Cognitive Psychology **38,** 495–553 (1999)
Article ID cogp.1998.0712, available online at http://www.idealibrary.com on IDEAL®

Food for Thought: Cross-Classification and Category Organization in a Complex Real-World Domain

Brian H. Ross and Gregory L. Murphy

Department of Psychology, Beckman Institute, University of Illinois

Seven studies examined how people represent, access, and make inferences about a rich real-world category domain, foods. The representation of the category was assessed by category generation, category ratings, and item sortings. The first results indicated that the high-level category of foods was organized simultaneously by taxonomic categories for the kind of food (e.g., vegetables, meats) and script categories for the situations in which foods are eaten (e.g., breakfast foods, snacks). Sortings were dominated by the taxonomic categories, but the script categories also had an influence. The access of the categories was examined both by a similarity rating task, with and without the category labels, and by a speeded priming experiment. In both studies, the script categories showed less access than the taxonomic categories, but more than novel ad hoc categories, suggesting some intermediate level of access. Two studies on induction found that both types of categories could be used to make a wide range of inferences about food properties, but that they were differentially useful for different kinds of inferences. The results give a detailed picture of the use of cross-classification in a complex domain, demonstrating that multiple categories and ways of categorizing can be used in a single domain at one time.

INTRODUCTION

People know a lot about food. We eat food, smell it, plan meals, read about it, talk about it, see it advertised, etc. How is this rich set of knowledge represented and used? This paper provides a preliminary examination of this

This work was partially supported by Grant MH41704 from NIMH and Grant SBR 97–20304 from NSF. Research for this paper was conducted at the Beckman Institute for Advanced Science and Technology. We especially thank Lawrence Hubert for his very generous help with analyzing and examining the sorting data. We also thank Lawrence Barsalou, Barbara Malt, Douglas Medin, and Thomas Spalding for comments on an earlier draft of the manuscript and our resident nutritionist, Cheryl Sullivan, R.D., for advice. Amanda Lorenz, Amanda Schulze, Hayley Davison, and Amy Anderson provided excellent help in conducting and scoring these experiments.

Correspondence and reprint requests may be addressed to Brian H. Ross, Beckman Institute, University of Illinois, 405 N. Mathews Ave., Urbana, IL 61801 or via email to bross@s.psych.uiuc.edu.

0010-0285/99 $30.00

question, focusing on issues and methods that will address how real-world concepts and categories are represented and used.

Most research on categories has focused on experimenter-defined categories in order to test models of classification (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1988; for reviews see Medin & Smith, 1984; Ross & Spalding, 1994; Smith & Medin, 1981). However, there has also been work on real-world concepts, such as animals and plants, examining how people represent the categories that they have learned through experience. These categories often have much richer correlational structures and longer learning histories than can be captured in the laboratory (see e.g., Brooks, Norman, & Allen, 1991; Johnson & Mervis, 1997, 1998; Lopez, Atran, Coley, Medin, & Smith, 1997; Malt, 1994; Malt & Johnson, 1992; Medin, Lynch, Coley, & Atran, 1997; Rips, 1989; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tanaka & Taylor, 1991).

This work has increased our understanding of conceptual representation, but has often suffered from three limitations: a single hierarchy, a single function, and isolated knowledge. First, many items belong to multiple hierarchies, but much of the literature on concepts examines only a single hierarchy for each domain (e.g., Rosch et al., 1976). For example, they might consider the category *mammals,* some types of mammals (e.g., *dogs*), and some more specific subcategories of each type (e.g., *spaniels, collies*). Examinations of single hierarchies have led to a number of interesting findings, but ignore the many cases in which we have alternative organizations, which are often called *cross-classifications.* Cross-classifications have been recognized and mentioned occasionally in the concept literature (e.g., Barsalou, 1982; Murphy, 1993) but have not been extensively investigated. An examination of cross-classification is important in part because it is a widespread phenomenon in many domains (e.g., person categories—such as a person who is a Democrat, fiscal conservative, feminist, and golfer). There is little understanding of how cross-classifications are represented and used.

Second, most studies of real-world concepts have focused on a single function, classification. As we argue in the next section, classification is an important aspect of conceptual representation, but only one of the functions for which concepts are used. A full picture of concepts and their uses requires considering other functions as well.

Third, most studies of real-world concepts have focused on knowledge that is isolated from much of our other knowledge of the world and of ourselves. Many projects have examined our representations of animals and plants. Although a few animals and plants may be seen every day, for most urban (and even suburban) dwellers they simply are not central to our thoughts and activities. In addition, much of our knowledge of such categories comes from observation and communication from others, not from extensive, important interactions. These biological categories are excellent for investigating some aspects of real-world conceptual representations because they do have clear taxonomic hierarchies that are shared by many people, but they are not, for most people, highly integrated with other knowledge or activities.

Goals of the Current Investigation

The current work examines the representation and use of a real-world concept, food, to overcome the three limitations just mentioned. We discuss each of these properties for food categories: cross-classification, multiple functions, and integrated knowledge.

Cross-classification. The primary goal of this project is to examine cross-classification in a complex real-world domain. As mentioned, it is common for an item to belong to multiple categories that represent alternative conceptual organizations. For example, one can classify people by their age groups, political party affiliation, and country in which they were born. We know very little about how such alternative organizations are represented and how they are used for various conceptual functions. Food is an excellent domain for examining these issues. There is a rich set of ways of cross-classifying many foods (e.g., *bagel* is not just a bread, but may also be considered a sandwich food, a breakfast food, a Jewish food, a snack food, etc.).

In many domains, such as foods, different conceptual organizations may be quite different from one another. What are the different conceptual organizations in a domain such as foods? How are such cross-classifications represented? Does the organization that is used in usual situations reflect just the dominant organization or is it some combination of the different conceptual organizations? These are basic questions about conceptual knowledge for which we do not have answers.

Cross-classifications bring a host of additional issues concerning how conceptual knowledge affects people's understanding of the environment and their actions in the world. What functions might different organizations serve? Are all the categories a food belongs to accessed when the food is experienced or named? For inferences, do people use one of the categories or more than one? If more than one, how are they combined to make an inference, and if just one category, how is it chosen? Again, these are basic questions that have not yet been answered.

Some previous work points out the importance of cross-classifications and provides a beginning for this research. Barsalou's (1983, 1985, 1991) well-known work on goal-derived categories suggests that people can form alternative organizations in response to some goal, such as ''things to take out of your house in case of a fire.'' Medin et al. (1997) found that landscapers' sortings were influenced by the landscaping utility of trees (e.g., shade trees, ornamental trees). Based on other findings (described below), it was clear that landscapers had a more standard taxonomic representation as well. In addition, there has been a variety of work in social cognition that addresses cross-classifications, because of their pervasiveness in person categories (e.g., Nelson & Miller, 1995; Smith, Fazio, & Cejka, 1996; Zarate & Smith, 1990). This research has provided a number of interesting ideas concerning cross-classifications and their selection and use in inference and will be discussed further in the General Discussion. Thus, some past work suggests the importance of cross-classifications in a variety of situations, but many of the basic questions outlined earlier remain unanswered.

Multiple functions. Although we have learned much about category representation from the study of classification, it is also important to examine other functions that categories may serve, such as induction, explanation, problem solving, category formation, and communication. These other functions provide alternative windows on category representation and are of considerable interest in their own rights. In many cases, classification provides access to categorical knowledge, but that knowledge may then be used in a variety of different ways. For example, problem solvers often classify problems because it allows them to access information (e.g., formulae) about how to solve problems of that type.

Investigations of real-world concepts, like laboratory studies, have often examined the representation of the category structure as revealed by the classification function of the concepts. Interesting work has examined what types of properties are important for determining category membership (e.g., Malt, 1994; Malt & Johnson, 1992; Rips, 1989; Smith & Sloman, 1994), the influence of correlated properties (Malt & Smith, 1984), and classification at different levels of abstraction (Lassaline, Wisniewski, & Medin, 1992; Murphy & Brownell, 1985; Rosch et al., 1976; Tanaka & Taylor, 1991; see Murphy & Lassaline, 1997, for a review). The examination of classification has included categories other than object categories, such as diseases (Brooks et al., 1991) and problem categories (e.g., Chi, Feltovich, & Glaser, 1981; Schoenfeld & Herrmann, 1982). These studies have focused on how new instances are classified, rather than examining how these classifications might be used.

A number of studies have examined nonclassification uses of categories. Perhaps the best known of this work investigates how we use categories to make inductions to new instances or other categories (e.g., Gelman & Markman, 1986; Heit & Rubenstein, 1994; Kalish & Gelman, 1992; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990). In these studies, the classification is often given and the question is how people make an inference on the basis of the category. Some recent work on category-based inductions examines the inferences made when the classification is uncertain (Malt, Ross, & Murphy, 1995; Murphy & Ross, 1994; Ross & Murphy, 1996). In these studies, category membership has been found to be critical to inductions about an item.

Different purposes and tasks may lead to different ways of processing the category representation so that a more complete understanding of the representation may require the use of multiple tasks. This need for multiple examinations of the representation may be especially true for investigations of complex real-world concepts. For example, Blessing and Ross (1996) investigated how experienced problem solvers are affected by the content of a problem (i.e., whether the content is typical for problems of this type). They found different patterns of the effect of content depending upon whether they examined classification or problem solving. That is, the classification results suggested that problems with neutral contents were classified as well as problems with typical contents, but the problem solving results showed clear differences in how content led to knowledge about the problem category and its use. Thus, using multiple tasks provided a more complete picture of how content affected the performance of experienced problem solvers. Barrett and Keil (1996) found that people have two parallel concepts of God that may be evoked in different situations—a theological one that they use to answer many general questions and an anthropomorphic representation that may be more important for some online tasks such as story understanding. Most directly related to the current investigation is the work of Medin et al. (1997). They examined the category representations and inductions of three different types of tree experts: botanists, park maintenance workers, and landscapers. The results showed that the sortings of the first two groups were similar to those of the scientific taxonomy, but, as mentioned earlier, the landscapers' sortings were influenced greatly by the utility of the different trees in landscaping. However, when asked to make inductions about biological properties from one tree category to another, the landscapers' judgments did not appear to be a function of this utilitarian representation but rather closely followed that predicted by the scientific taxonomy (see also Coley, Medin, & Atran, 1997). This finding points out that people may be quite flexible in how they use their representations for different purposes (e.g., Lopez et al., 1997). An examination of people's representations and uses of categories may need multiple measures to get a full picture.

Recent research also suggests that how we use categories can affect the representations of these categories for both classifications and inductions (Barsalou, 1983, 1991; Lopez et al., 1997; Markman, Yamauchi, & Makin, 1997; Medin et al., 1997; Ross, 1996a,b, 1997, in press-a, b; Yamauchi & Markman, 1998). Although most laboratory studies examine classification, many categories are not learned principally for classification. For example, Barsalou (1983, 1985) has shown that people readily form new categories that address specific goals. In everyday life, such categories would be used primarily as part of a planning process rather than for categorization (Barsalou, 1991). Purely goal-derived categories do not have strong correlational or family-resemblance structures. Instead, knowledge about an item can be processed in various ways so that the item's appropriateness to fulfilling a goal can be assessed. Clearly, we do not learn types of food primarily to classify—the classification is in the service of nutritional, hedonic, and social goals. Food categories are an interesting case because they clearly have both correlational structure and are used in a variety of goals (see General Discussion for elaboration).

In the current experiments, we examine the category representations of foods, but then we make use of these proposed representations to examine how category information is accessed and used in induction. Thus, the understanding of how people represent the categories is tied to how they access these categories and make inductive inferences from them. Such an examination is particularly important for items that are readily cross-classified (as foods will be shown to be), because the presentation of an item may access multiple categories, and it is not clear how the categories accessed will influence the inductions. For example, does the presentation of *bagel* lead people to access knowledge of both breads and breakfast foods? If so, how might each be used in making an induction about some new property? The main point is that as we consider additional functions of categories, a number of new issues arise. This situation is both more difficult and more illuminating for a full understanding of cross-classifications and their use.

Integrated knowledge. A third goal of this project is to examine a real-world concept that is well integrated with human knowledge and activities. The earlier work with real-world concepts has focused on relatively isolated biological categories, and the experimental work often uses artificial categories, which are not at all related to other knowledge or activities. Food is something that is used every day and is an integral part of human life. Our knowledge of food is extensive and is accessed many times per day (e.g., Rozin, Dow, Moscovitch, & Rajaram, 1998). It is not some isolated body of knowledge but part of many aspects of our physical and social life. We know which foods to eat for energy and which may upset our stomachs. We know the foods that are likely to be served at various holidays and social events. We know which foods we can afford to buy and how long it will take to prepare them. Our knowledge of food is connected to much of our other knowledge.

In addition, knowledge of food is learned and used in an incredibly large number of ways and contexts. Our knowledge of many biological categories (trees, nonhuman animals) often comes largely from observation and communication from others. Foods, however, are interacted with extensively and in many ways. Besides eating foods, many of us plan meals, cook, and shop for foods. Newspapers and magazines are filled with articles on cooking and on the health implications of different foods. The representation of foods is bound to be affected by the large number of interactions we have with them, as well as the wide diversity of these interactions.

