized version. Subjects worked through the booklets at their own pace and had as much time as necessary to complete the experiment.

Results

Unclear cases. Agreement for category membership for a given item in an item set was determined as follows. If a subject had circled an item, it was scored as +1; if the subject had not, it was scored as −1. The scores for the item were then summed across subjects. If all 12 subjects circled the item, it received a score of +12; if no subject circled the item, it received a score of −12; if half the subjects circled the item, it received a score of 0. The absolute value of the score was divided by 12. This measure ranged from 0 to 1, 0 indicating no agreement (for an unclear case) and 1 indicating complete agreement. The absolute value was taken to measure agreement for nonmembers as well as for members of the category. The overall agreement for an item set was simply the average of the agreement scores for the six items in that set.

It should be noted that the average number of items circled per item set was 3.15, with subjects circling

Table 1
Examples of Materials Used in Experiments 1 and 4

AD HOC CATEGORIES

Ways to Make Friends

Martin had moved from the midwest to the west coast over a year ago. He had encountered much trouble making friends since he had arrived in California and could not think of anyone he presently considered a good friend. He decided it was time to do something about it.

Experiment 1 Item Set

join a card playing club
get convicted for murder
don't take a bath more often than once a month
go back to school
have a garage sale
get convicted for burglary

Experiment 4 Item Set

get involved in local politics
get rich
have a garage sale
go to school

Ways to Escape Being Killed by the Mafia

Roy was in big trouble. The Mafia had a contract out on him for double-crossing them. He knew he couldn't continue living in Las Vegas or he'd be dead in a week. So he started thinking quickly about alternatives.

Experiment 1 Item Set

change your identity and move to the mountains of South America
move to the remote reaches of Wyoming*
stay where you're presently living in Las Vegas
move to Reno*
move to the mountains of Mexico
change where you're living in Las Vegas

Experiment 4 Item Set

move to the remote reaches of Wyoming
sail around the world
go to Mexico
become a drunk in Detroit

COMMON CATEGORIES

Fruit

Dan thoroughly enjoyed food. His favorite time of the year was summer because of the abundance of fresh food that was available.

Experiment 4 Item Set

apple
orange
banana
peach

RANDOM CATEGORIES

Horace was designing a computer system that would operate the traffic signal system in a major urban area. He had to find a competent group of programmers to help him do the programming for the system.

Experiment 4 Item Set

blue
erase
riddle
monkey

Note—Only ad hoc categories were used in Experiment 1. The category labels were not presented in Experiment 4.
**Unclear case.*

3 items 70% of the time and circling 2-4 items 95% of the time. It can be easily shown that the overall agreement score should be .04 if subjects were guessing and had circled 3.15 items on the average.[1]

The mean agreement across item sets for this analysis was .88. The average agreement for item sets not having unclear cases was .97, compared to .79 for item sets with unclear cases (Mann-Whitney $U = 0$, $p = .014$). An analysis of the item sets having unclear cases showed that lower agreement for these sets was entirely the result of poor agreement for five of the eight a priori unclear cases; the average for these five items was .17, compared to .96 for all other items from these four sets. This demonstrates the presence of unclear cases in ad hoc categories; namely, subjects can be divided about whether certain items are members of ad hoc categories.[2]

Graded structure. Agreement for graded structure was determined by computing Kendall's coefficient of concordance across the subjects by items matrix of rankings for each item set. A transformation of this statistic (Guilford & Fruchter, 1973) estimates how much a given subject agrees with every other subject on the average. It is an estimate of subject agreement and not the stability of an item set's means. If ad hoc categories do not possess salient graded structures, then subjects should show no agreement, and this statistic should approach 0. To the extent there is a salient graded structure perceived by all subjects, this statistic should approach 1. Across categories, the average agreement was .87. Subjects' high agreement demonstrates ad hoc categories possess salient graded structure.

Agreement was also computed for clear category members alone and for clear category nonmembers alone. Average agreement was .54 for category members and .37 for category nonmembers, revealing both internal and external graded structure. Agreement for internal structure was greater than for external structure in only three of the eight item sets.

The presence of unclear cases did not affect subjects' typicality rankings. The average agreement was .92 for item sets having unclear cases and .88 for those without; this difference was not significant (Mann-Whitney $U = 4$, $p = .171$). So, unclear cases led to less agreement for category membership but not for typicality. Subjects may agree on the graded structure underlying a category but be uncertain where to draw the category boundary along this continuum.

Discussion

Ad hoc categories can possess graded structure. Subjects showed high agreement for overall graded structure in general and for internal and external graded structure in particular. Further evidence of graded structure in ad hoc categories stems from the presence of unclear cases. Excluding unclear cases, however, subjects were in excellent agreement over which items were and were not category members.[3]

EXPERIMENTS 2A AND 2B

As shown by Experiment 1, ad hoc categories can possess graded structure. But how do these structures compare with those found in common categories? To what extent might well established category representations in common categories cause their graded structures to differ from those of ad hoc categories?

As mentioned earlier, salient typicality gradients have always been found in common categories. In addition, common categories exhibit another form of graded structure: When subjects generate instances to a category name, instances vary in how frequently they are generated; some instances are better examples during generation than others. These two forms of graded structure, typicality and production frequency, are well correlated for common categories (Barsalou, 1981; Mervis et al., 1976).

Experiments 2a and 2b compare these two forms of graded structure for common and ad hoc categories. Experiment 2a compares the distributions of responses given to both category types during exemplar generation. In particular, are some ad hoc category instances more dominant than others during generation, as is the case for common categories? If so, are the most dominant responses for ad hoc categories as dominant as the most dominant responses for common categories? Experiment 2b compares typicality gradients for the two category types. In particular, do subjects agree as much in their typicality judgments for ad hoc categories as they do for common categories? Also, do ad hoc category instances vary as much in typicality as common category instances do? That is, are typicality gradients in ad hoc categories as "steep" as those in common categories?

The comparison-network model predicts that typicality gradients, as indexed by typicality judgments, should be very similar for ad hoc and common categories. This follows from the assumptions that (1) the same similarity comparison process constructs graded structure for both category types and (2) how well established a category is in memory does not affect this process. In contrast, the model predicts that subjects should show less agreement when generating category members for ad hoc than for common categories. This follows from the assumption that common categories have well established concept-to-instance associations to their typical instances, whereas ad hoc categories do not. Consequently, most subjects should access the typical instances of the common categories and show high agreement, whereas subjects' search of ad hoc categories should be more random and, therefore, show less agreement.

Method: Experiment 2a

Twenty subjects generated the first four exemplars that came to mind for nine common and nine ad hoc categories. Half the subjects received the 18 categories in one random order, and half

received them in another. Subjects were introductory psychology students participating to receive course credit. The common categories were 9 of the 10 common categories used by Rosch (1975a): "birds," "sports," "fruit," "weapons," "vegetables," "vehicles," "clothing," "furniture," and "tools." The ad hoc categories were what intuitively appeared to be atypical and infrequently used categories. Four of these were drawn from the previous experiment: "ways to make friends," "ways to escape being killed by the Mafia," "things that can fall on your head," and "plunder taken by conquerors." The other five were "things that can be walked upon," "things that can float," "things that are poisonous," "things that can attack something," and "things that have a smell."

