When we hear a category word in natural language such as *furniture* or *bird* and understand its meaning, what sort of cognitive representation do we generate? A list of features necessary and sufficient for an item to belong to the category? A concrete image which represents the category? A list of category members? An ability to use the category term with no attendent mental representation at all? Or some other, less easily specified, form of representation?

The nature of cognitive representations of semantic categories has direct relevance to two important areas of psychological inquiry: One concerns the structure of categories and concepts and has implications for the way in which concepts and concept attainment should be studied in psychological research (Rosch, 1973, in press-a, in press-c). The other area is the nature of mental representations in general, a prominent concern in the postbehaviorist study of mental events (Cooper & Shepard, 1973: Posner, 1973; Segal, 1971). The present research was designed to investigate both issues within the same series of experiments. The first four experiments focused on the question of the nature and structure of categories and the effect of that structure on the tasks used. The succeeding five experiments used the effects of internal category structure found in the first experiments (reliable empirical effects which did not depend on any particular, possibly controversial, interpretation of internal structure) to explore several more general questions concerning the nature of mental representations of categories: questions regarding the type and degree of concreteness of the representations, and questions regarding the presence or absence of differentiation of the representations into pictorial and verbal forms.

With respect to the issue of the nature of categories, Rosch has previously argued (Rosch, 1973. 1975, in press-c) that many traditions (if thought in philosophy, psychology, linguistics, and anthropology imply that categories are Aristotelian in nature— that is. that categories are logical, clearly bounded entities, whose membership is defined by an item's possession of a simple set of criteria! features, in which all instances possessing the criterial attributes have a full and equal degree of membership. While such a structure may, in fact, characterize the artificial categories employed in much concept formation research (Bourne, 1968) and research on artificial languages (Ginsburg, 1966), many natural categories, that is, concepts designatable by words in natural languages, appear to possess structures of a quite different character (Labov, 1973; Lakoff, 1972; Rosch, 1973). The domain which has most readily lent itself to the demonstration of a type of categorical structure contradictory to the Aristotelian is that of color. There is now considerable evidence that color categories are processed by the human mind (learned, remembered, denoted, and evolved in languages) in terms of their internal structure; color categories appear to be represented in cognition not as a set of criterial features with clear-cut boundaries but rather in terms of a prototype (the clearest cases, best examples) of the category, surrounded by other colors of decreasing similarity to the prototype and of decreasing degree of membership (Berlin & Kay, 1969; Heider, 1971, 1972; Mervis, Catlin, & Rosch, 1975 ; Rosch, 1973, 1974, 1975, in press-c, in press-d).

Color, however, is a perceptual domain; its categorization is probably physiologically based (McDaniel, 1972; Rosch, 1974, in press-b, in press-d). Not all categories have an obvious perceptual basis and many categories may be culturally relative. Is the concept of internal structure applicable to other types of categories?

At this point, it is necessary to elaborate on the logic of the concept of internal structure as it is used in the present research. The hypothesis that categories have internal structure is not a theory which specifies, in advance of the collection of data, a precise model (with (low charts, steps, etc.) which is then tested by means of data from particular tasks. Rather, internal structure refers to that general *class* of conceptions of categories in which categories are not represented only as criterial features with clear-cut boundaries and in which items within categories may be considered differentially representative of the meaning of the category term. Progressive clarification of the meaning of internal structure is intended to occur through operational definitions based on converging experiments. Furthermore, initial experiments are intended to be in the form of "class questions" (Are categories of the general form x or y*)* rather than in the form of tests of specific stimulus or processing models. The research reported in this article (and in Rosch's other articles on the nature of human categorization which are referred to throughout this article) provides