There are studies that examine concepts that the subjects are likely to have well integrated to some of their prior knowledge. For example, Medin et al.'s (1997) tree experts presumably have extensive interactions with trees for many hours every day. The Lopez et al. (1997) experiments used animals that were an important part of life for one of the groups, Itzaj-Mayan Amerindians. The social cognition work examines person categories that are clearly very integrated with world knowledge we have. The present study will go beyond those studies by examining a rich domain from the perspective of multiple functions. It will document the cross-classification of foods, investigate the accessibility of these different categories, and examine the use of such categories in induction.

Current Experiments

The goal of the current studies is to investigate the phenomenon of cross-classification in a complex real-world domain, foods, and to attempt to explain how the cross-classification is represented and used in a variety of tasks. We ask two specific questions. First, what categories do people use for thinking about foods? It is likely that people employ taxonomic categories that capture the compositional similarities of foods (e.g., fruits, breads) but might there be additional organizations of their food concepts? People certainly have intuitions that foods may be cross-classified, but the current work goes beyond these intuitions to address more specific issues about the cross-classifications. In particular, assuming there is an alternative organization (or even organizations), what is it, how common is it, and how consistent is it across people? In addition, how is such an organization related to the interactions people have with foods? The answers to these questions go beyond simply demonstrating cross-classification.

Second, if people have different types of categories about food, what roles do these different categories play beyond classification? In particular, we will examine what categories might be brought to mind when encountering the different foods and how different categories might be used in inductive inferences. The accessibility of categories is a crucial question when considering cross-classified items. As will be discussed before Experiment 4, there is evidence about the accessibility of taxonomic categories for objects, but little is known about categories from alternative organizations. Given that the categories activated by an item will influence the comprehension of the current situation, as well as inferences and predictions, this is an important issue. The use of cross-classified items for inductive inferences has the same central import for our understanding of category representation and use. There are many possible ways that inductive inferences might be made when multiple categories are activated (discussed more in the introduction to Experiment 6), but the main issue is whether the taxonomic organization controls almost all the inferences or whether the alternative organization might influence some inferences. If the latter, what determines which inferences are influenced and how is the influence from the different organizations combined? Again, to the extent that cross-classification is common, these are basic questions that need to be addressed.

CATEGORY REPRESENTATIONS

The first set of studies examines the organization of food categories. To this end, we used techniques of category generation, ratings, and sortings. However, it is possible that the categories discovered in these studies do not accurately reflect the categories that are actually used in thinking and making judgments about food but are artifacts of the generation and comparison processes involved. Thus, the second set of studies examines whether the categories discovered are actually activated when people process the food items. The third set of studies tests whether these categories are used in inductive reasoning about properties of foods. Thus, the first experiments on category representation will provide the foundation for the later studies. Although these first experiments are largely descriptive, they provide a much clearer picture of the cross-classification of a real-world domain than has past research.

Experiment 1: Category Generation

To begin an analysis of the representations of food categories, we first examined what kinds of categories people have about foods. Thus, in this first study, we gave people a list of basic food types, like apple and chicken, and asked them to generate some categories for each of the foods. Given the very large number of foods, it is not possible to include more than a fraction of them. We tried to select foods that we thought would be generally representative for a college-aged American subject population. Although one can question any selection, it is important to note that we did not approach this work with any strong preconceived ideas about the kinds of categories that people have about foods. Rather, the issues of cross-classification and the examination of different types of categories described in the Introduction arose largely from the results of Experiment 1. To select the foods, we chose examples from familiar food types: beverages, breads, dairy foods, fruits, grains, meats, and vegetables. We selected several examples of each kind so as to ensure diversity. We also attempted (with help from undergraduate informants) to choose examples that were eaten at different meals and as snacks, especially by college students. We generally avoided combined food dishes as items (e.g., beef stew), with the possible exception of pizza, which is an extremely common food choice for our subject population and which is probably thought of as a single item rather than a composite. The full list of 45 foods is given in Table 1. This same set of foods was used for all

TABLE 1
List of Foods Used in Experiment 1
for Category Generation

Carrots	Spaghetti
Lettuce	Bread
Corn	Muffin
Potato	Rice
Onions	Pizza
Broccoli	Cereal
	Pancakes
Apple	Crackers
Orange	Bagel
Pineapple	Oatmeal
Banana	
Watermelon	Cake
	Pie
Chicken	Cookies
Hamburger	Doughnuts
Salmon	Ice cream
Lobster	
Steak	Potato chips
Pork	Pretzels
	Nuts
Milk	Chocolate bar
Eggs	Popcorn
Yogurt	Granola bar
Butter	
Cheese	
Soda	
Water	

the studies on category representation (Experiments 1–3), though additional items were required for later studies.

Method

Subjects. The subjects were 13 undergraduates who received course credit or pay. The experiment took about 30 min.

Materials. As mentioned, 45 foods were chosen that spanned a variety of categories and that we believed would be familiar to undergraduate students. They were randomized and typed 5 to a page. The booklets consisted of a cover page with the instructions and the nine randomized pages of foods.

Procedure. The instructions informed subjects that the goal of the study was to find out how people think about categories of foods. They would be given a number of food terms. For each term, they were to think about the food for a while and then write down what categories they think of that food as belonging to. An example was given of a dog belonging to a large number of different categories (pet, canine, animal, domestic animal, mammal).

Subjects were allowed 30 s for each food term and were asked to write down as many categories as they could think of. If they were not sure a particular category was appropriate, they were asked to write it down anyway. At the end of 30 s, the experimenter said ‘‘Next,’’ and they were to finish up what they were writing and proceed to the next term.

Design. All subjects generated categories to the same food terms, though the pages were randomly ordered for each subject.

Results and Discussion

The number of categories written down varied, but most responses had between 2 and 5 categories. For each food term, we tabulated the number of times each category was given. Because the goal of this study was to get an idea of the kinds of categories that people commonly use, we focused on the 5 most frequently generated categories for each food term (all categories tying with the 5th category were also included). This list included 1403 responses, covering 312 different categories (counting each category separately for each food term—thus, if fruit was generated for two different food terms it would count as two food categories).

Perhaps because subjects were encouraged to write down any response they were thinking of, a number of the responses were not categories. For example, there were properties of the foods (e.g., salty, green), associated items (e.g., cheese for the food term crackers), and subcategories (e.g., marble cake for cakes). When these responses were eliminated, 826 responses remained, which we divided into three main kinds of categories. First, there were the superordinate level taxonomic categories, which were largely the ones used in generating the list: beverages, breads, dairy foods, fruits, grains, meats, and vegetables. Of the 826 responses, 403 (.49) were of these food types, which we will call *taxonomic* categories. Second, subjects listed ‘‘proteins’’ and ‘‘carbohydrates,’’ which provide an alternative organization of foods by their macronutrients, which comprised 73 of the responses (.09). Third, there were categories that did not group together foods of the same constitutive kinds, but instead referred to the situation in which the food was eaten, such as breakfast foods (bagel, eggs, banana, yogurt) or snacks (popcorn, yogurt, muffin, apple), or referred to the healthiness of the foods, such as healthy foods (orange, chicken, granola bar) or junk foods (pretzels, potato chips, ice cream). Thus, these categories often included items from very different taxonomic categories. They constituted 350 of the responses (.42). By grouping together the situational and healthiness categories, we are not arguing that they are all necessarily the same kinds of categories, but for initial purposes we will consider them as a group that is different from the taxonomic categories. We will call them *script* categories, because they usually indicate a time or situation in which the food is eaten.

Thus, these results suggest that people may have alternative organizations or cross-classifications of foods. The script categories are particularly interesting for three reasons. First, they demonstrate the existence of categories based on interactions with the food rather than on its composition. In contrast to the script categories, the taxonomic categories are much more like the traditional similarity-based categories, in that they share multiple properties that are not as dependent on human interaction (see Lin, 1996). The taxonomic categories, with the exception of beverages, represent very different macronutrient profiles (i.e., proportions of carbohydrates, proteins, and fats). In fact, these are approximately the categories used in cases in which biochemical properties of food need to be carefully regulated, such as diabetic exchanges (as discussed in many nutrition texts, e.g., Wardlaw & Insel, 1990). Second, as will be discussed later, the script organization of foods may be especially helpful in generating plans for deciding about what foods to eat, a crucial function for food categories. Third, these script categories were generated almost as frequently as the taxonomic categories, suggesting that they are a fairly salient way of thinking about foods (also see Nelson, 1996).

The remaining studies in this article examine more closely these different ways of conceiving of foods. Of particular interest is to better understand the different representations of foods and how category information may be activated and used in making inferences.

Experiment 2: Category Ratings

The category generation task suggests that people have both taxonomic and script categories for foods. However, generation tasks are often suspect, because they may create an implied demand to produce a number of responses, and after generating the taxonomic categories, subjects may generate answers that they do not really believe are categories (see Hampton, 1979, or Tversky & Hemenway, 1984, for discussion of ratings vs. production measures). For example, as mentioned, a number of the answers were associates of the food, such as cheese for crackers. In this study, we provided both the item and the category and asked subjects to rate how good an instance of the category the item is. Such ratings, made without any time pressure, are more likely to indicate whether the categories generated are viewed as true superordinates of the food items. Thus, two questions motivated this study. First, are foods rated as belonging to script categories, as suggested by the generation data? Second, how do the ratings of the script categories compare to those of the taxonomic categories? In particular, are script categories thought to be just as good superordinates of the foods as more traditional taxonomic categories?

Method

Subjects. The subjects were 10 undergraduates who received course credit or pay. The experiment took about 45 min.

Materials. Sixteen categories that were common responses from the category generation

TABLE 2
Categories Used in Experiment 2 Ratings

Taxonomic	Script	Macronutrients
Beverages	Appetizers	Proteins
Breads and grains	Breakfast foods	Carbohydrates
Dairy foods	Desserts	
Fruits	Dinner foods	
Meats	Healthy foods	
Vegetables	Junk foods	
	Lunch foods	
	Snack foods	

task were chosen (though we combined breads and grains into 1 category). The categories, which are shown in Table 2, include 6 taxonomic categories and 8 script categories.[1]

Each page had one category and all 45 instances from Experiment 1 (see Table 1). At the top of a page was a scale (with three labels), ranging from 0, which was labeled ‘‘NOT a Member,’’ to 3, ‘‘Fairly Good Member,’’ to 7, ‘‘Excellent (Very Typical) Member.’’ Below the scale (5 cm from the top of the page) was one of the categories printed in boldface and underlined. Below the category were the 45 instances, randomly ordered in three columns of 15 each. Two random orders were used, but for a given subject, all the pages had the same random order of the 45 food names. The booklets consisted of a cover page, an instructions page, and then 16 pages of categories, randomly ordered.

Procedure. The instructions informed subjects that the study was to find out what people think about types of foods. They would be given a number of food categories, and their task was ‘‘to rate each of the foods on the page in terms of how good an instance of the category it is,’’ using the 0 to 7 scale at the top of each page. They were given an illustration with a nonfood category, vehicle, rating flagpole as 0, car as 7, and skateboard as 2 or 3.

Design. All subjects made the same ratings, though the pages were randomly ordered for each subject, and half the subjects received each random ordering of the foods on the page.

Results and Discussion

Two results are of main interest. First, do subjects rate foods as belonging to the script categories? Second, how do these ratings compare to those of the taxonomic categories? To answer these questions, we averaged the ratings for each food in each category and arbitrarily set 4.0, on a 7-point scale, as the boundary for being included as a good member of the category (though we describe later results with a stricter boundary).

[1] Although we also included the two macronutrient categories, proteins and carbohydrates, we did not use these categories in subsequent research, and we will not present the data from these categories. We initially considered these categories to be a kind of food type. However, further consideration suggested that they only characterize one component of a food—whether it includes a certain amount of this one ingredient. Many foods contain both protein and carbohydrates. In that sense, then, protein and carbohydrate are more like food properties, which were discarded after the feature-listing stage. Although such properties may be of interest, they are not like either our script categories or the food types, which include other information, such as origin, color, typical means of cooking, etc.

TABLE 3
Number of Foods (of 45) Given a Mean Rating of 4.0 or Greater in Experiment 2

	Number of categories				
Type of category	0	1	2	3	4
Taxonomic	10	34	1	0	0
Script	2	9	16	15	3

Note. The entries are the number of items that reached criterion in the indicated number of taxonomic or script categories. For example, 16 of the 45 items had a rating of 4.0 or greater for two script categories.

First, it is clear that people do view the foods as belonging to the script categories, as can be seen in Table 3. For the script categories, 43 of the 45 foods were rated as belonging to at least one of the script categories. (The 2 foods that were not were butter and onions.) In fact, 34 of the 45 foods were rated as belonging to at least two script categories. For example, corn was considered a healthy food (mean rating of 6.5) and a dinner food (6.0). Thus, subjects clearly believe that foods belong to these script categories.

Second, although foods were often rated as belonging to the script categories, their ratings did differ somewhat from those of the taxonomic categories (see Table 3). Foods were often (34 of 45 foods) viewed as belonging to just one of the taxonomic categories (e.g., corn was judged to be a vegetable, with a mean rating of 5.5), whereas with the script categories, as just mentioned, they often belonged to two or more.[2] Only 9 of the foods were viewed as belonging to just one script category.