Results: Experiment 2a

Within the 80 exemplars generated for each category, the average number of different exemplars per category was 20.33 for the common categories and 37.00 for the ad hoc categories [$t(16) = 6.54$, $p < .001$]. The distributions for average number of items per category as a function of average number of subjects generating an item are shown in Table 2. The functions for both the common and ad hoc categories were generally decreasing and negatively accelerated. However, there was an increase in frequency for the common categories for items generated by 16-20 subjects. Also, the function for the ad hoc categories decreased more rapidly than the function for the common categories. The average proportion of subjects producing the ith most generated item per category is shown in Table 3 (for $i = 1$-5). For each of the five levels shown, the proportion is higher for the common categories, this being significant by a sign test ($p < .05$). These three results—the average number of different items generated, the characteristics of the frequency as a function of agreement distributions, and the proportion of subjects generating the five most generated items—are all in agreement with the conclusion that exemplar production is more consistent for common than for ad hoc categories. Ad hoc categories, however, clearly show graded structure in the sense that some instances are more dominant than others.

Table 2
Effect of Category Type on Production Frequency Distribution: Average Number of Items per Category (Experiment 2a)

N	Category Type	
	Common	Ad Hoc
1-5	15.77	34.00
6-10	2.56	2.21
11-15	.77	.77
16-20	1.21	.00

Note—N = number of subjects (of 20) generating an item.

Table 3
Average Proportion of Subjects Producing the ith Most Generated Item per Category (Experiment 2a)

ith Most Generated Item	Category Type	
	Common	Ad Hoc
1	.84	.56
2	.69	.39
3	.46	.31
4	.36	.26
5	.27	.20

Method: Experiment 2b

For each of the 18 categories in Experiment 2a, the exemplars were rank ordered by production frequency (i.e., the number of subjects, from 1 to 20, generating an item). Six exemplars were then selected from each category, one from the highest level of production frequency, one from the lowest level (always 1), and the remaining four from, as much as possible, equally spaced intervals between the highest and lowest levels. Two versions of the stimuli were constructed. Each had the 18 categories in a different random order; within each version, the six exemplars from the same category were in a different random order. The six exemplars appeared in a column below their category label.

The instructions defined typicality and directed subjects to judge the typicality of each exemplar with respect to its category label. Twelve subjects rated the exemplars' typicality on a scale from 1 to 7, where 1 referred to a most unusual exemplar and 7 to one of the best. Twelve other subjects ranked the exemplars for typicality; no ties were allowed. Six subjects in each task received each of the two list versions. Subjects were 24 introductory psychology students participating to receive course credit.

Results: Experiment 2b

Ratings. For each of the 18 categories, the intraclass correlation coefficient (Guilford & Fruchter, 1973) was computed for the subjects by items matrix of ratings. This coefficient can be interpreted as measuring how much subjects agree on the typicality of a category's exemplars. Specifically, the coefficient estimates how much a given subject's ratings correlate with another's on the average; it does not measure how stable the mean for each item is. Its average value across categories within category types is shown in Table 4. Subjects agreed equally well for both category types [$t(16) = .94$, $p > .20$].

The average level of typicality differed marginally for common and ad hoc categories [$t(16) = 1.93$, $.10 > p > .05$; see Table 4]. The standard deviation of the mean typicality ratings for each category's items was computed, and there was no difference between category types [$t(16) = 1.34$, $p > .20$; see Table 4]. In fact, the average standard deviation for the ad hoc categories was slightly larger than that for the common categories. This latter finding indicates that the range of typicality values for items in common and ad hoc categories was comparable. This also indicates that a difference in range did not bias the comparison between intraclass correlations for the two category types. If the ranges for one category type had been less on the average, this could have relatively reduced the correlations for that type.

Rankings. For each of the 18 categories, a transformation of Kendall's coefficient of concordance was computed for the subjects by items matrix of rankings (Guilford & Fruchter, 1973). This statistic, like the intraclass correlation, can be interpreted as measuring how much subjects agree on the typicality of a category's exemplars. Unlike the intraclass correlation, Kendall's

Table 4
Summary Data for Experiment 2b

Measure	Category Type: Common	Category Type: Ad Hoc
Ratings		
Intraclass Correlation	.50	.56
Typicality	5.08	4.68
Within-Category SD	1.44	1.65
Rankings		
Coefficient of Concordance	.57	.54

Note—All measures are averaged across categories.

coefficient does not have the problem of possible differences in range. This is because the standard deviation of a subject's rankings within a set is constant. The average value of this statistic across categories within category types is shown in Table 4. Again, subjects agreed equally well for both category types [$t(16) = .36$, $p > .50$].

Intermeasure correlations. The following six correlations across item averages are all significant at the $\alpha = .001$ level. Typicality ratings correlated with typicality rankings .89 for the common categories and .89 for the ad hoc categories. The ratings correlated with production frequency (from Experiment 2a) .80 for the common categories and .66 for the ad hoc categories; the difference between these latter two correlations approached significance ($z = 1.55$, $p = .12$). The rankings correlated with production frequency .90 for the common categories and .65 for the ad hoc categories; the difference between these two correlations was significant ($z = 3.45$, $p < .001$). This difference suggests that high- and low-typicality instances of common categories may differ more in how strongly they are associated to their category concepts than do high- and low-typicality instances of ad hoc categories.

Discussion

As predicted by the comparison-network model, typicality gradients derived from typicality ratings and rankings were equally salient for the ad hoc and common categories. This suggests, first, that subjects use the same similarity comparison process to construct graded structure for both category types and, second, that this process is not affected by how well established a category is in memory.

Similar to common categories, ad hoc categories also showed graded structure as indexed by production frequency. Some members of ad hoc categories are more dominant during exemplar generation than others. As predicted by the comparison-network model, however, there was less consistency in exemplar production for ad hoc than for common categories. This suggests that the category types differ in the extent to which their concept-to-instance associations are established in memory. People have more experience with common categories and, therefore, establish stronger associations to these exemplars. Also, cultural forces may focus on typical members such that these instances become particularly well associated to their categories.

To provide further support for the claim that concept-to-instance associations are better established in common categories than in ad hoc categories, additional data were collected. Twelve new subjects generated exemplars to the nine common categories and seven of the ad hoc categories for a fixed time period (two ad hoc categories, "ways to make friends" and "ways to keep from being killed by the Mafia," could not be used, since their exemplars are typically described by more than one word). A tape recorder presented the category names to the subjects intermixed in one of two random orders. For 15 sec following the end of each name, subjects wrote down as many category instances as they could think of. Subjects generated 5.67 exemplars on the average for the common categories and 4.22 exemplars for the ad hoc categories [$t(14) = 5.58$, $p < .001$]. There was no overlap between the distributions of mean exemplars generated per category for the two category types. Stronger concept-to-instance associations for common than for ad hoc categories enabled faster access to the instances of common categories.