illustrations of this strategy of conducting research. [v] This article is concerned with the question of whether the general concept of internal structure, previously specified only for perceptual domains such as color and form, is applicable to other types of categories; specifically, whether it is applicable to the semantic classifications of common objects in everyday use. Applicability refers to two general issues: Can subjects make meaningful judgments about internal structure—the degree to which instances are good or poor members of categories; and can a reasonable case be made that internal structure affects cognition with respect to categories? Some evidence that semantic categories fulfill both criteria for internal structure already exists. Rosch (1973) showed that subjects reliably rated the extent to which an item represented their idea or image of the category name for six items from each of eight categories. Such ratings were predictive of subjects' reaction times in a sentence verification task of the form "A (member) is a (category)" (Rosch, 1973). Rips, Shoben, and Smith (1973) confirmed and extended Rosch's (1973) findings, and Smith, Shoben, and Rips (1974) have argued that a model based on ratings and sentence verification tasks such as these can account for most of the findings in the literature on retrieval from semantic memory.

With regard to the nature of categories, the purpose of the present research was to extend the range of processing tasks shown to be affected by category structure and, terms, elaborate further the concepts of category structure and prototypes. It should be emphasized that the research was designed to investigate the nature (the structure and content) of cognitive representations of semantic categories, not the dynamic processes (activation, spread of energy, evaluation of paths, etc.) by which those representations are formed. That is, the studies were designed to focus on the what and not the how of cognitive representations of semantic categories. Specifically, the present research was not an attempt to construct a model of semantic memory; the fitting of dynamic memory models to the data of the present research is seen as a further stage of investigation. In Experiment 1, norms were collected for ratings of the goodness of example of a large number of instances from 10 systematically chosen classifications of common concrete objects. Succeeding experiments explored the manner in which those ratings of internal structure affected outcomes in tasks which were designed to explore, aspects of the structure and content of cognitive representations of categories.

With regard to more general questions about the nature of mental representations, these issues have been the subject of philosophical and psychological debate since the initiation of both fields of study. The present research was concerned with a specific type of category—the higher-level (super-ordinate) classifications into which the common concrete objects of our culture are divided—and with some of the classic questions concerning the nature of mental representations as these can be applied to such categories. Do our cognitive representations involve concrete imagery as Titchener (1909) claimed, or, as the Wurzburg school countered, is thought essentially abstract and imageless; do all meanings have a common abstract form regardless of their content, or are meanings represented in n variety of modes, for example, by different forms of representation for verbal materials and for pictures (Boring, 1950) ? The purpose of the present studies was to obtain data relevant to these interesting issues.

The questions that can be asked and the kinds of answers that can be obtained about the nature of mental representations of categories depend upon the particular research procedures by which representations are studied. The present studies are based on a method of priming (Beller, 1971) in the type of matching paradigm described by Posner and Mitchell (1967), which Rosch (in press-d) had already demonstrated to be particularly suitable to a study of the nature of the mental representation of color categories.

In the original matching research, subjects were required to decide as rapidly as possible whether two simultaneously presented visual letters were the same or different; under some conditions same was defined as physical identity (e.g., AA), and under others as possession of the same name (e.g., Aa). If a subject is provided with some of the information which he needs to make the match in advance of presentation of the pair, matching speed should be facilitated. Beller (1971) primed subjects with a letter (the letter was presented 2 sec in advance of the pair) and found that physically identical, as well as same-name, pairs improved—even when physically identical pairs were in the opposite case from the case presented in the prime (e.g., if the prime was A, the pair aa, as well as AA, was facilitated). Posner (1969) argued that this showed that subjects were not simply retaining a literal representation of the presented letter but were

generating an abstract expectation or representation which did not depend on case. It should be noted, however, that, as in early studies of set (Haber, 1966), this reasoning assumes that, because the prime is presented prior to the stimulus, it must have its influence on cognitive processes which occur prior to— rather than during or after—presentation of the stimulus.