One possible explanation is that foods are very good members of one taxonomic category but are less good members of a number of script categories. However, the data are not consistent with such an explanation. If one examines only foods with a mean rating of 6.0 or greater (very good to excellent members), then 22 of the 45 foods are rated this highly for some taxonomic category and 25 are for some script category. Although it is true that there are two more script categories than there are taxonomic, the mean number of highly rated items per category also shows only a small difference (3.67 for taxonomic vs. 3.13 for script). Thus, it is not the case that the foods are poorer members of the script categories.

[2] We were surprised that 10 foods were not considered good members of any of the taxonomic categories. However, as mentioned earlier, the selection of the items included both ones generated from the taxonomic categories we thought of and foods that are common ones for undergraduates. Thus, there was no assurance that we had an exhaustive set of taxonomic categories. The food items not considered good members of any of our six taxonomic categories were cake, chocolate bar, cookies, nuts, pancakes, pie, pizza, popcorn, potato chips, and spaghetti (though pancakes and spaghetti were close to the criterion for breads and grains). As one can see from this list, some of the items do not fall clearly into any of the taxonomic categories.

TABLE 4
Proportion Distribution of Mean Ratings in Experiment 2

	Mean ratings						
Type of category	0–0.9	1.0–1.9	2.0–2.9	3.0–3.9	4.0–4.9	5.0–5.9	6.0–7.0
Taxonomic	.82	.03	.01	.00	.03	.02	.08
Script	.33	.16	.11	.12	.10	.06	.11

A fuller picture of the ratings can be seen in Table 4, which shows the distribution of mean ratings for the two different types of categories. That is, these are the proportions of mean ratings over the six taxonomic or eight script categories. Although the proportions of high ratings are about equal for the two types of categories, the script categories have far more foods in the intermediate ratings. As mentioned, foods are often rated as being in more than one script category, but usually they are rated as 6.0 or more in only one script category (that is true for 40 of the 45 foods). For example, rice is rated as an excellent member of the dinner foods category (6.4) but also as a good member of the lunch foods category (4.2).

Another difference between the category types is that the taxonomic categories had fewer members; as can be seen from Table 4, the taxonomic categories have a much greater proportion of foods that are rated as nonmembers of the category—indeed, the majority of their ratings were less than 1.0. Script categories had only a third of the foods rated less than 1.0. As one extreme example, the script category lunch foods had a rating of less than 1.0 for only 1 of the 45 foods (lobster). The number of script categories' nonmembers ranged from 1 to 26 per category, while the taxonomic categories' nonmembers ranged from 27 to 42, so the distributions were nonoverlapping. As one can see from Table 4, much of this difference was due to script categories having far more marginal members (mean ratings of 1.0 to 3.99) than did the taxonomic, but as mentioned, the script categories also had a slightly higher proportion of good members.

The results of this experiment provide support for the category generation findings for script categories. Food items were found to be typical of both taxonomic and script categories. In addition, the distribution of membership appears to be different for these two kinds of categories, at least for our food sample. Taxonomic categories have a number of very good members, a very large number of nonmembers, and only a few members in between. Script categories, on the other hand, have relatively few nonmembers and a larger proportion of poor and fairly good members. The script categories have at least as many excellent members (ratings of 6 or higher), but a much wider distribution, with many items near the boundary of membership. One interpretation of these data is that the taxonomic categories appear to have a more well-defined criterion for category membership—a food is either a good member of the category or it is not a member. The script categories, however, appear to have much more ambiguity about category membership. For example, what qualifies as a lunch food? Although there are certainly very typical lunch foods, such as soup, hamburgers, and the like, many things can be eaten for lunch, and indeed, most people have probably had pancakes or corn at lunch even if they are not at all usual. It is very difficult to rule something out as a lunch food, since lunch is primarily determined by the time of day it is eaten, rather than in terms of the kind of food consumed then.

This interpretation provides one account of the data, but other interpretations are also possible (Lawrence Barsalou, personal communication). For example, it may be that there is no difference in how well defined the different types of categories are, but rather the features may be differentially diagnostic (i.e., more diagnostic for taxonomic than script categories). Another possible interpretation for the differences in the distribution of ratings is that the competition among script categories may lead to reduced ratings for some category members. For instance, in deciding on the rating for bagel as a snack food, if one took into account that it was really a very good breakfast food, then one might give a lower rating for snack food (though this would require considering script categories separately from taxonomic categories). Although the exact interpretation of the difference in ratings distribution awaits further research, it is clear that foods are viewed as belonging to script categories.

Experiment 3: Category Sortings

The category rating task indicates that people believe foods are members not just of taxonomic categories but also of script categories. Although the rating task is informative, it does not show that these script categories constitute an important part of the representations of foods. For example, people would be able to rate foods along a number of property dimensions (size, color, cost), but these properties might be a relatively unimportant part of the representation of foods. (Of course, the script categories were consistently produced by subjects in Experiment 1, unlike these other dimensions, suggesting that the script categories may have a more prominent role.)

In this study, we examined people's sortings of food terms as an additional indication of their underlying organization of the category foods (as Lopez et al., 1997, and Medin et al., 1997, did). The sortings in conjunction with the ratings provide a helpful look at how people organize food categories. There were three groups of subjects, with one group instructed to sort by taxonomic categories (types), one by script categories, and one asked only to sort the foods into groups that go together. The main question of interest concerns the data from this last group not given any specific basis for their sortings—do the script categories appear to influence the sortings of this group? The first two groups provide an idea of what the taxonomic and script organizations would be like for these items, which will be helpful in interpreting the results from the nondirected group.

Consider the possible outcomes for this nondirected group. First, they may sort the food items just as do the taxonomic subjects, with no influence of script categories. Second, they may sort by scripts, with no influence of taxonomic categories. Third, and predicted from the generation and rating resultss, their sorts may be influenced by both the taxonomic and script categories. Fourth, they may sort in some other way that has little relation to either the taxonomic or script category sorts.

Method

Subjects. The subjects were 94 undergraduates who received course credit or pay. There were three groups: those who sorted by the taxonomic instructions ($n = 29$), those who sorted by the script instructions ($n = 27$), and those who sorted by the default (neither taxonomic nor script) instructions ($n = 38$).

Materials. The same 45 foods were used from the earlier two experiments (see Table 1). Each food term was typed on a 3 × 5-in. (7.6 × 12.7-cm) white index card.

Procedure. The instructions informed subjects that the study was to find out how people categorize types of foods. They were to read through all the cards once and then were to divide them into groups. The taxonomic group was told to divide into groups ''of *similar food types.* That is, you should group together items that are the same kind of food.'' The script group was told to divide the foods into groups ''of *foods that are eaten at the same time or in the same situation.* That is, you should group together items related by when and how they are encountered.'' The default group was told to simply divide the foods into groups ''of things that go together.'' Subjects were told to make as many groups as they liked and to move the cards around until they were satisfied. The instructions required subjects to make at least two groups and put at least two cards in each group.

After the subjects had sorted all the cards, they were asked to say why they made each group (''what about these objects made you want to put them together?''), and the experimenter wrote down their answers. The experiment took about 15 min.

Results

We first present some descriptive analyses of these sorts for the different groups and then provide further analyses to more finely examine the underlying representations. The descriptive analyses concern the number of piles the subjects sorted the 45 foods into and the labels they gave to these piles.

Across all groups, the number of piles that the subjects sorted the foods into ranged from 2 to 21, but the means were similar: 8.7 for the taxonomic group, 7.8 for the script group, and 8.1 for the default group (the corresponding medians were 8.0, 7.6, and 8.3). These means did not differ statistically, $F(2, 91) = .54$, $MSe = 9.52$.

The explanations subjects gave for each pile were classified as being taxonomic or script. Most of the explanations were one of the taxonomic or script labels given in Table 2 (though sometimes at a lower level, such as ''white meats''). A few additional taxonomic (e.g., seafood) and script categories (e.g., movie food) were mentioned, but they accounted for a very small num-

TABLE 5
Proportion of Sortings Given Taxonomic or Script Labels in Experiment 3

	Labels given	
Sorting instructions	Taxonomic	Script
Taxonomic	.60	.22
Script	.11	.68
Default	.56	.30

ber of observations. Table 5 shows the different proportions of labels that were classified as taxonomic or script for the three different sorting conditions.[3] Three points are worth noting. First, it is clear that the taxonomic and script groups labeled their groups very differently. The instructional manipulation appears to have affected the sortings. Second, although each of these two groups primarily sorted into their respective kinds of categories (i.e., taxonomic or script) a substantial minority of the sorts were of the other kind (i.e., script or taxonomic, respectively), suggesting that both forms of organization are salient ways of conceiving of food. Third, the default group's labeling fell in between the two other groups, but clearly was much closer to the taxonomic condition. Nonetheless, 30% of the piles in the default group were labeled with script category names (especially snacks, junk foods, breakfast foods, and desserts).

The intercorrelations among the sorts provide additional support for these observations. For each instruction group, we constructed matrices of the proportion of subjects who sorted each pair of items together and computed the correlations among these matrices The default and taxonomic instruction matrices were very similar, with a correlation of .95. (Remember, however, that even the taxonomic instruction group sortings included script categories.) The script instruction matrix correlated less well with the default (.60) and taxonomic (.54), but still indicated considerable similarity among the different groups. The details of these are explored with further analyses.

Similarity scaling. These sorting data were analyzed by finding (through least-squares fitting) a set of ‘‘Robinson matrices’’ that best fit the original proximity matrix, using the method developed by Hubert and Arabie (1994). This method provides a representation of the similarity relations between foods that will allow us to address critical questions of how the different category types influence the sorting. In particular, the output matrices capture

[3] The other piles were given labels relating to: macronutrients (carbohydrates, proteins), ethnic foods (Italian), cooking technique (baked foods), evaluation (foods I like to eat), or how it is used with other foods (pizza topping, put sauces on). Only carbohydrate and Italian food were used as labels by more than a few subjects.

the similarities among the foods and provide an easy-to-read graphical presentation in which more similar foods are nearer each other in the matrices.

This technique begins with a similarity matrix of the raw proximity data—in our case, the proportion of times every pair of items was placed into the same category. The general goal of the method is to model this matrix by the sums of (reordered) matrices that have Robinson form. A matrix has Robinson form if the cell entries along a row or column never increase as one moves away from the diagonal. (This means that rows and columns are ordered so that the further apart they are, the less similar the items are.) Normally, the raw proximities do not have this property, as the rows and columns may initially be ordered in a way that has little to do with the proximity structure (perhaps alphabetically). Thus, the first step of the technique is to find an order of the rows and columns that results in a matrix that is close to being in Robinson form. (In a similarity matrix, the rows and columns represent the same items. For example, if the entries for ‘‘bagel’’ are in row 2, then column 2 is also ‘‘bagel.’’) Computationally, the rows and columns are permuted until a reordered matrix that is close to Robinson form is found. (Note that reordering the rows and columns does not change any of the proximity information in the matrix.)

Such a permutation generally does not produce a perfectly Robinson matrix, however. As a result, it is necessary for an algorithm to go through each column and row and find places where a value increases rather than decreases over adjacent cells closer to the diagonal. Both of these cells are replaced with the mean of the two, bringing them into (local) conformity with the Robinson rule. This process (which our description is simplifying a bit) must be done iteratively until the entire fitted matrix is in Robinson form. The Hubert and Arabie (1994) algorithm ensures fitted matrices that are a least-squares minimum.

Such a procedure captures some aspects of the proximity structure of the stimuli (as will be explained shortly). However, because of the changes to the cell entries done to make the matrix have Robinson form, the new matrix does not represent all the information in the original matrix. Therefore, this process can be performed again, on the residual matrix, that is, the matrix resulting from the original proximity matrix minus the fitted matrix. The second matrix (and third, etc.) includes variance that was not explained in the first matrix. In practice, the matrices are simultaneously fit to the data so that changes to one fitted matrix then result in changes in the others. The result of this is that the fitted matrices tend to capture different aspects of the data, but together they converge on a least-squares minimum solution (see Hubert & Arabie, 1994, for a detailed description). In our analyses, the first matrix accounted for over 80% of the variance, but the information in the second matrix was often interpretable as well, revealing structure not apparent in the first matrix. Because the two matrices together accounted for more than 95% of the variance in each condition, we stopped at two matrices.