There is an alternative explanation for the differences between ad hoc and common categories in production consistency and access time: These ad hoc categories may have contained more exemplars than the common categories. Regarding consistency, there may simply have been a lower probability of sampling an item from an ad hoc than from a common category. With respect to access time, greater numbers of exemplars may lead to more interference and, therefore, slower access. Although both familiarity and category size may affect exemplar production, two of the present results provide evidence for the familiarity explanation. (1) In Table 3, the number of exemplars generated for common categories by 16-20 subjects was greater than the number generated by 11-15 subjects. This could not be the result of a difference in sampling probabilities, since the sampling explanation predicts frequency to be a monotonic decreasing function of the number of subjects. Rather, this "bump" appears to be the result of familiarity with the most prototypical exemplars. Notably, no such bump occurs for the ad hoc categories. (2) The correlations between typicality and production frequency were higher for common than for ad hoc categories. This suggests that the high- and low-typicality instances of common categories differ more than those of common categories in the strength of their concept-to-instance associations. If so, then the high-typicality instances of common categories may well have stronger concept-to-instance associations than the high-typicality instances of ad hoc categories. Experiments 3 and 4 provide further evidence that common categories are much better established in memory than are ad hoc categories.

It is of interest to note that robust typicality gradients occurred in categories defined by necessary and sufficient conditions. Many of the ad hoc categories were of the form "things that exhibit X," in which X was a

necessary and sufficient condition (e.g., things that have a smell, things that can float). Thus, unequal category membership exists even in well defined categories. This could be the result of either of two factors. First, a necessary and sufficient condition may be possessed by exemplars to varying degrees, in different manners, or with different frequencies. "Milk" and "coffee" vary in degree of "smell"; "perfume" and "skunk" vary in manner of "smell"; and "basketball" and "sailboat" vary in frequency of "floats." Such differences in possession of category criteria may be correlated with typicality. Seocnd, irrelevant properties associated with a necessary and sufficient condition may enter into typicality judgments. The criteria for "medical doctor" are well defined, but someone possessing irrelevant properties of the doctor-personality stereotype may be a better example of "doctor" than someone who does not. Barsalou (1981) offers direct evidence that these factors determine typicality in categories defined by necessary and sufficient conditions.

EXPERIMENT 3

The production frequency results of Experiment 2a suggest that concept-to-instance associations are better established in memory for common than for ad hoc categories. The current experiment tests this hypothesis more directly and also tests the hypothesis that the category concepts of common categories are better established than are those of ad hoc categories. One group of subjects received words from common categories, a second group received words from ad hoc categories, and a third group received clusters of words that were unrelated. For the common and ad hoc category lists, a category's instances were grouped and preceded by their category label. All subjects performed a free recall of the words, and the common and ad hoc subjects performed a subsequent cued recall.

Two measures of recall from categorized lists are of interest (see Tulving & Pearlstone, 1966): (1) the number of categories accessed during recall (a category is accessed if at least one of its exemplars is recalled), and (2) the average number of exemplars retrieved per accessed category. The predictions of the comparison-network model are as follows.

Category access. If common categories have better established category concepts than ad hoc categories, then the category concepts for common categories should be accessed more easily during free recall. This is in accordance with the well known fact that high-frequency responses are better recalled than are low-frequency responses (Hall, 1954). It can be explained by assuming that better established concepts are more "visible" to a scanning process or have more pathways into them from other information in memory. A consequence of category concepts being better established for common than for ad hoc categories is that it should be easier to retrieve the exemplars of common categories during recall via their category concepts. Consequently, common categories should be accessed more often than ad hoc categories. In addition, if ad hoc categories are accessed no more frequently than random categories, this would suggest that the concepts for ad hoc categories are indeed not well established in memory.

It should be noted that the exemplars of the common, ad hoc, and random categories were made equally accessible in this experiment by equating their word frequency and imageabliity. Any differences observed in category access, therefore, can be attributed to differences in how well established the category concepts are for each category type, but not to differences in the accessibility of their exemplars.

Exemplar retrieval. Two predictions follow from the hypothesis that concept-to-instance associations are better established for common than for ad hoc categories. First, more correct exemplars should be retrieved from common than from ad hoc categories for accessed categories. Second, there should be a higher intrusion rate for common categories. Both predictions follow from the assumption that stronger concept-to-instance associations result in instances more likely being activated above threshold during retrieval. Increased availability of instances via these associations may also facilitate organization at encoding.

Method

Design. Three groups of subjects each received 48 words partitioned into 12 clusters of 4 words each. One group received a list of common category exemplars blocked by category, a second group received a list of ad hoc category exemplars blocked by category, and a third group received a list of clusters containing unrelated words. Half the subjects for each category type received the word sets in one random order, and half received them in another. Within each order, the words for the same set appeared in a different random order. Six subjects were nested in each of the category type by order cells of the design.

All subjects performed a free recall of the list after it was presented. Subjects in the common and ad hoc category conditions were then given the category labels as cues for further recall. (Random category subjects could not perform a cued recall, since their list did not contain category labels.) In the common and ad hoc category conditions, half the subjects in each category type by order cell received the cues in one random order, and half received them in another. Before learning the critical lists, all subjects received and free recalled the same practice list, which contained 36 unrelated words similar to those in the critical lists.

Materials. The common categories, drawn from Battig and Montague (1969) and Rosch (1975a), were "furniture," "clothing," "vehicles," "birds," "sports," "fruit," "vegetables," "insects," "trees," "animals," "musical instruments," and "colors." The ad hoc categories, which seemed to be atypical and infrequently used, were "where eating can occur," "can fall on your head," "has a smell," "can be walked upon," "plunder taken by conquerors," "can attack something," "can be used for hitting," "manufactured by humans," "is poisonous," "can be eaten," "is a liquid," and "can float." All the exemplars were common single-word nouns. Exemplars from the common and ad hoc categories were typical of their respective categories. Those for "vegetables" were "potato," "corn," "celery," and "spinach"; those for "manufactured by humans" were "telephone," "helicopter," "camera," and "refrigerator"; those for one of the random categories were "grease," "spider," "admiral," and "copper."

The three lists were equated for Kučera and Francis (1967) frequency (21.96, 22.52, and 21.19 for the common, ad hoc, and random categories, respectively; $F < 1$). To insure that the stimulus sets were equivalent in imagery, 12 undergraduates rated how easy it was to image the referent of each word in the three lists. On a 1-7 scale, with 7 meaning highest imagery, the average ratings were 5.42, 5.32, 5.44, for the common, ad hoc, and random categories, respectively ($F < 1$).