The basic logic of the priming technique as it was used by Rosch (in press-d) and as it is used in the present experiments, is that a prime can only facilitate a response if it contains some of the information needed to make the response. If the prime facilitates a response when the prime is presented prior to the stimulus but not when presented simultaneously with the stimulus, it can be inferred that the information which facilitates the response is located in the representation generated by the prime in advance of information provided by the stimulus. If, under such conditions (i.e., a prime is effective when presented prior to, but not simultaneously with, a stimulus), a prime affects reaction time to physically identical pairs of stimuli, it may be concluded that the representation generated by the prime contains some of the information used by subjects in the perceptual encoding of the stimulus[1]. Since it is assumed that no facilitation occurs if the representation has not been allowed sufficient time to develop, it is possible to assess the amount of time taken to generate the representation by systematically reducing the length of time between the prime and the stimulus until the prime loses its effectiveness.

There is another possible outcome in which the effects of priming may include a different process. A prime may be effective when presented simultaneously with a stimulus as well as when presented prior to the stimulus. In such cases, it can be inferred that at least part of the relevant information in the prime acts on decision processes regarding the stimulus after the stimulus is available to the subject. However, because of possible interactions between stages of processing (see, e.g., Sternberg. 1969), it may be impossible to determine, from a priming experiment alone, which processing events are affected and in what manner by the prime.

A study of colors (Rosch, in press-d) illustrates how results found by priming can be of direct relevance both to the nature of color categories and to the nature of cognitive representations. Advance priming with the category name facilitated responses to good examples of color categories and hindered responses to poor examples for physically identical pairs of colors both under instructions which defined same as belonging to the same category and under those which denned same as physical identity." "The selective effect of the prime was entirely eliminated by simultaneous presentation of prime and stimuli. Following the previous logic, it was concluded that the cognitive representations of color category names contained information used in the perception of actual color stimuli. Furthermore, since the effect of the prime was selective, it could be concluded that the internal structure of color categories affected the representation generated by the category name; that is, the representation appeared to contain information used in the perception of good members of color categories but which was sufficiently unlike poor members of the category as to interfere with their perception. Representations of color categories were generated quite rapidly— selective effects of priming were already significant at 200 msec, which is less than half the time usually estimated to be required for the formation of a visual image (Posner, Boies, Eichelman, & Tayior, 1969). In sum, the effectiveness of the priming technique for an investigation of cognitive representations demonstrated by Rosch's (in press-d) studies of color categories made reasonable the present attempt to use a similar logic and similar procedures for an investigation of cognitive representations of semantic categories.

The nine experiments of the present research were designed to form a logical sequence. In Experiment 1, reliability of subjects' ratings of internal structure was verified, and norms were collected for ratings of the goodness of example of 50-60 members of 10 systematically chosen categories. In Experiment 2, items (represented by pictures in one condition and by words in another condition) of high, medium, and low goodness of example, according to the norms provided by Experiment 1, were used in a priming study with a 2-sec interval and same-

[1] [1] We may make such an assumption because the only operations which a subject can perform upon a physically identical pair are (a) the perception (visual encoding) of the stimuli and (b) the perception that the members of the pair are identical. Even if the perception of physical identity is a decision separate from perceptual encoding of the stimulus itself, a prime consisting of a category name does not contain within it information about the nature of identity as such (an especially untenable assumption if, as was found to be the case, the prime differentially affects different types of stimuli). Thus, the only possibility, under these conditions, is that the prime affects the visual encoding.

category instructions. The conditions of Experiment 2 were designed to determine whether, under maximally favorable conditions, any effect of internal category structure on the representation of the category could be observed. Experiment 3 was performed to check whether effects obtained in Experiment 2 were the artifactual result of the fact that subjects did not consider the poor examples of the categories members of those categories at all. Because an interaction between goodness of example and priming was achieved in Experiment 2 which did not appear to be an artifact. Experiment 4 was performed in which essential aspects of Experiment 2 were replicated under conditions in which the prime and stimulus were presented simultaneously. Following the logic previously outlined, the results of Experiments 2 and 4 indicated that the effect of internal structure on same category and different responses lay partially in the decision processes after the stimulus pair was presented and that, in those conditions, effects of the prime on encoding and on subsequent decisions could not be unequivocally disentangled. However, the effects of the prime on physically identical pairs in these experiments showed a selective effect of the representation generated by the category name prior to presentation of the stimulus—an effect very similar to that which had been found for colors by Rosch (in press-d).