The Robinson form is a useful one, because it results in similar items being close to one another in the matrix. That is, by ensuring that each row is nonincreasing as one moves away from the diagonal, the values near the diagonal must be high, and the cells near the diagonal represent the similarity of items that are near one another in the matrix. For example, the similarity of item n to item $n + 1$ is indicated in a cell adjacent to the diagonal, and the similarity of item n to item $n + 2$ is one step away from the diagonal, etc. So, creating a Robinson form results in a matrix where adjacent items are very similar, and similarity decreases between items that are farther and farther apart. By reading the entries in the matrix, one can see how similar each item pair is. Because the matrix can be difficult to read, however (especially when there are 45 items, as in our stimuli), we used a graphical representation that identifies similarity levels above various thresholds (Hubert & Arabie, 1994), as exemplified in Fig. 1. In this figure, cells that have a square indicate that the items in the corresponding row and column are highly similar (z score of 3.0 and greater), those that have a circle are highly to moderately similar (z score of 2.0–2.99), those with a triangle are moderately similar (z score of 1.0–1.99), and those with only a line are simply above-average similarity (z score of .0–.99). This division highlights high similarities. Those cells with no marks (the majority of the matrix) have below-average similarity. One can see from this matrix that the items form fairly distinct clusters of similarity groupings. Furthermore, by comparing the different symbols, one can distinguish various levels of similarity within those groupings. (Note that this information requires only one triangular half of the matrix, but we have duplicated it on both sides of the diagonal so that the pattern is easier to see.) One can also see interesting results that are not apparent in most clustering, tree, or multidimensional scaling solutions. For example, note that rice is somewhat similar to items in two large clusters, suggesting that it is a ‘‘spanner,’’ connecting the two. Such an item might simply appear as a singleton in a tree structure, but this does not represent the fact that it is actually sorted with two or more clusters.[4] Such items will be important in understanding the structure of food categories we present below. Our focus will be on the default sortings, because these instructions did not bias subjects toward any particular groupings. However, we will also briefly discuss the taxonomic and script sortings.

The matrices for the default sortings are presented in Fig. 1 and 2. Together, these matrices accounted for .985 of the variance in the data. First,

[4] One reviewer was concerned about whether this analysis encourages finding spanners. If the sortings of two categories are clearly distinct, then no spanner will be found. However, if some subjects include an item in one category and some in the other, then it will be shown as a spanner. This analysis distinguishes between items that are sorted in this way (as fairly good members of two groups) and items that are in between two groups but are not members of either group.

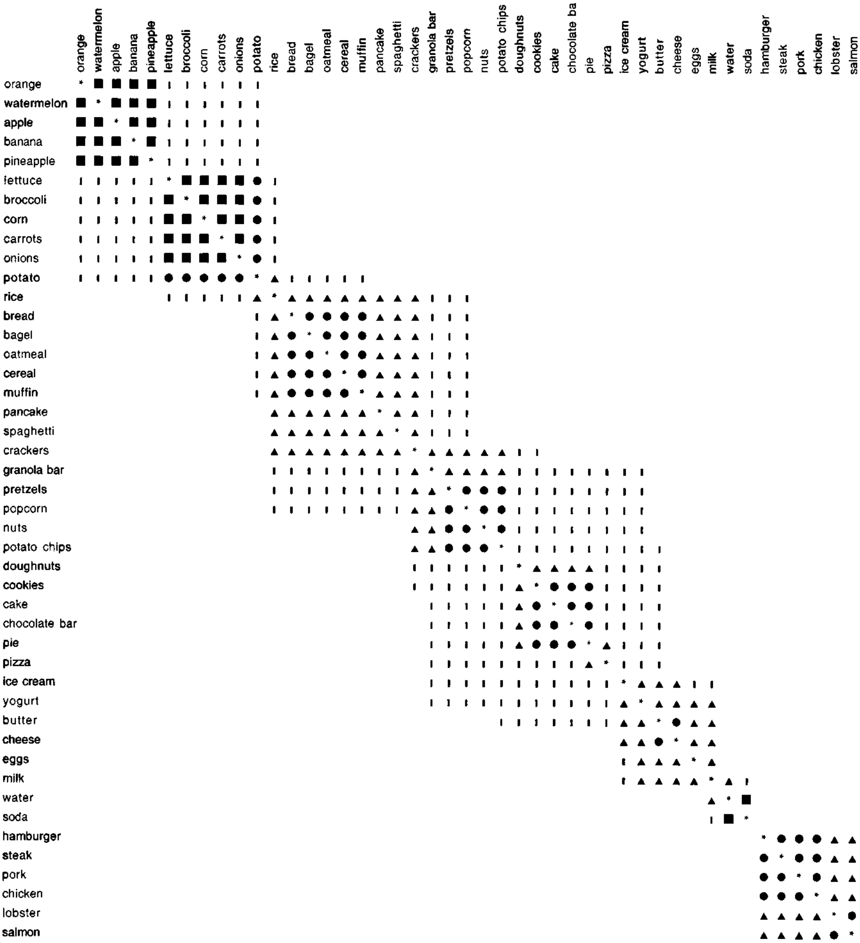

FIG. 1. Primary Robinson matrix (see text for explanation) of food items for Default sorting instructions in Experiment 3. The symbols indicate the extent of similarity between the corresponding row and column items (from square as most similar, to circle, triangle, and line, with blank cell entries for pairs with below average similarity).

let us concentrate on Fig. 1, which is the primary Robinson matrix for the default sorting. The taxonomic categories have a very strong influence on this sorting, but there is some evidence that the script information is influencing the sort as well. Looking at the upper left, one can see two strongly connected clusters, fruits and vegetables, that are weakly connected to each other. Looking at the lower right, one can see three clusters that appear to be dairy foods (from the category ratings we know that about half the undergraduates consider eggs to be a good example of a dairy food), beverages,

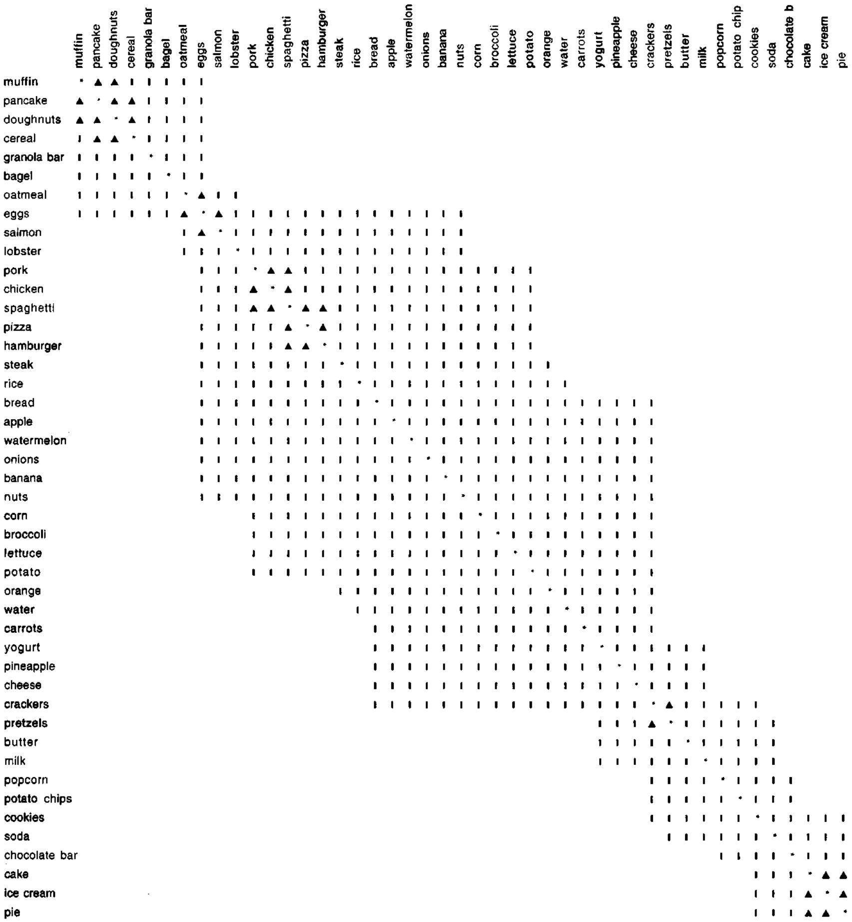

FIG. 2. Secondary Robinson matrix (see text for explanation) of food items for Default sorting instructions in Experiment 3.

and meats (with perhaps a fish/seafood subgroup apart from other meats). These five clusters all show the strong influence of the taxonomic categories on the default sortings, and in fact the taxonomic sorting primary matrix is very similar to the default one (see Appendix A). The middle part of Fig. 1 includes many of the breads and grains, but three clusters stand out that appear to be breakfast foods, snacks/junk foods, and desserts/sweets. These clusters, therefore, have some script qualities. This leads to muffins being clustered with breakfast foods like cereal, oatmeal, and pancakes, while doughnuts is grouped with desserts like cookies, cake, chocolate bar, and pie, and nuts are put in with snacks like pretzels, popcorn, and potato chips. Thus, there is evidence of script category influence on the default sortings.[5]

These analyses of the sorting data necessarily collapse over data for individual subjects, so it is important to examine whether the influence of script categories on the default sortings occurs at the level of the individual data as well. There are many ways that this could be addressed, but an examination of the labels given the piles in the default sortings provides a clear and easy to understand answer. Of the 38 subjects, each of the six taxonomic categories were mentioned by between 21 and 28 subjects. For the script categories, the three mentioned most often were snack foods (23 subjects), breakfast foods (19), and desserts (19). Thus, half or more of the subjects mentioned each of these script categories. In addition, 33 of the 38 subjects mentioned at least one script category. Thus, the conclusion from the overall data that script categories influenced default sortings is also true at the individual level.

Two additional points need to be discussed. First, as mentioned, a few items seem to span two clusters. For example, rice appears to be a spanner, connected somewhat to potato and a little to the cluster of vegetables, while also related to the more bread (and breakfast) products such as bread, bagel, oatmeal, spaghetti, and crackers. Similarly, crackers and granola bar seem to span the breakfast and snacks clusters, while doughnuts span the snacks and desserts clusters and ice cream spans desserts and dairy foods. Spanners give evidence that some foods are cross-classified very strongly. That is, they are simultaneously salient members of more than one group. For example, the ratings indicate that ice cream is a good member of both desserts (6.9 of 7) and of dairy foods (4.8). Second, in addition to the clusters in Fig. 1, one can also look across the whole ordering and see, roughly, a division between plant-based foods (fruits, vegetables, and breads and grains) and animal-based foods (dairy foods and meats). These may indicate the existence of superordinate or covert categories that encompass the more specific clusters of items. In addition, note that this way of representing the data allows a clear depiction of overlapping clusters, unlike simple multidimensional scaling and hierarchical clustering.

The secondary default matrix, given in Fig. 2, provides further evidence of an influence of script categories. First, in the upper left, there is a cluster of muffin, pancake, doughnuts, and cereal (with less connection to granola bar, bagel, oatmeal, and eggs), which appears to be a breakfast cluster. Second, the cluster of meats now has spaghetti and pizza as well (not present in Fig. 1), suggesting a main course grouping. Thus, a fuller representation

[5] We also conducted a number of other analyses of these data, but found that the Robinson form matrices presented the most comprehensive description. The other analyses did agree with the general findings. For example, the Additive Tree analysis solution (Corter, 1982) also had ice cream clustered with desserts and away from the dairy foods.

of these foods is gained from examining this additional matrix. Third, toward the lower right there is a cluster of desserts/sweets, in which ice cream is now separated from the other dairy foods.

The matrices for the taxonomic and script sortings accounted for .988 and .975 of the variance, respectively. The primary taxonomic matrix is very similar to the primary default matrix, as might be expected by the earlier descriptive statistics. There are strong fruit, vegetable, dairy food, and meat clusters. However, even in the taxonomic matrix there is some evidence that the sortings were not made strictly on the basis of the constitutive basis of the food, which also supports the label information in Table 5 that was discussed earlier. For example, ice cream appears to be in a dessert cluster, rather than in the dairy food cluster. In addition, the secondary taxonomic matrix shows further evidence of a script influence, with a cluster including bagel, muffin, doughnuts, pancake, eggs, and cereal. Although most of these are breads and grains, the inclusion of eggs (and the absence of other breads and grains) suggests a breakfast grouping. In addition, spaghetti and pizza are grouped here with the meats. Thus, there was some influence of script categories in spite of instructions to sort taxonomically (by ‘‘similar food type’’). Third, the script sorting matrices do show that people can sort many of the foods by scripts. For example, in the primary script matrix (Fig. 5) there is a cluster that appears to be dinner foods, because it is mainly meats, but hamburger is absent, and spaghetti is present. This cluster is somewhat connected to a vegetable cluster that also includes rice, again suggesting a dinner interpretation. There also appear to be dessert, snack, and breakfast clusters. Not surprisingly, the dinner and breakfast items are farthest apart in this figure, unlike in Fig. 1.

Discussion

The sorting results provide further evidence that people’s representations of foods are dominated by taxonomic categories but are strongly influenced by script categories as well. The default group’s sortings contain many clearly identifiable taxonomic clusters but also some clusters that are much closer to script categories (the matrices in Appendix A may be consulted for further evidence of this). Most importantly, the results give considerable evidence for cross-classification. Subjects are clearly able to categorize items in two rather different ways, for example, placing pizza with ice cream in one case but with hamburgers in another. In fact, the overall correlation of the taxonomic and script sorts was .54, suggesting some major differences between the two sorts. Also, these two ways can be found together in the same sorts and so are apparently not thought of as incompatible or contradictory. For example, the primary default sort appears to show a cluster of meats (a taxonomic category) along with a cluster of snack foods (a script category).