Subjects and Procedure. The subjects, 36 undergraduates participating to receive course credit, were told they would be presented with two lists to learn for free recall. The practice list was presented by a slide projector, one word at a time at a 2.5-sec rate, followed by a buffer task lasting 3 min. Subjects were then asked to recall, for 5 min, as many words as possible in any order from the list.

Subjects were then informed about the categorized list to come. Subjects in the common and ad hoc category conditions were told not to recall the category labels, but that these would be useful in organizing and learning their list. Subjects in the random category condition were told to learn each cluster of four words as a group. All subjects were told about the format of their list, which was then presented by a slide projector, one item at a time at a 2.5-sec rate. For the common and ad hoc category conditions, a category label was presented, followed by the individual presentation of each exemplar. After presentation of the fourth exemplar, the label of the next category was presented and followed by the individual presentation of its exemplars, this cycle continuing until all categories had been presented. For the random category list, a string of Xs appeared in place of the category labels.

After working on a buffer task for 3 min, subjects attempted to recall, for 5 min, as many exemplars as possible in any order from the second list only; they did not recall the category labels. Subjects in the common and ad hoc category conditions were then given the category labels and again asked to recall, for 5 min, as many words as possible in any order from the list.

Results

There are four dependent measures of interest. "Overall exemplar recall" refers to the proportion of 48 category exemplars recalled by a subject. "Category access" refers to the number of categories for which a subject recalled at least one of the four exemplars. "Exemplars per category recall" refers to the average number of exemplars per category recalled by a subject (for only those categories in which one or more exemplars were recalled). "Intrusions" refers to the total number of intrusions made by a subject.

For each measure, a two-way ANOVA was performed on the free and cued recall data for the common and ad hoc category subjects. Comparisons involving the free recall data for the random category subjects were computed from a one-way ANOVA performed across the free recall data for all three category types. Except where noted, each ANOVA was performed twice for each dependent measure: once across subject averages and once across item (i.e., category) averages. The results from the subject and item analyses were then combined to perform min F', planned comparisons of interest (Clark, 1973). Analyses of the intrusion data were performed on frequencies; all other analyses were performed on proportions transformed using arcsins, as suggested by Winer (1971). The subject means from all analyses are shown in Table 5. There was no difference between the three list conditions on the practice list [min $F'(2,62) = 1.60$, $p > .10$].

Overall exemplar recall. Common category subjects recalled more exemplars than did ad hoc category subjects during free recall [min $F'(1,39) = 16.72$, $p < .001$] and also during cued recall [min $F'(1,39) = 13.40$, $p < .001$]. Cued recall was superior to free recall both for common category subjects [min $F'(1,38) = 14.70$, $p < .001$] and for ad hoc category subjects [min $F'(1,39) = 18.18$, $p < .001$]. There was no Recall Type by Category Type interaction (min $F' < 1$). Random category subjects recalled fewer exemplars during recall than did common category subjects [min $F'(1,64) = 8.65$, $p < .01$]; the difference between random category and ad hoc category subjects was not significant [min $F'(1,62) = 1.16$, $p > .25$], although it favored the ad hoc category subjects.

Category access. Common category subjects accessed more categories than did ad hoc category subjects during free recall [min $F'(1,36) = 9.51$, $p < .01$]. There was no significant difference for cued recall [min $F'(1,38) = 2.73$, $p > .10$], although it was in the same direction as that for free recall. No difference was expected, since providing cues should have equalized accessibility. More categories were accessed during cued recall than during free recall both for common category subjects [min $F'(1,36) = 15.14$, $p < .001$] and ad hoc category subjects [min $F'(1,35) = 28.29$, $p < .001$]. There was no Recall Type by Category Type interaction [min $F'(1,33) = 1.01$, $p > .25$]. The difference between random category subjects and common category subjects for free recall approached significance [min $F'(1,62) = 3.40$, $.10 > p > .05$]. Notably, there was no difference between random category subjects and ad hoc category

Table 5
Effects of Category Type on Free and Cued Recall (Experiment 3)

	Category Type				
	Free Recall			Cued Recall	
Measure	Random	Ad Hoc	Common	Ad Hoc	Common
Overall Proportion of Exemplars Recalled	.33	.42	.58	.59	.73
Number of Accessed Categories (of 12)	7.33	7.33	9.42	10.75	11.75
Average Number of Exemplars/Accessed Category (of 4)	2.15	2.72	2.94	2.59	2.96
Total Intrusions	1.08	.92	2.42	1.44	3.50

Note–All measures are the average per subject.

subjects (min $F' < 1$). In fact, the number of categories accessed was identical.

Exemplars per category recall. Common category subjects retrieved more exemplars per accessed category than did ad hoc category subjects, both during free recall [min $F'(1,37) = 5.30$, $p < .05$] and during cued recall [min $F'(1,42) = 18.64$, $p < .001$]. There was no difference between cued recall and free recall for the common category subjects (min $F' < 1$) or for the ad hoc category subjects [min $F'(1,43) = 2.35$, $p > .10$]. There was no Recall Type by Category Type interaction [min $F'(1,44) = 1.85$, $p > .10$]. Random category subjects during free recall retrieved fewer exemplars per accessed category than did common category subjects [min $F'(1,65) = 7.67$, $p < .01$] and ad hoc category subjects [min $F'(1,65) = 4.04$, $p < .05$].

Intrusions. Common category subjects produced more intrusions than did ad hoc category subjects, both during free recall [min $F'(1,43) = 7.63$, $p < .01$] and during cued recall [min $F'(1,43) = 14.72$, $p < .001$]. The difference between cued recall and free recall was marginally significant for common category subjects [min $F'(1,43) = 3.98$, $.10 > p > .05$], but not for ad hoc category subjects (min $F' < 1$). There was no Recall Type by Category Type interaction (min $F' < 1$). Intrusions did not differ during free recall between random category subjects and common category subjects [$F(1,33) = 2.22$, $p > .10$] or between random category subjects and ad hoc category subjects ($F < 1$).[4]

Discussion

Overall, common category subjects recalled more words than did ad hoc category subjects both for free and cued recall. Decomposing overall recall into category access and exemplar recall yielded similar results: Common category subjects accessed more categories than did ad hoc category subjects during free recall and retrieved more exemplars per accessed category during both free and cued recall. Common category subjects also produced more intrusions during free and cued recall than did ad hoc category subjects. All these findings are consistent with the conclusion that ad hoc categories are not as well established in memory as common categories are. The concepts for common categories and their concept-to-instance associations are better established in memory than are those for ad hoc categories.[5]

There is an alternative explanation for the difference in accessibility between ad hoc and common categories. Namely, the category labels were longer for the ad hoc categories and, therefore, may have been more difficult to retrieve. But if retention is related to length of category label, then this relation should hold within each category type. The correlation between number of letters for each category label and probability of accessing the category was computed separately for the ad hoc and common categories. The correlations were .04 [$t(10) = .13$, $p > .50$] and $-.16$ [$t(10) = .52$, $p > .50$], respectively. Thus, label length does not appear to have been a factor. Furthermore, subjects were instructed not to recall the category labels, but only to use them for organizing the word sets. This alternative explanation has no bearing on exemplar recall, which was computed only for accessed categories.