Subsequent experiments focused on obtaining further information about the nature of the representation as it appeared to. If affect physically identical pairs when the prime was presented prior to the stimulus. In Experiment 5, physical identity instructions were employed to test the concrete-ness of the representation; the disappearance of the effects of advance priming under such instructions indicated that different levels of encoding of the stimuli were being used. This hypothesis was further verified by a recall task in Experiment 6. Experiment 7 approached the question of whether the meaning which appeared to be part of the representation generated by the category name was differentiated into pictorial and verbal form. In this experiment, both pairs of words and pairs of pictures were used as stimuli for the same subject, presented in a format such that subjects were uncertain about the form of the stimulus pair with which they would be presented on each trial. The results for Experiment 7 did not indicate a differentiation of the code for words and pictures. However, since the length of the 2-sec prime interval /nay have masked subtle degrees of such differentiation, Experiment 8 was performed to chart the time course of generation of word and picture representations both under conditions in which the form of the ensuing stimulus was certain and uncertain. This experiment produced some evidence of differentiation of representations required in processing verbal and pictorial form, and some evidence that representations for pictures were generated more rapidly than those for words. The final experiment was designed to study the effect of long practice and thorough memorization on a small set of good and poor examples of categories.

EXPERIMENT 1

Before attempting to examine the effect of internal structure on mental representations of semantic categories, it was necessary to establish the reliability of judgments of internal structure for these categories. That is, it was necessary to gather reliable normative data on subjects' ratings of the extent to which instances of semantic categories represent their idea or image of the meaning of the category name. In addition, if the nature and extent of concreteness of representations were to be explored, it was necessary to use concrete, picturable categories in the studies. Previous subjects' ratings of internal structure (Rosch, 1973) were performed on only six items per category which had been prechosen to represent an expected wide range in goodness of example. Furthermore, the categories used in that study had not been systematically chosen and included abstracts as well as concrete categories. Rips, Shoben, and Smith (1973) used 12 instances of three categories for their basic material. The purpose of the present experiment was to determine reliability of ratings and gather normative data for a systematically sampled set of concrete noun categories in common use in English. If reliable, these norms were intended for use in the subsequent experiments on the nature of mental representations.

Method

Stimuli

The categories for which ratings of instances were to be gathered were chosen in the following manner: The population of categories of concrete nouns in common use in English was determined by drawing all concrete nouns with a word frequency of 10 or greater from the Kucera and Francis (1967) sample of written English. A category was considered in common use if at least five items from the category appeared in that list. A category was considered concrete only if the items in it could be unequivocally represented by pictures—thus, common categories such as *relative* or *number* were not considered concrete. Categories were eliminated if (a) the items bore a part-whole relationship to the only reasonable superordinate (e.g., parts of the body, parts

of buildings), (b) if there was linguistic ambiguity among possible superordinates (e.g., *animal* is commonly used as a synonym for *mammal),* and (c) if the superordinate crosscut a large number of other taxonomic structures (e.g., *food).* Surprisingly, only 17 concrete categories met the initial frequency requirement, 7 of which were eliminated by the other criteria. The remaining 10 categories which were used in the normative study were: fruit, bird, vehicle, vegetable, sport, tool, toy, furniture, weapon, and clothing.