Two other points should be noted about interpreting these data. First, there may be many reasons that food items are put together, and interpreting the clusters as taxonomic or script may be missing the influence of the other. The categories formed from these two perspectives are often correlated. Meats (or at least most meats) may be together in the script sorting because they also tend to be dinner foods, and a cluster of breakfast foods may have many breads and grains. It is possible that the interpretation of the default matrix in terms of the taxonomic categories underestimates the script influence.

Second, the sorting task may lead to more conformity to a single perspective than might be true of the underlying representations. The requirement to form distinct piles means any food can only be put into one group. Thus, once some foods are sorted it may be that the others are sorted by the same perspective. Suppose a subject sorts some of the foods by their taxonomic categories and then tries to sort the remaining foods. They cannot be put in the same piles as their script-related neighbors, because those are already taxonomically sorted. For example, if one has made a pile of beverages, then one cannot put cereal with milk, since milk is already with soda and water. Our point is simply that very salient taxonomic similarities (e.g., fruits or vegetables) may be leading to more taxonomic sorting of less strong categories (e.g., breads and grains) than might be representative of people's conceptual organization of food. (This problem might have been avoided by obtaining pairwise similarity ratings, but each subject would have had to make 990 judgments, which was prohibitive.)

CATEGORY ACCESSIBILITY

The results of the first set of studies have provided evidence that people do cross-classify foods and that these cross-classifications have rather different organizations. However, it does not follow from this finding that people activate both kinds of categories when actually thinking about specific foods. One possibility is that the script categories are activated only in special contexts or only in tasks in which general questions about similarity or category membership are posed (Smith & Sloman, 1994). For example, perhaps when trying to generate categories for nuts and popcorn, the concept of movie food comes to mind, but when thinking about either one alone, this category is not actually activated or used. The next two experiments focus on whether taxonomic and script categories are activated when making similarity judgments of pairs of items or when making category judgments for single items.

Experiment 4: Similarity Ratings

In Experiments 4 and 5, we examined what knowledge may come to mind when a food is encountered. When an object is perceived or a word is read, concepts related to that item may become more accessible or activated (Barsalou, 1982; Murphy, 1991). This accessibility means that the related concept would be available to aid in comprehension, to make inferences, etc. Of special interest here is whether the script category information may be activated when a food item name is read. Barsalou (1982) distinguished information about an item that is activated across all encounters (i.e., context-independent information) and information that is activated only in specific situations (i.e., context-dependent information). For example, the concept of a basketball may always activate the property ''round'' but activate the property ''floats'' only in specific contexts. We asked whether the taxonomic and script categories were context-independent or context-dependent information for the food items.

Barsalou (1982, Experiment 2) investigated the distinction between context-independent and context-dependent categories by having people rate the similarity of items from the same category, preceded or not by the category label. His reasoning was that if processing the items would lead to the category label being activated anyway, then the label's presence should have little effect on the similarity rating. If, however, subjects do not normally activate the category, then presenting the category label might well affect the rating. Barsalou contrasted two types of categories, common taxonomic categories (e.g., birds, furniture) and ad hoc categories (e.g., plunder taken by conquerors, can be a pet). These ad hoc categories are ones that people readily can use but are not usually activated by the presentation of the items (see Barsalou, 1983). Thus, he had some pairs such as ''desk–sofa'' from the taxonomic category furniture and some pairs, such as ''raccoon–snake'' from the ad hoc category can be a pet. He found that the similarity ratings of the common taxonomic categories were little affected by the presence of the category label, whereas the ratings of ad hoc categories showed a large difference.

In this experiment, we followed Barsalou's technique using three kinds of food categories: script, taxonomic, and ad hoc categories. In each case, we asked subjects to judge the similarity of items from a category with and without the category name. If the items themselves activate the categories, then providing the names should not influence their judgment. However, if the items do not spontaneously lead subjects to notice the categories, then providing the category names should influence their judgments—in particular, they should increase the similarity rating, since it will have been brought to their attention that the items share category membership. Based on Barsalou's (1983) findings, we predicted that providing taxonomic category names will not influence ratings very much since they are automatically activated (though this has not yet been demonstrated for food categories) but that providing ad hoc category names will increase perceived similarity. The main question of interest, then, is the effect for script categories. If the script categories are always highly activated by the presentation of the items, then they

TABLE 6
Examples of Pairs Used in Experiment 4

Categories	Items	
Taxonomic		
Fruit	Watermelon	Strawberry
Dairy food	Yogurt	Cheese
Script		
Breakfast food	Bagel	Bacon
Snack food	Apple	Pretzel
Ad hoc		
Foods that are often cooked in water	Spaghetti	Broccoli
Foods that squash easily	Tomato	Marshmallow

should show a result similar to that of the taxonomic categories, that is, no effect of providing the category name. If the script categories are never activated, then they should show as big an effect of the category name as the ad hoc categories do. If they are sometimes or partially activated, then they should show an effect between those of the taxonomic and ad hoc categories.

Method

Item selection. Before obtaining similarity ratings, we needed to get instances of the three kinds of categories (taxonomic, script, and ad hoc) that were rated as good members of their respective categories. We chose six taxonomic categories, eight script categories, and eight ad hoc categories, with six to eight instances per category, for a total of 141 category–instance pairs. Ten subjects were given booklets consisting of these pairs, with the category on the left and the food instance on the right in boldface (e.g., vegetable **corn**). Their task was to give a 0 to 7 typicality rating for each category–instance pair, with a 0 meaning that the instance is not a member of the category, a 3 meaning it is a fairly good member of the category, and a 7 meaning it is an excellent (very typical) member of the category, as in Experiment 2. These data were tabulated and used to select six categories of each kind (taxonomic, script, and ad hoc) and four instances of each category. The mean ratings for taxonomic, script, and ad hoc instances chosen were 6.22, 6.18, and 6.20, respectively. Thus, the items tested in the main experiment were equally typical of the three category conditions. Some examples are given in Table 6. The full list of categories and pairs is given in Appendix B.

Subjects. The subjects were 49 undergraduates who received course credit or pay. The experiment took about 15 min.

Materials. The 18 categories (6 each of taxonomic, script, and ad hoc, as just described) had four instances each which were split into 2 pairs, for 36 total pairs. In constructing these pairs, care was made to avoid pairing two instances that were also in another type of category. For instance, in pairing items for the script category breakfast foods, pancake and waffle would not be put together because they also belong to the taxonomic category breads and grains. These 18 pairs were split into three pages, with 2 pairs of each category kind on each page and no category occurring more than once per page. The pages were randomly ordered for each subject.

For the condition without category names, the pages consisted of six pairs, with the two

TABLE 7
Mean Similarity Ratings for Food Pairs in Experiment 4, with and without the Category Label

Categories	With category label	Without category label	Difference
Taxonomic	4.40	4.32	.08
Script	3.80	2.77	1.03
Ad hoc	3.41	1.94	1.47

Note. A 1 to 7 scale, with 7 most similar, was used.

words in each pair next to each other and the pairs separated by two blank lines. In the condition with category names, the category label was to the left of the food pair and printed in boldface. The same word pairs were used for all subjects, with the manipulation of presence or absence of category names made between-subjects.

Procedure. The instructions informed subjects that the goal of the study was to find out how people think about the similarities of foods. They would be given the names of two foods and were to think of the two foods to which the words refer. Their task was to rate how similar the foods are, using a 1 to 7 scale, with 1 meaning the foods are not at all similar and 7 meaning the foods are very similar. In the category–pairs condition, instead of simply being told they were to be given pairs of foods, subjects were told that they would be given a category and two foods from that category. No indication was given as to how the category should be used in making the judgments (following Barsalou's, 1982, procedure).

Results and Discussion

The result of main interest is how the similarity ratings in the three kinds of categories changed as a function of whether the category label was presented. The similarity ratings were averaged for each kind of category for each subject, and then the ratings of the two groups of subjects were compared. As can be seen in Table 7, the change was greatest in the ad hoc category and least in the taxonomic category, as predicted. The change in the script category was intermediate, though closer to the ad hoc change.

A two-way ANOVA was conducted on the data with a between-subjects factor of presence of the category name and a within-subject factor of category type (taxonomic, script, and ad hoc). We also carried out item analyses on the results to be sure that the effects were not due to a small number of unrepresentative food terms. However, we should note that because the ad hoc categories are a kind of artificial construction (e.g., subjects did not list them as categories of our food types in Experiment 1), it is not clear that the question of generalizing to the population of ad hoc categories (the usual goal of item analyses—see Clark, 1973) is a well-founded one. Nonetheless, the item analyses are useful in demonstrating the relative robustness of effects across our stimuli.

Not surprisingly, there was an overall effect of category type, with items in taxonomic categories rated as more similar than items in script categories, which were more similar than items in ad hoc categories, $F(2, 94) = 94.20$, *MSe* = .378, $p < .0001$, with all pairwise differences significant by a Neuman–Keuls test. This overall effect is difficult to interpret because there were different items in the three kinds of categories and they were not selected randomly from their respective categories. It may well be that items from taxonomic categories are more similar than those from script categories in general, but the data here do not allow a definitive answer to that question. More importantly, there was a reliable interaction of label condition and category type, indicating that the categories were differentially affected by the manipulation of including the category label, $F(2, 94) = 16.52$, *MSe* = .378, $p < .0001$. To better address the main question of how the script categories fared relative to the taxonomic and ad hoc categories, two further analyses were carried out to compare the effect of the label on script category ratings versus the other two category types. First, the difference between the two labeling conditions for taxonomic categories was found to be less than the difference for script categories, $F(1, 47) = 16.13$, *MSe* = .344, $p < .001$. Second, the difference for the script categories was less than that for the ad hoc, $F(1, 47) = 4.55$, *MSe* = .269, $p < .05$. Item analyses showed similar results, but the script versus ad hoc difference was only marginally reliable: $F(2, 33) = 17.65$, *MSe* = .173, $p < .0001$, for the overall interaction; $F(1, 22) = 18.55$, *MSe* = .147, $p < .001$, for the taxonomic versus script comparison; and $F(1, 22) = 2.98$, *MSe* = .199, $p < .10$ for the script versus ad hoc comparison.

This experiment shows that script categories do appear to be activated by the presentation of the items. The amount of activation was significantly less than that for taxonomic categories, but significantly more than for ad hoc categories. It may be that the script categories are activated less strongly than taxonomic categories, less often, or both.

Experiment 5: Priming and Speeded Category Verification

Experiment 4 found that the script categories are activated by pairs of items in the same category, but it did not show that a single item activates its script category (or categories). In addition, the judgment in Experiment 4 was not speeded, so subjects might have activated this information only after considerable thought (see Smith & Sloman, 1994). It is possible that knowledge about script categories is not usually activated but can be with extended time to think about the items. In this experiment, we examined the accessibility issue for script categories from a single item using a speeded category verification judgment.

To get a measure of accessibility in this experiment, we again adapted a procedure from Barsalou (1982). Barsalou (1982, Experiment 1) had subjects make speeded judgments about whether an item had a particular context-independent property (e.g., ‘‘Skunk—Has a smell’’) or context-dependent property (e.g., ‘‘Roof—Can be walked upon’’). These judgments were made immediately after subjects read a sentence that primed the property (e.g., ‘‘The skunk stunk up the entire neighborhood’’ or ‘‘The roof creaked under the weight of the repairman’’) or did not prime the property (e.g., ‘‘The skunk was under a large willow tree’’ or ‘‘The roof had been renovated prior to the rainy season’’). Barsalou argued that if a property is context-independent for an item, then it should be activated whether the sentence primes it or not, leading to the prediction that the time to verify that a skunk has a smell should not be affected by which skunk sentence was read. However, context-dependent properties are activated only if the context primes them, so the judgment about whether a roof can be walked upon will be faster if the preceding sentence primed this property for roof than if it did not. This is the pattern he obtained, with context-independent properties showing no effect of priming and context-dependent properties showing a large priming effect.

In Experiment 5, people made speeded judgments to category verification questions (e.g., about whether a bagel is a breakfast food) for taxonomic, script, and ad hoc categories. Preceding this judgment, they read a sentence that either primed that category for that food (e.g., ‘‘The bagel was what he had when he woke up’’) or was neutral with respect to that category (e.g., ‘‘The bagels were in the last aisle in the store’’). Based on earlier results, we expected ad hoc categories to show a large effect of priming and for taxonomic categories to show little if any priming effect. The question of interest was whether the script categories show a priming effect. If reading ‘‘bagel’’ leads to the activation of the script category breakfast foods, then it should not matter whether the sentence involves bagel as a breakfast food (similar to the context-independent properties and the taxonomic categories). If, however, breakfast food is not activated by reading ‘‘bagel,’’ then the preceding priming sentence should lead to a faster verification than the preceding neutral sentence. If the script category is activated partially or only some of the time by the food name, then the prior context should prime it somewhat: less than for ad-hoc categories but more than for taxonomic categories.

Method

Item selection. Before conducting the priming study, we chose materials on the basis of two ratings to ensure that the results could be interpreted as intended. First, the materials for the three types of categories (taxonomic, script, and ad hoc) were equated for how good category members the items were, as in Experiment 4. Second, and as in Barsalou (1982), the strength of the priming manipulation (primed versus neutral sentences) was equated for all three types of categories so that any reaction time differences across category types cannot be attributed to the priming sentences being more related to the category.