Ad hoc categories have some capability as mnemonic devices. This was evidenced by ad hoc categories showing greater exemplar recall per accessed category than random categories. Although ad hoc categories are not well, if at all, represented in memory, they are at least able to provide organizational schemes for presented information. The concepts for these categories, however, were no more accessible than the representations of random groups of words. In fact, the average numbers of categories accessed for these two category types were identical, 7.33 and 7.33.

EXPERIMENT 4

The previous two experiments show that category concepts and concept-to-instance associations are better established in common than in ad hoc categories. The current experiment attempts to show that instance-to-concept associations are better established for common than for ad hoc categories. Subjects received sets of exemplars to categorize. For example, "What category do moth, bee, gnat, and ant all belong to?" "What category do coffee, perfume, leather, and skunk all belong to?"[6] Each set contained exemplars from a single common category, exemplars from a single ad hoc category, or unrelated items. Half the subjects received a context vignette prior to each set that described a character trying to achieve a goal. The category instantiated by the exemplars was usually instrumental to achieving the goal; that is, the context primed the category. The remaining subjects received no such contexts. All subjects tried to generate a category label for each set and rated how easy it was to do so.

If instance-to-concept associations are not well established for ad hoc categories, then subjects without context should have difficulty labeling the ad hoc category sets and should show much variability in the labels they generate. This is because the ad hoc category concepts should be difficult to activate, and subjects may activate a wide range of concepts in the process of trying to classify these sets. Difficulty should be reduced, however, when these sets are preceded by contexts that prime the category concepts. In contrast, the common categories should be easy to label even without context, since their instance-to-category associations are well established in memory. As each item in a set is encoded, it should activate the same highly associated category concept. Consequently, subjects should find it easy to categorize these sets and should show much consistency in the labels they generate. It would

not be surprising if the common category sets were as easy to categorize without context as with it.

Method

Design. Subjects attempted to categorize 14 item sets. Eight were from ad hoc categories, three were from common categories, and three, as best as possible, were from no category. For each set, subjects provided (1) a label categorizing all four items (if one could be discovered), (2) a rating of how easy it was to find the label, and (3) a rating of how confident they were that the label was appropriate.

Half the subjects received a context vignette before each item set, and half received no specified context. For the ad hoc and common categories, each context provided a relevant setting for the category. The random categories bore no relation to their contexts. Half the subjects received the ad hoc categories in one random order, and half received them in the inverse order. For all subjects, the common categories occurred at Positions 1, 7, and 10, and the random categories occurred at Positions 2, 5, and 11. These categories provided subjects with examples of the easiest and most difficult categories initially and intermittently during the experiment; this allowed subjects to respond to the ad hoc categories relative to the anchor points established by the common and random categories.

Materials. Each subject received a booklet of instructions and category materials. For the context condition, each page of materials contained the context vignette, a blank line on which to write a category label (if discovered), a column of randomly ordered labels for the four exemplars, a response scale for indicating ease of category discovery, and a response scale for indicating how appropriate the label was for the four items. The two response scales were the integers from 1 to 7, 7 representing maximum ease and appropriateness, respectively. For the no-context condition, the category materials did not contain the context vignettes, but they were otherwise identical to the materials for the context condition.

The eight ad hoc categories were the same as those in Experiment 1. The common categories were "fruit," "birds," and "sports." The random categories were sets of items that intuitively did not constitute any category. Each context vignette described a character engaged in a goal-directed activity and primed the subsequent category. None of the vignettes contained the category label for the respective item set. Table 1 contains examples of the context and item sets.

Procedure and Subjects. Subjects in the context condition were asked to read each vignette and to find, if they could, a category to which all four items belonged. In the no-context condition, subjects were instructed to study the four items and to find a category to which all belonged. All subjects were told that phrases, as well as single words, could serve as category labels. They were also told that some item sets did not form a category, and that if no category was apparent, they should write "O" on the blank line above the items. Once they had generated a label or given up, they were to choose values on the ease and appropriateness scales.

The subjects, 24 undergraduates participating for either course credit or pay, were randomly assigned, 6 each, to the four context by order conditions. They worked through the booklets at their own pace.

Results

Ratings of ease and appropriateness. For each measure, one ANOVA was performed on the ad hoc categories and another on the common and random categories together. Planned comparisons contrasted means of interest between the two analyses. The results for the ratings of ease and appropriateness were equivalent in effects. Therefore, only the tests for ease will be reported, although the means for both ratings are shown in Table 6. As recommended by Clark (1973) and Winer (1971), subjects and categories were both treated as random effects in these analyses when appropriate. For the ad hoc categories, subjects found it easier to discover a category label with context then without [$F'(1,12) = 16.02$, $p < .01$]. There was a Context by Categories interaction [$F(7,140) = 6.52$, $p < .001$]: Some ad hoc categories were more difficult to discover than others, but less so with context than without. This interaction occurs in all analyses for ad hoc categories reported here, and its interpretation is the same in all cases.

Context had no effect on the ease of categorizing the common and random categories [$F'(1,9) = 1.00$], and there was no Context by Category Type interaction ($F' < 1$). However, common categories were much easier than random categories [$F'(1,140) = 308.39$, $p < .001$]. Although position in the list was not controlled for the common and random categories, varying this factor for the ad hoc categories did not have an effect or interact with context. This suggests that position was not a factor for the common and random categories as well, and that the context and category type results are valid for the common and random categories.

With context, the ad hoc categories were as easily discovered as the common categories [min $F'(1,37) = 1.88$, $p > .10$] and were more easily discovered than the random categories [min $F'(1,31) = 123.27$, $p < .001$] Without context, the common categories were easier to discover than the ad hoc categories [min $F'(1,25) = 10.32$, $p < .01$], and the ad hoc categories were easier than the random categories [min $F'(1,30) = 6.16$, $p < .05$].

Subjects' categorizations. Subjects generated labels for the common categories 100% of the time. For the ad hoc categories, they did so 97% of the time with context and 83% of the time without. For the random categories, they generated labels 14% of the time with context and 31% of the time without. With so few labels, the random categories were not considered in the following analysis.

Table 6
Effects of Context and Category Type on Average Ratings of Category Discovery (Experiment 4)

	Ease			Appropriateness		
Condition	Random	Ad Hoc	Common	Random	Ad Hoc	Common
Context	1.31	6.33	6.83	1.31	6.08	6.72
No Context	1.97	4.29	6.94	1.89	3.43	6.83

How much did different subjects' labels for a given item set denote the same category? For this analysis, the 12 (or sometimes fewer) category labels generated with context for each item set were typed in a random order on a page, and those generated without context for the same item set were typed on another. The resulting 22 pages were randomly ordered, half the booklets having the inverse order of the others.