All of the categories which met the criteria were categories which had been included in the Battig and Montague (1969) normative tabulations of the frequencies with which instances were produced in response to the category name (e.g., all of the *tools* appear in the Battig & Montague norms under carpenter's tool). Thus, the Battig and Montague lists could be used as a basis for selecting those members of the categories which were to be rated in the present experiment. For each category, all of the items which had been produced by 10 or more subjects in the Battig and Montague study were included. Additional items to make a total of 50-60 instances were taken randomly from the items produced by fewer than 10 subjects in the Battig and Montague norms. Items that were jokes or obvious mishearings of the category name were excluded.

Subjects

Subjects were students in three psychology classes who filled out the rating forms as part of their class work. On a front sheet, subjects listed their country of birth and the country(ies) or state (s) in which they had lived since birth. Subjects who were not native speakers of English or who did not complete the forms were eliminated from the study. Analysis was based on the forms of the 209 remaining subjects.

Procedure

All members of a category were listed below the category name. Subjects were asked to rate, on a 7-point scale, the extent to which each instance represented their idea or image of the meaning of the category term. Specific instructions were:

This study has to do with what we have in mind when we use words which refer to categories. Let's take the word red as an example. Close your eyes and imagine a true red. Now imagine an orangish red . . . imagine a purple red. Although you might still name the orange red or the purple red with the term red, they are not as good examples of red (as clear cases of what red refers to) as the clear "true" red. In short, some reds are redder than others. The same is true for other kinds of categories. Think of dogs. You all have some notion of what a "real dog," a "doggy dog" is. To me a retriever or a German shepherd is a very doggy dog while a Pekinese is a less doggy dog. Notice that this kind of judgment has nothing to do with how well you like the thing; you can like a purple red better than a true red but still recognize that the color you like is not a true red. You may prefer to own a Pekinese without thinking that it is the breed that best represents what people mean by dogginess.

On this form you are asked to judge how good an example of a category various instances of the category are. At the top of the page is the name of a category. Under it are the names of some members of the category. After each member is a blank. You are to rate how good an example of the category each member is on a 7-point scale. A 1 means that you feel the member is a very good example of your idea of what the category is. A 7 means you feel the member fits very poorly with your idea or image of the category (or is not a member at all). A 4 means you feel the member fits moderately well. For example, one of the members of the category fruit is apple. If apple fit well your idea or image of fruit, you would put a 1 after it; if apple fit your idea of fruit very poorly you would put a 7 after it; a 4 would indicate moderate fit. Use the other numbers of the 7-point scale to indicate intermediate judgments.

Don't worry about why you feel that something is or isn't a good example of the category. And don't worry about whether it's just you or people in general who feel that way. Just mark it the way you see it.

Approximately half (116) of the subjects received one random order of instances under the category name; the other half a different random order. Categories were typed on separate pages and each subject received a different order of categories (taken randomly from the 10! possible orders of 10 categories).

Results and Discussion

Rank orders and mean ratings of goodness of example for all instances of all categories are shown in Table Al. To test the reliability of ratings both Spearman rank-order correlations and Pearson product-moment correlations of the mean ratings were obtained (a) between split halves of the sample of subjects divided at random, (b) between subjects who rated the two different item orders, and (c) between subjects who had lived predominantly on the west (n = 131) versus east (n = 78) coasts of the United States. Consistency was extremely high. All split-half correlations were .97 or higher, and all correlations between the west and east coast samples were .92 or higher. Agreement between subjects was particularly high for the items rated as very good examples of the category ; for example, for 9 of the 10 categories, 95% of the subjects agreed in giving the item with the mean best example rating the same score, that of 1. In addition, 6 categories occurred in both the present study and in Rosch (1973), making it possible to compare the rank order of the ratings which items received when they were the only 6 instances of the category present and when they were embedded in almost the total Battig and Montague (1969) list; for all 6 categories, all rank orders were identical.

The results of this study clearly indicate that semantic categories do have internal structure: (a) Subjects consider it a meaningful task to rate members of such categories according to how

well they fit the subjects' idea or image of the meaning of the category name and (b) there is high agreement between subjects concerning these rankings.