The first rating was done exactly as in Experiment 4, with subjects judging the goodness of category members. (We also used the results from the earlier ratings, but we needed additional items both because there were more items in this study and because no food item was presented on more than one trial.) Ten subjects were given booklets of 124 item–category

TABLE 8
Samples of Materials used in Experiment 5

	YES responses
Taxonomic: Soda—Beverages	
Primed:	The soda was poured into a glass by the waiter.
Neutral:	The soda was so popular the store was sold out of it.
Script: Pretzels—Snack foods	
Primed:	The pretzels were what he ate between meals.
Neutral:	The pretzels had a much higher price than usual.
Ad hoc: Soup—Foods that you eat with a spoon	
Primed:	The soup was put into a bowl for eating.
Neutral:	The soup was ready to eat.
	NO responses (Primed refers to whether the sentence primed some other category)
Lemonade—Foods that you can throw far	
Primed:	The lemonade was what he liked when he was thirsty. (‘‘Beverages’’ primed)
Spaghetti—Breakfast foods	
Neutral:	The spaghetti was something that her aunt always liked.

Note. The primed and neutral versions for YES responses were shown to different subjects.

pairs, exactly as for the ratings of Experiment 4, and judged typicality. Combining these and the earlier ratings, we took 4 items from each of the 18 categories—6 each of taxonomic, script, and ad hoc—to have 72 critical food items. The mean typicalities of the items in these three types of categories were 6.21, 6.21, and 6.24, respectively.

The second rating was to ensure that the priming sentences were equally good primes for all category types (and that the neutral sentences were equally neutral). For the main experiment, we needed 72 predicates (the sentences without the food item subject) to have 2 priming and 2 neutral predicates for each of the 18 categories. We constructed 164 pairs of predicates and categories, with 4 to 9 priming predicates and 4 or 5 neutral predicates for each of the 18 categories. (See Table 8 for some examples of the chosen predicates.) Following Barsalou (1982), subjects rated how much a predicate made them think of the food category (from 1, ‘‘not at all,’’ to 7, ‘‘immediately’’). The predicate was shown on the computer screen for 2 s and then the food category was presented just below the predicate. For example, the predicate ‘‘was poured into a glass by the waiter’’ was followed by the category ‘‘beverages.’’ Subjects typed in their ratings. We then chose the 72 predicates so as to equate the different categories as much as possible. The mean ratings for the chosen priming predicates were 6.2, 6.1, and 6.1 for the taxonomic, script, and ad hoc categories, respectively, and the corresponding means for the neutral predicates were 3.0, 2.5, and 2.6. (Although we chose the lowest rated neutral predicates for the taxonomic condition, it was not possible to find ones that were rated as low as for the script and ad hoc categories.)

Subjects. The subjects were 20 undergraduates who received course credit or pay. The experiment took about 45 min.

Materials. The critical materials were 72 sentences constructed from the 72 food items and 72 predicates chosen in the two ratings described above. For each of the three category types (taxonomic, script, and ad hoc), we chose 6 categories that were the same as the ones used in Experiment 4 (except that the ad hoc category Foods that require little preparation was substituted for Foods that you carry in a paper bag). There were 4 predicates for each of the 18 categories, 2 primed and 2 neutral. All sentences had the structure: ‘‘The’’ + the food item name + the predicate. The food item name was sometimes pluralized to make the sentence read more smoothly. Some examples are provided in Table 8.

All of the critical sentences are followed by category verifications for which the correct response is ''yes,'' so an equal number of ''no'' responses was needed. We took 72 additional predicates that consisted of 2 priming and 2 neutral for each category and paired them with 72 additional food items (4 from each category). Thus, the study sentences for these ''no'' responses had roughly the same characteristics as the critical sentences, though they were not equated by ratings as the critical sentences were. In addition, the same categories were presented in the verification part (and each category was again presented four times). However, the categories chosen were ones that the food item in the sentence did not belong to, so that the correct response would be ''no.'' Some examples are given in Table 8.

A second list of materials was constructed to counterbalance the particular food items and priming condition. In particular, for half the subjects the critical food items in a category that had been presented in primed sentences were now presented in neutral sentences and vice versa. Thus, across subjects, each food item occurred equally often in the primed and neutral conditions.

Eighteen practice items were also constructed to give subjects some experience in making these speeded judgments. Half the sentences were primed and half neutral, and half the category verifications were true and half false. Each category was used once during this practice, so subjects saw all the categories before the critical trials began.

To ensure that subjects read the whole sentence, recognition tests for the sentences were constructed with half true and half false sentences. Following the 18 practice trials, a 4-item recognition test was given. Following the 144 experimental trials, a 20-item recognition test was given.

Procedure. The instructions informed subjects that the experiment examined how people decide that a particular food is a member of a food category. A sentence would first be displayed on the computer screen with a food name as its subject (e.g., ''The soda was poured into a glass by the waiter''). They were to read the sentence carefully, because they would be tested on their memory for these sentences. After a few seconds, the sentence would disappear and they would be shown a food category, such as ''beverages'' (they were given foods to take on a picnic and proteins as examples). Their task was to decide whether the food from the sentence is a member of this category (i.e., is soda a member of the beverages category?) and press a button labeled YES (with the right index finger) or NO (with the left index finger). An illustration was given using the category vehicles. Subjects were told to respond as fast as they could while maintaining a high level of accuracy.

The experiment was programmed using PSYSCOPE (Cohen, MacWhinney, Flatt, & Provost, 1993) and run on Macintosh computers with a button box. The sentence was displayed above the middle of the screen for 3 s and was erased as the category was displayed in the middle of the screen. Subjects responded by pressing a button labeled NO on the left or YES on the right. If the response was correct, the next trial began 1 s later. If the response was incorrect, the phrase ''Incorrect response'' was displayed for 2 s and erased, and the next trial began 1 s later. If no response was made within 5 s, a message was displayed asking subjects to respond more quickly, and the next trial was presented 1 s later.

Following the 18 practice trials, subjects took the 4-item recognition test. This test was to remind them to read the sentences carefully, and they were told again that another recognition test would be given later. The 20-item test was given after all of the 144 experimental trials.

Design. The experiment had within-subject manipulations of category type (taxonomic, script, and ad hoc) and priming (primed versus neutral), with six categories of each type and two sentences of each priming type for each category. In addition, between subjects, the food items for the primed and neutral sentences were switched for counterbalancing.

Results and Discussion

As expected, the taxonomic categories showed little effect of priming and the ad hoc categories showed large effects of priming on both reaction time (RT) and accuracy. The script categories showed intermediate levels of prim-

TABLE 9
Mean Reaction Times (in ms) and Error Proportions for Experiment 5[a]

	Primed		Neutral		Difference	
Category type	RT	error	RT	error	RT	error
Taxonomic	850	.049	849	.062	−1	.013
Script	926	.058	950	.102	23	.044
Ad hoc	1180	.056	1352	.124	172	.068

[a] The RTs and proportion errors are back-transformed from the ln and arcsine transformations, respectively.

ing. More specifically, they showed little effect of priming on RTs (and significantly less than the ad hoc categories), but they did show an effect of priming on accuracy, falling between the effects of taxonomic and script categories. Overall, there does appear to be a priming effect on script categories, but it is less than that of ad hoc categories, suggesting again that the script categories are partially activated (or sometimes activated) when food items are presented.

Reaction times. To reduce the effects of outliers, all analyses were conducted on natural log transformations of the observations (the reported numbers are the back-transforms of the means). For each subject, the mean of correct RTs for the conditions was calculated (each category type × priming), and these are given in Table 9. An ANOVA (3 × 2 × 2) was conducted on these RTs with category type and priming as within-subjects variables and counterbalancing set as a between-subjects variable. There were significant effects of category type ($F(2, 36) = 159.04$, $MSe = .011$, $p < .0001$) and of priming ($F(1, 18) = 5.77$, $MSe = .015$, $p < .05$). The priming by category type interaction was reliable ($F(2, 36) = 5.80$, $MSe = .009$, $p < .01$) with the taxonomic and script categories showing little effect of priming but ad hoc categories showing a large effect. This interaction effect supplied the pooled *MSe* (.009) that was used in all of the planned contrasts: priming effects for each category type (taxonomic, script, and ad hoc) and the interaction of priming with category type for script categories versus taxonomic categories and script categories versus ad hoc categories.

The priming effects of taxonomic, script, and ad hoc categories (−1, 23, and 172 ms, respectively) were examined, but only the last showed a significant effect ($F(1, 36) < 1$ for both taxonomic and script categories, but $F(1, 36) = 20.27$, $p < .0001$, for ad hoc categories). The priming effect for script categories was not different from that for taxonomic categories ($F(1, 36) < 1$) but did differ from that of ad hoc categories ($F(1, 36) = 6.81$, $p < .05$).

The data were also collapsed over subjects to examine item effects, and the pattern of results was similar, though the inferential statistics were not always statistically significant (probably because category type was tested within subjects but between items). Overall, there were effects of category type ($F(2, 69) = 82.70$, $MSe = .025$, $p < .0001$) and priming ($F(1, 69) = 5.40$, $MSe = .027$, $p < .05$), but the interaction of category type and priming was only marginally reliable, $F(2, 69) = 2.71$, $MSe = .027$, $p < .10$. Because category type was varied between items, individual analyses were conducted of the priming effects (with corresponding means of 4, 31, and 195 ms) and the priming effect by category type interactions for the script categories. As in the subject analyses, the taxonomic and script categories did not show effects of priming (F's < 1 for both analyses), but the ad hoc categories did show a significant effect of priming ($F(1, 23) = 6.91$, $MSe = .040$, $p < .05$). The priming effect of the script categories was not significantly different from that of taxonomic categories ($F(1, 46) < 1$, $MSe = .020$). Despite the 164-ms difference in priming effect, the script priming effect was not significantly less than that of ad hoc categories due to increased variability $F(1, 46) = 2.60$, $MSe = .033$, $p < .12$). Thus, although the item effects were not as strong as the subject effects, the pattern is quite clear: Ad hoc categories showed a large effect of priming, while the taxonomic and script categories did not.

Errors. The proportion errors were transformed by arcsines (Winer, 1971), and the back-transformed proportions are given in Table 9. There were significant effects of category type ($F(2, 36) = 3.89$, $MSe = .042$, $p < .05$), with fewer errors for taxonomic categories, and of priming ($F(1, 18) = 28.85$, $MSe = .025$, $p < .0001$), with higher error rates following neutral sentences. The overall priming by category type interaction did show a significant effect ($F(2, 36) = 3.86$, $MSe = .022$, $p < .05$). The priming effects for taxonomic, script, and ad hoc categories were .013, .044, and .068, respectively (untransformed, they were .021, .067, and .096). This ANOVA supplied the pooled MSe (.022) that was used in all of the planned contrasts. The priming effect was not reliable for the taxonomic categories ($F(1, 36) = 1.47$), but the effect was significant for both the script categories ($F(1, 36) = 11.88$, $p < .01$) and the ad hoc categories ($F(1, 36) = 26.29$, $p < .0001$). The priming effect of the script categories was not significantly different from that of the taxonomic categories ($F(1, 36) = 2.49$, $p < .15$) nor the ad hoc categories ($F(1, 36) = 1.41$). Thus, the errors show effects in both the script and ad hoc categories, though the script category priming effect is intermediate between that of taxonomic and ad hoc categories.

The item effects were very similar in pattern. Overall, there were effects of category type ($F(2, 69) = 4.34$, $MSe = .040$, $p < .05$) and priming ($F(1, 69) = 18.25$, $MSe = .039$, $p < .0001$), but the interaction of category type and priming was not quite reliable, $F(2, 69) = 2.36$, $MSe = .039$, $p < .11$. The priming effect was not statistically significant for taxonomic categories ($F(1, 23) = 2.91$, $MSe = .008$, $p < .11$), but it was significant for both script categories ($F(1, 23) = 7.18$, $MSe = .044$, $p < .05$) and for ad hoc categories ($F(1, 23) = 8.57$, $MSe = .064$, $p < .01$). The script category priming effect was marginally greater than that of taxonomic categories ($F(1, 46) = 3.25$, $MSe = .026$, $p < .10$) and not different from that of ad hoc categories ($F(1, 46) < 1$, $MSe = .054$).

Thus, across both RT and errors, the pattern that emerges is that the script categories show some priming effect but it appears to be intermediate between that of taxonomic categories, which showed little effect of priming, and that of ad hoc categories, which showed large effects of priming.[6] Similar to Experiment 4, these results indicate that script categories are partially activated (or sometimes activated) by the presentation of food item names.

INFERENCE

One of the most important functions of categories is to allow inferences and predictions. If you know an animal is a dog, what can you infer about its size, ferocity, eating habits? Classification itself is rarely the ultimate goal of categorizing: Knowing an animal is a dog is of little use unless that classification allows one to accomplish some goal, such as deciding whether the animal can be approached or if it is likely to be a pet. Because of its centrality in category use, the inferential function of categories has received increasing attention from psychologists in recent years (Carey, 1985; Keil, 1989; Gelman, 1988; Gelman & Markman, 1986; Murphy & Ross, 1994; Osherson et al., 1990; Ross & Murphy, 1996).