Six judges were asked to group all the labels on a page that denoted the same category for as many categories as were perceived. To be grouped, two or more labels had to be very similar conceptually, but they did not necessarily have to share the same linguistic form. For each category concept having two or more instantiations, the judge wrote the category concept and indicated the labels instantiating it.

A normalized measure of agreement, A, was computed for each judge's analysis of the labels for each item set in each context condition. A is defined as [(number of category labels + 1)–number of category concepts]/number of category labels. "Number of category concepts" equaled the number of concepts a judge thought were instantiated by one or more labels for an item set. A equaled 1 when all the labels were instantiations of one and only one category concept. A equaled 1/number of category labels, at the lower bound, when no category concept had more than one instantiation.[7] The means from these analyses are shown in Table 7.

For the ad hoc categories, agreement was greater with context than without (.90 and .59, respectively) [$F'(1,9) = 16.08$, $p < .01$]; this corresponds to slightly more than two concepts per set with context and slightly less than six concepts without, for a set of 12 labels. There was also a Context by Categories interaction [$F(7,28) = 7.44$, $p < .001$].

Surprisingly, for the common categories, there was more agreement without context than with context; this difference was marginally significant [$F'(1,6) = 5.28$, $.10 > p > .05$]. Subjects' labels with context were primarily phrases relating the common category label, a single word, to the relevant character or context (e.g., John's favorite fruit, birds Mike saw). In contrast, subjects' labels without context were almost always the single label for the category (e.g., fruit, birds). Subjects in the context condition, by being more specific, were more variable in their labelings.

The relation between category discovery and labeling. Categories that were difficult to label were also the ones that showed the least consistency in labeling. Ease of perceiving a category was positively correlated with how well subjects agreed in labeling it. A correlated .94 with ease and .92 with appropriateness (both significant at the $\alpha = .001$ level). These correlations show that increased difficulty in discovering a category led to increased variability in subjects' categorizations.

Table 7
Effects of Context and Category Type on Agreement for Category Labeling (Experiment 4)

Condition	Ad Hoc	Common
Context	.90	.89
No Context	.59	.98

Note–See text for description of agreement measure.

Discussion

These data support the conclusion that instance-to-concept associations are better established in memory for common than for ad hoc categories. Without context, the ad hoc categories were difficult to identify, and subjects were highly variable in the categories they discovered. These category concepts only became obvious and agreed upon in relevant contexts that primed the concepts.

Because ad hoc categories are so specialized, it may be optimal that perceiving an entity does not activate all the ad hoc categories to which it belongs. Seeing a chair and having categories such as "emergency firewood," fits in the trunk of a car," and "used to prop doors open" come to mind would be highly distracting when these categories are irrelevant. Ad hoc categories should come to mind only when primed by current goals. Such priming does occur, as found in this experiment.

In contrast to ad hoc categories, context had no impact on ease of discovering common categories. The concepts for these categories were as available without context as they were with context. This shows that instance-to-category associations are much better established in memory for common than for ad hoc categories. Interestingly, subjects with relevant contexts were more variable in their categorizations of common category item sets than were subjects without relevant contexts. It appears that the contexts caused subjects to be more specific in these categorizations and that people in general may often tailor common categories to current contexts. That is, categories like "fruit," "furniture," and "clothing" may often be incorporated into ad hoc categories relevant to current purposes (e.g., fruit for dessert, furniture to be moved, clothing in the laundry).

A given entity can be cross-classified into an indefinitely large number of categories. For example, "apple" can be cross-classified into "fruit," "things to take on a picnic," "things that could fall on your head," and so on. The data from the current experiment, in conjunction with Barsalou's (1982) distinction between context-independent and context-dependent properties, suggest the following account of cross-classification. During the classification of an entity, categories with strong instance-to-category associations may be automatically activated. (e.g., "apple" may automatically activate "fruit"). Alba, Chromiak, Hasher, and Attig (1980) and Ross and Barsalou (note 1) provide further evidence that such automatic classifications exist. In contrast, categories weakly associated to an instance (e.g., "things to take on a picnic" for "apple") may be activated only in

contexts that require use of the category (e.g., going on a picnic). These context-dependent categories are not activated by the instance alone, but only by the conjunction of the instance and a particular context. In general, classifications highly associated to an instance are available across all contexts, whereas weakly associated classifications are only available in contexts that prime them.

GENERAL DISCUSSION

Categories other than the common taxonomic categories usually studied possess graded structure. Moreover, the ad hoc categories observed here had typicality gradients as salient as those in common categories: Ad hoc category instances varied as much in typicality as common category instances, and subjects showed equal agreement when judging typicality for both category types. As suggested earlier, the same similarity comparison process appears to construct graded structure in both common and ad hoc categories. Interestingly, this process appears unaffected by how well established a category is in memory.

Although the ad hoc and common categories were similarly structured, they differed in category representation. First, strong concept-to-instance associations for the common categories resulted in high consistency and fast access during exemplar production and facilitated the encoding and retrieval of relevant information presented for learning. Second, strong instance-to-concept associations in common categories resulted in highly available category concepts that facilitated categorization. And third, well established category concepts in common categories facilitated locating these categories during memory search. In contrast, lack of strong concept-to-instance associations for ad hoc categories resulted in less consistent and slower instance retrieval during exemplar production and made it harder to remember relevant information presented for learning. Lack of strong instance-to-concept associations in ad hoc categories resulted in less available category concepts, thereby making categorization difficult. And the category concepts for ad hoc categories were so poorly established that they were no more accessible than the representations of random word groups. Although ad hoc and common categories may be constructed by a shared process that bestows the same graded structure on each, other processes appear to result in the representations of common categories becoming better established in memory.

The Loss of Ad Hoc Status

Some ad hoc categories may be processed so frequently that their category concepts, concept-to-instance associations, and instance-to-concept associations all become well established in memory. At this point, these categories are no longer ad hoc by the definition I have been using. Even though they still violate correlational structure, their representations in memory are much more like those of common categories. Certain ad hoc categories appear to make this transition. "Things to sell at a garage sale" may start out as ad hoc for someone's first gargae sale but then become well established with subsequent ones. Similarly, someone taking up camping may have the category of "things to take on a camping trip" shift from being ad hoc to well established.

A phenomenon discovered by Alba et al. (1980) can be used to demonstrate this shift. These investigators presented subjects with instances of common categories ostensibly to learn for free recall. The number of instances presented per category ranged from three to nine. Immediately following presentation, subjects were unexpectedly asked to estimate the number of instances presented for each category (i.e., category frequency). Across a wide range of instructions, list organizations, and retrieval settings, subjects showed an unchanging and excellent sensitivity for category frequency. Alba et al attribute this to automatized instance-to-concept associations: Every time an instance is encoded, it automatically activates its category concept. Sensitivity to category frequency results from a count being kept at each category concept of how often it has been activated.

Ross and Barsalou (Note 1) performed a similar experiment with ad hoc categories. They initially found no sensitivity for category frequency, which they attribute to ad hoc categories not having strong instance-to-concept associations. But if subjects repeatedly process these categories for a week before the surprise frequency test, strong instance-to-concept associations develop, and equal sensitivity for ad hoc and common categories results. What were once ad hoc categories are no longer ad hoc.

Similar to the potential for ad hoc categories to become well established in memory is the necessity of common categories being poorly established in early childhood. Horton (1982) demonstrated this by showing that children's poor performance with common categories results at least in part from a lack of well established memory structure. Children assessed as having well established category representations did much better on standard taxonomic tasks than children assessed as having weak category representations.

As discussed earlier, ad hoc and common categories differ in how well they reflect correlational structure. Ross and Barsalou's (Note 1) category frequency results, together with Horton's (1982) developmental findings, demonstrate that how well a category reflects correlational structure does not affect fundamental ways the category is processed. Both ad hoc and common categories can behave as categories that are either poorly established or well established in memory. It does not appear that the differences between ad hoc and common categories in Experiments 2a, 3, and 4 regarding estab-

lishment of category representation stem from how well the two reflect correlational structure. Instead, these differences most likely result from common categories' having received much more processing prior to the experiments than ad hoc categories. Furthermore, Experiments 1 and 2b demonstrate that reflecting correlational structure is not necessary for a category to exhibit graded structure. In general, the same similarity comparison and memory processes appear to operate on categories that do and do not reflect correlational structure.

Determinants of Typicality

As discussed earlier for common categories, how similar an instance is to all other category instances (i.e., its family resemblance) determines its typicality. The comparison-network model explains this relationship by assuming that a category concept is the average of all category instances and that an instance's typicality increases as it becomes more similar to the category concept. This model goes on to explain typicality in ad hoc categories in a similar manner; namely, typicality depends on similarity to category concepts. Missing from this account, however, are assumptions regarding the content of category concepts for ad hoc categories. Are they, too, averages of their category instances? Barsalou (1981) provides evidence that they are not. In two studies, family resemblance bore no relation whatsoever to typicality for categories that violate correlational structure. Instead, these categories are structured by dimensions relevant to the goals the categories serve. For example, the dimensions of "calories" structures "things not to eat on a diet," with the typicality of an instance increasing as its number of calories increases. This dimension is important, presumably because it is relevant to the goal the category serves, namely, losing weight.

The comparison-network model accounts for this finding as follows. The category concept for an ad hoc category does not contain the average properties of its instances but, instead, only contains properties of the instances relevant to the goal the category serves. Since only "edible" and "high in calories" are relevant to losing weight, all other properties in "things not to eat on a diet" are not included in the category concept. During typicality judgments, the primary way instances can vary from this concept in similarity is simply with regard to calories. Instances having few calories are not as similar to the category concept as those high in calories and are therefore less typical. Family resemblance does not enter into typicality because the category concept does not contain averages across all properties. All properties except those relevant to the goal are effectively weighted to zero during the similarity comparisons. Although typicality derives from the same comparison process for common and ad hoc categories, fundamentally different category concepts result in fundamentally different forms of graded structure.

Constructing Category Concepts for Ad Hoc Categories

Perhaps one of the most difficult problems regarding ad hoc categories is explaining how goals make particular ad hoc categories relevant. In the previous section, I assumed that the concepts for these categories contain properties relevant to the goals these categories serve. But, given a goal, how can we predict which properties should be activated and associated to each other to form the concept for an ad hoc category? The solution to this problem will most likely need to be developed in the context of a theory of problem solving.

A more tractable problem is discovering how new concepts can be constructed to represent ad hoc categories. One way is to retrieve well established concepts and alter them by adding or deleting properties. In finding furniture for an office, "desk" and "lamp" may become "desk with a large surface" and "lamp that has adjustable positions and provides bright light." Adding properties was observed in Experiment 4 when subjects provided highly specific labels for common category item sets instead of using the one-word labels. Properties can also be deleted from concepts when contrary to current goals. A dieter might order a "salad without dressing," or someone who likes to work on furniture might buy an "unupholstered sofa."

Novel concepts can also be constructed from property information in memory. Any property, X, can be used to construct the category of "things possessing X." For example, the property of "flammable" could be used to represent the category of "things that are flammable" for someone trying to prevent fires. Such concepts may also contain sets of properties (e.g., things that are flammable and near a heat source). Sets of properties can be organized as conjunctions (e.g., expensive and unusual), as disjunctions (e.g., expensive or unusual), or as combinations of other forms.

Once a new concept has been constructed, it may be reconstructed on subsequent occasions if it continues to be relevant. Such reconstructions may cause the concept to become well established in memory. The associations between its properties (if there are more than one) may become stronger, and its overall accessibility may become higher.

Exemplar Production from Ad Hoc Categories

Once a concept has been constructed or retrieved, it can be used as a cue to retrieve category instances. For a category well established in memory, direct associations from its category concept to its category members are activated, thus activating the concepts for these members. But for categories not well established in memory, there are no well established concept-to-instance associations that serve this purpose. How, then, do people retrieve exemplars from ad hoc categories? One possibility is that they use a generate-test procedure. The associative structure of related, well established categories may be used to generate possible instances of a poorly established category. As each item

is retrieved, it is then tested for membership in the poorly established category. To find instances of "restaurants to watch a sunset in," instances from well established restaurant categories could be retrieved (e. g., local Indian restaurants, local seafood restaurants). Each instance would then be checked for properties such as "has a western exposure," "has large windows," and "has an unobstructed view." Instances having these properties become members of the category, and instances not having them become members of its complement.

As a category becomes frequently instantiated using a generate-test procedure, direct associations should become well established between the category concept and its instances. This change in category representation would eventually make the generate-test procedure unnecessary, since the more efficient lookup procedure could now operate.

Cross-Classification into Ad Hoc Categories

Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) argue that people prefer to classify entities initially with basic category names. For example, people prefer to call an inanimate object with four legs, a seat, and a back in a particular configuration a "chair" as opposed to an "office chair" or "furniture." Given this initial classification, however, there are numerous ways an entity can be cross-classified during subsequent classifications that serve particular goals. For example, a chair could subsequently be classified into "things that can be used for emergency firewood," "things that can be stood on," "gifts," and so on, depending on the current goal of the perceiver. An outline of a cross-classification model was sketched in the discussion of Experiment 4.

How well someone can generate secondary cross-classifications appears to vary substantially. In fact, the ability to do this well has often been considered a sign of creative ability. In the Unusual Uses Test (Cronbach, 1970), subjects are given the name of an entity and asked to generate as many uses of it as they can think of, each of which can be considered a cross-classification. The more cross-classifications a person generates, and the more novel these classifications are, the more creative the person is assessed as being.