The question arises, then, whether the multiple categories that people apparently have for foods are useful for inference. It is conceivable that people can identify items as meats or appetizers but do not use these categories for making further inferences. For example, perhaps knowing that something is an appetizer says only that it is eaten before the main meal, without carrying other information. This question also may serve to separate the two kinds of categories we have been comparing. There may be a form of cognitive economy in which taxonomic categories are used to carry the bulk of the information about foods, and script categories are used only to indicate a minimal commonality. So, although it is useful to know that nuts are a typical appetizer, so that we can serve them before the meal, perhaps there is little else that this classification tells us.

In the next two experiments, we investigate the use of taxonomic and script categories for making inferences. We will examine whether both kinds of classification are used to form inferences and whether there are any distinctions to be made about what kinds of inferences they support.

[6] An interesting, but tangential, question is whether these results from Experiments 4 and 5 support Barsalou's findings of no priming for the taxonomic categories, over which there has been some controversy (e.g., Greenspan, 1986). As one can see in Tables 7 and 9, any effects of priming were near zero for similarity ratings (Experiment 4) as well as for reaction times and errors (Experiment 5). The corresponding inferential statistics show no evidence of an effect: $F(1, 47) = .19$, $MSe = .378$; $F(1, 36) = .001$, $MSe = .009$; and $F(1, 36) = 1.47$, $MSe = .022$.

Experiment 6: Absolute Inference Judgments

In Experiment 6, we investigated the extent to which people use food categories to make a variety of inferences. In particular, if you know that a food is from a specific category, what do you know about issues such as how much fat it has, how expensive it is, or how much effort it is to prepare it for eating? We chose 12 such inferences and paired them with 14 different categories (6 taxonomic, 8 script), for which subjects had to answer a question and give a probability judgment for their answer.

For example, a subject might be asked whether meats are low, medium, or high in vitamins (compared to all other foods). After choosing one of these responses, he or she would then give a probability for that answer (how probable is that to be the correct answer?). Our interest here was not primarily in the response selected, but in the probability judgment given. For example, suppose that one subject felt that meats in general are low in vitamins. This person should give a high probability for his choice, since any given meat would be likely to be low in vitamins. In contrast, if another subject felt that meats varied greatly in vitamins, she should give a low probability judgment for whichever response she selected, since one could not be sure that any given meat would have that vitamin content. Thus, the probability estimate is an index of the inductive strength of the category with regard to that property. If script categories are low in inductive strength, they should receive reliably lower probability judgments in this task.

We also expected that there might be an interaction, with different inferences being stronger for different types of categories. For example, Heit and Rubenstein (1994) found that subjects used taxonomic categories of animals to answer questions about anatomical similarities but used knowledge about locomotion similarities to answer questions about behavior. That is, different categories may serve different inferential functions. For the present stimuli, it is possible that taxonomic categories are more informative about the content of the foods, their origins, and perhaps their nutritive qualities. It is possible that script categories are more informative about the times and circumstances under which the foods are eaten, their costs, or their cooking methods. To investigate this possibility, we included a variety of different properties in order to discover whether different categories were systematically informative about one or another type of property, as the results of Heit and Rubenstein might suggest.

Method

Subjects. The subjects were 22 undergraduates who received course credit or pay. The experiment took about 30 min.

Materials. We used 14 categories, 6 taxonomic and 8 script categories, given in Table 2 (and used in Experiment 2). We chose 12 inferences that sampled a diverse set of properties. These inferences were later split into two types, as discussed under Results. The inferences are given in Table 10. The 12 inferences each had three possible responses. For example, for *Has fat,* the choices were Low, Medium, or High. For *Is good to eat when depressed,* the

TABLE 10
Inferences Used in Experiment 6

Inferences and choices	Rating: biochemical or situational[a]
1. Gives one energy: Low, Medium, High	Biochemical
2. Costs: Inexpensive, Moderate, Expensive	Situational
3. Is sweet: Low, Medium, High	Biochemical
4. Has vitamins: Low, Medium, High	Biochemical
5. Is eaten in the: Morning, Afternoon, Evening	Situational
6. Is good for you: No, Moderately, Yes	Biochemical
7. Is filling: No, Moderately, Yes	
8. Is good to eat when depressed: No, Moderately, Yes	Situational
9. Calories: Low, Medium, High	Biochemical
10. Has fat: Low, Medium, High	Biochemical
11. Effort to prepare for eating: Low, Medium, High	Situational
12. Has fiber: Low, Medium, High	Biochemical

[a] As explained in the text, a separate group of subjects rated the inferences as biochemical or situational. The only inference that could not be agreed upon was *is filling.*

choices were No, Moderately, or Yes. The inferences were arranged in a random order that was used in all pages (to make it less confusing for the subject). The order was reversed for half of the subjects.

The top of each page began ''Compared to all other foods, a . . .'' followed by a category name (e.g., Breakfast food or Fruit) in boldface. Below that were the 12 inferences with their three possible responses. To the right of each inference was ''Probability'' and a line for writing in the probability. The 14 category pages were randomly ordered for each subject.

Procedure. The instructions informed subjects that the goal of the study was to find out their beliefs about different types of foods. They would be asked to answer the 12 questions by judging one category of foods relative to all other foods. Their task was to circle the best choice for each question and then to give a probability between 0 and 100 to indicate ''how probable such a food type is to have the value that you circled.'' To explain the probability scale, the instructions said that if subjects thought that their answer would be true one-third of the time, they should write 33; if about half the time they should write 50; and if almost certain they were to write a number near 100. An illustration was also given (using finger foods and how often they are served in restaurants). They were asked to answer the questions in the order given and not to go back and change an answer.

Design. All subjects made all 12 inferences about each of the 14 categories. The main manipulation, varied within-subjects, was the type of food category, taxonomic or script. In addition, there was a counterbalancing variable of the order of the inferences with half the subjects getting one order for each category and the other half of the subjects getting the reversed order.

Results and Discussion

The main result of interest is how the judged probability (how probable the subjects thought their answers were) compares for the two different kinds of categories. If subjects believe that knowing the food categories is predictive of a property, then they should give higher probability judgments than if they do not think it is very predictive. There was no overall difference between the category types: For the taxonomic categories, the mean probability judgment (out of 100) was 75.9, while the corresponding mean for the script categories was 76.1, $t(21) = 0.24$ (by items, $t(11) = .14$). Thus, surprisingly, the two types of categories showed equal inferential power for these inferences.

A second means of addressing the issue is to examine subjects' choices, though this involves some additional assumptions. Because subjects were given three choices (e.g., Low, Medium, or High) and asked to compare a given food category to all other foods, we assumed that if the food category provided no information that subjects would select each choice about one-third of the time. (We realize that this is an idealized assumption, which is why we view the probability judgments as the main dependent measure.) We can then ask whether the distribution of answers over the three choices varied from this 33–33–33 pattern for each particular inference for each food category (where the selections are from the 22 subjects). To the degree that the choices differed from this pattern, the category must be providing information about that property. A χ^2 was computed for each inference for each category. Because the χ^2s are not independent, they cannot be added to test the overall pattern, so we simply calculated the proportion of these χ^2s that were statistically significant at the .05 level. For the taxonomic categories, the proportion was .79, and it was .80 for the script categories. These high proportions indicate that subjects believed that the categories were informative for making these inferences. However, again, there is no hint of a difference in the inferential power of these two types of categories. In short, subjects are just as likely to make strong inductions for script categories as for taxonomic categories.

The next analyses examined whether there might be an interaction between the type of category and the type of inference. For example, some of the inferences asked about the biochemical aspects of foods (vitamins, fat, fiber), which we thought might be related to the taxonomic categories, and some inferences seemed more related to the situations in which one eats them (costs, is eaten in the morning, takes effort to prepare for eating). A group of 20 subjects judged whether each inference is more likely to be related to the biochemical composition of the food or the situational aspect of the food. The biochemical makeup of the food was described as the stuff the food is made of. The situational aspect was described in terms of how it is used in our culture, where the use often puts together foods that can have very different biochemical composition. From these ratings we divided the inferences into seven that were clearly biochemical (as chosen by .85 to 1.0 proportions of subjects) and four that were situational (.80 to 1.0 proportions of subjects). The remaining inference (''is filling'') was not rated as clearly biochemical or situational and so was omitted from further analyses. These divisions are indicated in Table 10.

When the data were broken down in this way, there was now some difference between the taxonomic and script categories, though the effects were small. For the biochemical inferences, the respective probabilities were 77.5 and 76.0 and the small advantage for taxonomic categories was not reliable, $t(21) = 1.55$, $p < .15$ (by items, $t(7) = 1.86$, $p < .15$). The proportion of significant deviations from the χ^2s did not show large differences either, .83 to .86 ($t(7) = .35$). However, the situational inferences did show a significant difference, with the mean probability judgment for taxonomic categories, 72.3, smaller than the 75.7 mean for script categories, $t(21) = 2.91$, $p < .01$ (the item analyses for probabilities and the proportion of significant χ^2s showed small effects in the same direction, but were not significant due to only 3 *df*). Most importantly, there is a significant interaction: The 1.5 greater probability for the taxonomic categories in the biochemical inferences is different from the 3.4 lower probability in the situational inferences, $F(1, 21) = 12.34$, $MSe = 10.7$, $p < .01$.

Thus, although the overall results showed that both taxonomic and script categories could lead to inferences, there was an interaction, consistent with Heit and Rubenstein (1994): The taxonomic categories led to slightly higher probability judgments on the biochemical inferences and significantly lower judgments on the situational inferences. These differences, however, were very small. Both types of categories seemed to have much inferential power with both types of inferences (all the probabilities were over .70, far above the .33 that one might think of as chance in this procedure).

Experiment 7: Inference Triplets

Experiment 6 produced two results that will be followed up in this experiment: the equal inferential power of both category types and the small interaction of category and inference type. First, the script categories had considerable inferential power—knowing that a food was a breakfast food, for example, led to as much influence on the judgments and probabilities as did knowing that a food was from a taxonomic category, such as meat. One aspect of the experimental method that may have contributed to such an effect will be examined here. In Experiment 6, subjects gave probability judgments without any specific comparison in mind. That is, they rated the inference of a property based on a category, giving an absolute probability judgment. A more sensitive method may be to make relative judgments in which two categories are compared. This will allow a more direct comparison of the inferential power of taxonomic and script categories.

Second, it may be that the inferential power of the categories depends upon the type of inference being made (e.g., Heit & Rubenstein, 1994; Kalish & Gelman, 1992). That is, as we suggested earlier, the taxonomic categories may be especially useful for some types of inferences (e.g., biochemical) and script categories for other types of inferences (e.g., situational). There was evidence of such an interaction in Experiment 6, but it was small. In this experiment we used specific food items (e.g., bagel) rather than categories (e.g., breads and grains or breakfast foods) and pitted the two types of food categories against each other as in past work on category-based induction (e.g., Gelman & Markman, 1986), so any such interaction may be more likely to be observed.

For this experiment, we constructed triplets of foods and two types of inferences. The food triplets consisted of a target food, a taxonomic alternative (a food from the same taxonomic category), and a script alternative (a food from the same script category as the target food). For example, if the target food was cereal, then a taxonomic alternative might be noodles (since they are both breads and grains) and a script alternative might be milk (since they are both breakfast foods). There were two types of inferences: biochemical and situational. For the biochemical inference, subjects were told that there was an enzyme, metacascal, which had been found in the target food in a foreign country, and they were asked which food was more likely to contain metacascal, the taxonomic alternative or the script alternative. The situational inference questions claimed that the target food was eaten at the annual initiation ceremony in that country and asked which food would be more likely to also be eaten at the ceremony, the taxonomic alternative or the script alternative. In this circumstance (with the category types pitted against each other), we expected the taxonomic alternative to be chosen more often for the biochemical inference and the script alternative to be chosen more often for the situational inference.

However, other results are possible as well. For example, given that taxonomic categories seem to be the most salient ones (as shown in the default sortings of Experiment 3), it is possible that they will be the winners when pitted against the script categories. As mentioned earlier, a form of cognitive economy might predict that most inferences are made through a single most salient categorization of an item, rather than through multiple categories for every item. However, other research on multiple categorization of people (Nelson & Miller, 1995) has suggested that the most distinctive category is the source of inductive inference when there are two or more categories available. If taxonomic categories like meat and vegetables are the most common ones used for identifying items (as suggested by the sorting data), then script categories may be viewed as more unusual and distinctive. If so, they may dominate the inferences. The present experiment examines these possibilities by contrasting the two kinds of categories for different inference properties.

Method

Subjects. The subjects were 48 undergraduates who received course credit or pay. The experiment took about 25 min.

Materials. We constructed 20 triplets consisting of a target food, a taxonomic alternative, and a script alternative, as described above. The triplets are shown in Table 11.