In general, the construction and use of ad hoc categories appear to reflect creative aspects of human intelligence. Similar to the ability to cross-classify, the abilities to construct new concepts instrumental to achieving goals and to retrieve instances of these concepts without direct concept-to-instance associations are creative processes. All three enable the construction of new representations, each representation reflecting a new way of organizing the environment. Perceiving these new organizations may be necessary to achieving new goals or to approaching old ones in novel ways. Once an ad hoc taxonomy has been constructed, further use may cause it to become well established in memory. Understanding the construction, use, and establishment of ad hoc categories may not only turn out to be informative about human intelligence, understanding these processes may be central to it.

REFERENCE NOTE

1. Ross, B. H., & Barsalou, L. W. *Explorations into the nature of the category frequency effect.* Work in progress, University of Illinois, Champaign-Urbanna, and Emory University.

REFERENCES

Alba, J. W., Chromiak, W., Hasher, L., & Attig, M. S. Automatic encoding of category size information, *Journal of Experimental Psychology: Human Learning and Memory*, 1980, **6**, 370-378.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D. C: Winston, 1973.

Barsalou, L. W. *Determinants of graded structure in categories.* Unpublished doctoral dissertation, Stanford University, 1981.

Barsalou, L. W. Context-independent and context-dependent information in concepts. *Memory & Cognition*, 1982, **10**, 82-93.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80**, (Monograph Supplement 3, Pt. 2).

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229-240.

Bousfield, W. A., & Cohen, B. H. The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. *Journal of Psychology*, 1953, **36**, 67-81.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 323-343.

Clark, H. H. The language-as-fixed-effect-fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335-359.

Cofer, C. N. Does conceptual organization influence the amount retained in immediate free recall. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory.* New York: Wiley, 1967.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, **82**, 407-428.

Cronbach, L. J. *Essentials of psychological testing.* New York: Harper & Row, 1970.

Guilford, J. P., & Fruchter, B. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1973.

Hall, J. F. Learning as a function of word-frequency. *American Journal of Psychology*, 1954, **67**, 138-140.

Hayes-Roth, B. Evolution of cognitive structures and processes. *Psychological Review*, 1977, **84**, 260-278.

Horton, M. *Category familiarity and taxonomic organization in young children.* Unpublished doctoral dissertation, Stanford University, 1982.

Kučera, H., & Francis, W. N. *Computational analysis of present-day English.* Providence, R. I: Brown University Press, 1967.

Loftus, E. F. Spreading activation within semantic categories: Comments on Rosch's "Cognitive representations of semantic categories." *Journal of Experimental Psychology: General*, 1975, **104**, 234-240.

Lorch, R. F., Jr. Priming and search processes in semantic memory: A test of three models of spreading activation. *Journal of Verbal Learning and Verval Behavior*, 1982, **21**, 468-492

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press, 1967.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. Effects of organization and semantic similarity on recall and recognition. *Journal of Learning and Verbal Behavior*, 1969, **8**, 410-423.

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

McCloskey, M., & Glucksberg, S. Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, 1979, **11**, 1-37.

Mervis, C. B., Catlin, J., & Rosch, E. Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, 1976, **7**, 283-294.

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

Ortony, A. Beyond literal similarity. *Psychological Review*, 1979, **86**, 161-180.

Puff, C. R. Role of clustering in free recall. *Journal of Experimental Psychology*, 1970, **86**, 384-386.

Ratcliff, R., & McKoon, G. Does activation really spread? *Psychological Review*, 1981, **88**, 454-462.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 1-20.

Rosch, E. H. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

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

Rosch, E. H. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **1**, 303-322. (b)

Rosch, E. H. Reply to Loftus. *Journal of Experimental Psychology: General*, 1975, **104**, 241-243. (c)

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

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

Rosch, E. H., Simpson, C., & Miller, R. S. Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, **2**, 491-502.

Smith, E. E., & Medin, D. *Categories and concepts*. Cambridge, Mass: Harvard University Press, 1981.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, **81**, 214-241.

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, **69**, 344-354.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381-391.

Tversky, A. Features of similarity. *Psychological Review*, 1977, **84**, 327-352.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Zadeh, L. A. Fuzzy sets. *Information and Control*, 1965, **8**, 338-353.

NOTES

1. Let p, defined as the average proportion of items circled per item set, represent the probability of a subject's circling an item while guessing. Then if n is the number of subjects in the experiment, pn is the number of subjects circling a given item and $(1 - p)n$ is the number of subjects not circling the item. Agreement for the item, as defined in the text, is then$|(+1)pn+(-1)(1 - p)n|/n$, which is $2p - 1$. Averaging across items and item sets results in an overall agreement score for the experiment of $2p - 1$. Given that p in Experiment 1 was .52 (i.e., the average number of items circled per set, 3.15, divided by the number of items in a set, 6), then overall agreement should be .04 if subjects were guessing.

2. It should be noted that the unclear cases observed in these categories were unclear in the sense that some subjects thought these items were category members and other subjects did not. As noted by McCloskey and Glucksberg (1978), such cases are not necessarily unclear. They could falsely appear to be unclear because some subjects' definitions of a category include these items as category members, whereas other subjects' slightly different definitions do not. Most important, classifying these instances may be completely clear (as opposed to unclear) for all subjects. Given the presence of typicality gradients in these categories, however, it is likely that the unclear cases in this experiment really were items subjects were uncertain about. If subjects show decreasing confidence in the membership of clear category members, which they did, it is reasonable to believe their confidence would decrease to the point of being uncertain about the membership of unclear cases.

3. That subjects ably discriminate ad hoc categories is corroborated by the following demonstration. Six subjects generated four exemplars to each of the eight ad hoc categories used in Experiment 1. Four judges examined each of the 192 exemplars generated as to whether they were valid members of their respective categories. Three judges accepted all 192 exemplars as valid, and one judge accepted all but 4. Thus there appears to be substantial agreement on membership in ad hoc categories.

4. It was impossible to compute item averages for the random category intrusion data since it was not clear to which category a given intrusion belonged. Therefore, the one-way ANOVA for the intrusion frequencies was performed only across subjects.

5. Since number of correct exemplars retrieved and number of intrusions were both higher for common than for ad hoc categories, one could argue that common categories are not superior mnemonic devices. However, the total increase during free recall in the number of correct exemplars from ad hoc to common categories was 7.68, whereas the total increase in the number of intrusions was only 1.50. For cued recall, the increase in correct exemplars was 6.72, whereas the increase in intrusions was 2.06. The much larger increases for correct exemplars indicate that common categories do indeed make superior mnemonic devices.

6. The categories are "insects" and "things that have a smell," respectively.

7. The minimum increased as the number of labels for a set decreased. However, this worked against the hypothesis that there would be a context effect, since the no-context sets usually had fewer labels than the context sets.

(Manuscript received for publication March 26, 1982;
revision accepted February 7, 1983.)