The two inferences were as described above. For the biochemical inference, each question

TABLE 11
Food Triplets Used in Experiment 7

Target	Taxonomic alternative	Script alternative
Bagel	Cracker	Egg
Muffin	Spaghetti	Yogurt
Rice	Cereal	Potato
Hamburger	Bacon	Pizza
Butter	Yogurt	Bread
Cookie	Biscuit	Ice cream
Sausage	Steak	Waffle
Cereal	Noodles	Milk
Stuffing	Muffin	Turkey
Onions	Carrot	Hamburger
Custard	Cheese	Cake
Caviar	Trout	Champagne
Tuna	Swordfish	Mayonnaise
Raisins	Oranges	Nuts
Croutons	Bagel	Lettuce
Noodles	Bread	Gravy
Salami	Salmon	Mustard
Tortilla	Bagel	Beans
Chicken	Lobster	Lasagna
Milk	Sour cream	Cookie

stated: ‘‘Suppose that an enzyme, metacascal, had been found in [target food] in the country Quain. What food is more likely to contain metacascal:’’ and then the taxonomic and script alternatives were given. For the situational inference, the question read: ‘‘Suppose that [target food] is eaten at the annual initiation ceremony in the country Quain. What food is more likely to be eaten at that ceremony:’’ again followed by the two alternatives.

Each page consisted of 10 biochemical inference questions or 10 situational inference questions. Next to the alternatives was the word ‘‘Confidence’’ and a line for subjects to write their confidence in their selection.

Each booklet contained the instructions and then four pages: two pages of the biochemical inference questions followed by two pages of situational inference questions or the reverse. There were four orders of pages (i.e., the order of the two pages of each type was counterbalanced).

Procedure. The instructions informed subjects that the goal of the study was to find out their beliefs about different types of foods. They were to imagine that they were in a foreign country they had never been to and about which they knew nothing. They would be told a fictional fact about a food in this country, Quain, and they were to pick which food was more likely to share this property. Their task was to circle the best choice for each question and then to give a confidence rating in their choice between 0 (guessing) and 100 (certain). It was stressed that there was no right answer. This was followed by an illustration using the enzyme question. Subjects were told to answer the questions in order, not to go back, and to treat each question separately, as if the information given in that question was all they knew about foods in Quain.

Design. The main manipulation, varied within subjects, was the type of inference, biochemical (enzyme) or situational (initiation ceremony). As mentioned, the inference orders were counterbalanced.

Results and Discussion

The main result of interest is whether the food choice (taxonomic or script alternative) depends upon the inference question (biochemical or situational). It does. For the biochemical questions about enzymes, .83 of the responses were the taxonomic alternative, which is significantly different from chance, $t(47) = 12.03$, $p < .001$, while for the situational question about the initiation ceremony, .71 of the choices were for the script alternative, $t(47) = 6.06$, $p < .001$. The item analyses showed similar effects, both for the biochemical inference questions, $t(19) = 13.36$ (all 20 items leading to higher responses for the taxonomic alternative), and the situational inference question, $t(19) = 6.38$ (17 of 20 items leading to higher responses for the script alternative).

The results of Experiments 6 and 7 indicate that both taxonomic and script categories allow many inferences to be made, but also that people are sensitive to the different types of inferences that each best allows. In this study, the taxonomic alternatives were viewed as better choices for the biochemical inferences (related to an enzyme in the food), whereas the script alternatives were chosen more often for the situation inferences (related to an initiation ceremony). However, it is also clear that script categories are thought to be quite useful for making inferences: In Experiment 6, subjects were just as confident in the inferences they drew for the script categories as they were for the ones they made for the taxonomic categories.

GENERAL DISCUSSION

The goal of this project was to investigate complex conceptual structures that involve cross-classification and to examine how this cross-classification is represented and used in access and inferences. We chose food as the domain for this investigation because of its rich categorical structure and because it is a domain that people interact with daily and have considerable knowledge about. We asked which categories people use for thinking about food, and we found evidence for both taxonomic and script-based categories, which are quite different ways of organizing information about food. We also asked whether people access and use both kinds of categories and found that they do. However, the extent to which the two types of categories were used did seem to differ: The taxonomic categories tended to be preferred as the neutral organization and were more accessible as well. Here we address the implications of these findings for understanding conceptual structure and how cross-classification affects category use.

The strategy of the project was to examine three main aspects of category knowledge—representation, accessibility, and inference—and to examine each with multiple measures. We begin by summarizing the results related to cross-classifications and then consider cross-classifications in conceptual organizations, nonclassification functions of categories, and the implications for current views of categories.

Summary of Main Results

The idea that an item can belong to a number of different (nonhierarchical) categories simultaneously has been often mentioned, but not extensively investigated. Barsalou (1983, 1985, 1991) has convincingly shown that people can spontaneously form categories that cut across preexisting categories to address particular goals. These goal-derived categories do have graded structure, but their basis for typicality is an extreme ideal, not an average prototype (e.g., for foods to eat when on a diet, the ideal would be a food with zero calories). Because goal-derived categories are constructed as needed, they do not usually have a permanent representation in memory and so are not as interesting from the perspective of conceptual organization. Barsalou (1985) showed that well-established taxonomic categories were likely to be based on a prototype but that their typicality structures were also affected by ideals, hinting at possible multiple categorization. Medin et al. (1997) showed that a great deal of experience using a category may lead to an alternative organization that emphasizes this use, such as the landscapers developing an organization based on landscaping roles (e.g., shade trees, ornamental trees). They argued that the landscapers' organization is a special-purpose one that is used in inference tasks only to support inferences directly related to this special purpose. When asked to make biological inferences, these subjects did not use their landscaping categories but fell back on taxonomic categories (e.g., maples; see also Coley et al., 1997). The results of the current studies suggest that the cross-classifications of foods are common, quickly available from the food names, and may support a variety of inferences. Unlike Medin et al.'s data, our results show that subjects can use two different conceptual organizations to classify items *and* as a basis for inference.

Experiments 1–3 provided a clear picture of the importance of script categories. The category generation results of Experiment 1 showed that script categories were generated about as often as taxonomic categories were. The category ratings in Experiment 2 confirmed that people believe that food items are members of particular script categories. In addition, the distribution of these ratings was very different than that in taxonomic categories both because there were far more ratings in the middle range (so that items were not just either very good members or nonmembers) and because many items belonged to more than one script category. The sorting results of Experiment 3 suggest that both taxonomic and script categories influence how people sort food items as well. Most importantly, when subjects were given no instructions about sortings, the script categories influenced their sorts. The taxonomic categories did dominate, but there were clear cases of script groupings, even for subjects given taxonomic instructions. Thus, people believe that items are in both script and taxonomic categories and their sortings of foods are influenced by these categories. Also, some items (e.g., rice) appeared to span two categories in a single scaling solution, indicating that they are viewed as good members of both categories.

Experiments 4 and 5 found that taxonomic categories seem to be spontaneously activated when food items are encountered. For example, when subjects viewed a pair like banana–orange, or a single item like banana, they thought of the category of fruit. This was indicated by the fact that providing the category name had no effect on similarity ratings, and a helpful prime had no effect on category verification for these items. In contrast, ad hoc categories, which are not thought to be spontaneously evoked by an item, showed large effects of such priming. The script categories showed effects in the middle of these two, usually significantly different from the ad hoc categories. This pattern of results suggests that these categories are indeed spontaneously activated by presentation of the food item. However, this activation is not as strong or consistent as that of the taxonomic categories. Nonetheless, the results suggest that cross-classified items do actively evoke multiple classifications, which are then used in judgments.

Experiments 6 and 7 showed that both taxonomic and script categories are used in making inferences, though they are especially used for inferences related to their category types. In Experiment 6, both taxonomic and script categories were used to make inferences about various properties of foods (e.g., how much energy it would give), though there was a suggestion that the type of property made a difference in which type of category was most influential. Experiment 7 confirmed this interaction between category type and inference type—different category organizations may be used for making different types of inferences. However, across the two studies, it is impressive that both kinds of inferences can be made for both kinds of categories. That is, although subjects preferred to use script categories to make event-type inferences and food types to make biochemical inferences when forced to choose between them, Experiment 6 showed that subjects feel quite confident in making both kinds of inference based on both categories. This suggests that the information associated with the two organizations is not strictly segregated, as we will discuss below.

The results of all of these experiments provide strong evidence for the importance of script categories in people's representation of and inferences about food. This alternative organization appears to be a common one that affects a variety of category-related tasks. In the next section, we consider the importance of these script categories, but a note should also be made about the separation of taxonomic and script categories.

Clearly, foods are a messy domain (no pun intended). In particular, there are bound to be correlations between the category types. For example, most vegetables are healthy foods and few are breakfast foods, junk foods, or snacks. Thus, in some cases one might well attribute the influences of one category type to another (see Medin et al., 1997, for another example). For instance, when interpreting the sorting data in the default condition, we attributed groupings first to the taxonomic categories and invoked script categories only when the results were not consistent with the taxonomic categories. We do think that this was a reasonable way of interpreting the data, but this conservative procedure means that there may be some unidentified influences of script categories in these data. For example, bread, bagel, oatmeal, cereal, muffin, and pancake form a cluster in the first default sorting solution (see Fig. 1), probably because these are *both* grain-based foods and breakfast foods. Also, it may be that script categories reinforce the taxonomic divisions, making them stronger. If people tend to eat breads and grains at breakfast, this fact becomes a property of the taxonomic category. The groupings of foods into taxonomic and script types are clearly not completely independent. This may be part of the reason that subjects are quite confident in making inferences about both event-related and biochemical properties from both types of categories (Experiment 6). In short, our methods of analysis may be underestimating some script category influences.

On the other hand, one might argue that at times we interpreted taxonomic influences as script influences or that some of the script groupings may have been due to a salient property (e.g., saltiness or fattiness for snacks) rather than truly situation-based. However, we do not think that this would account for the observed effect of script categories. First, as mentioned, the sorting data were interpreted initially in terms of the taxonomic categories, with the script categories being used to interpret the remaining pattern. Second, script categories were often generated from the food items in Experiment 1, and Experiments 4 and 5 showed that the script category names appear to be activated by the food item names. Thus, the items appear to be related by category membership, not just by individual properties.

Cross-Classification in Conceptual Organization

Thus, the results provide support for the importance of both script and taxonomic categories in representation, access, and inferences. What is the nature of these two category types? Although we cannot offer a complete answer at this stage, we suggest the following. Taxonomic categories appear to be more oriented toward intrinsic properties of the foods, including their origins, composition, nutritional values, and so on. For example, meats come from animal sources and tend to be high in protein and fat. As mentioned earlier, these categories also approximately divide items by their macronutrient profiles (except for beverages). We are not suggesting, however, that these categories do not include functional properties or event-related information. Meats tend to be part of the main course of a meal, for example, in part because of their protein content. The power of taxonomic categories is that they bring together many different properties, which allow many different kinds of inferences (Markman & Callanan, 1983). Our suggestion is that the taxonomic categories are structured around the intrinsic properties of the food itself, rather than how it relates to other activities or events, as the script categories are.

The taxonomic categories subjects produced in Experiment 1 do not appear to be ‘‘natural kind’’ categories of the sort proposed by Putnam (1975; Malt, 1994), as oak, gold, or shark are. That is, although they capture regularities about the macronutrient profiles of the foods, categories like meat, vegetable, breads and grains, and beverages are not, in general, categories given to us by nature. They depend on human-defined properties and conventions as well as on their purely natural qualities. For example, vegetables do not correspond to any particular botanical class and include items that are botanically considered fruits (like tomatoes and zucchini). Thus in saying that it is the intrinsic properties of the foods that largely determine their taxonomic category membership, we are not saying that these categories are given by nature or that they have biochemical essences. We are simply contrasting categories that are largely defined by their origins, textures, tastes, and biochemical content with those that are determined more directly by human events and activities.

We have already discussed script categories in some detail throughout the paper. Our proposal here is that such categories derive largely from their place in more externally defined activities and schemata. Desserts, for example, are largely defined by their place as a sweet dish at the end of a meal. Although this clearly includes some intrinsic properties (primarily sweetness), a wide variety of different foods can fit this schema (e.g., pies, ice cream, fruit, cookies, candy, cheese). Ice cream and fruit clearly have little in common in their origins, taste, biochemical makeup, texture, and so on, except for the sugar that makes them sweet. However, a category like desserts can still be useful in drawing inferences and in planning (as we discuss below), because we know what kinds of things are conventionally served as desserts and what properties make them particularly suitable as desserts.

Given these two rather different ways of categorizing foods, the critical question is how this cross-classification is mentally represented. It seems clear that these are different kinds of organization, because the two specific sorting solutions (taxonomic and script) deviate in a number of important respects and are only correlated .54 overall. Thus, these could not be embedded within a single hierarchy. Rather, it seems necessary to conclude that different clusters of items exist simultaneously, presumably through different category links. For example, bagel may be connected to breads and grains as well as to breakfast foods and sandwich foods. There does not seem to be any kind of contradiction in one item being fairly strongly connected to multiple categories, as, for example, hamburger was rated as being an extremely good example of both meats and lunch foods. It is probably a mistake to think of a category as consisting of a cluster of items in semantic space, as there is no way to cluster hamburger and salmon together in some cases (as meats) and separately in other cases (one as a lunch food and one as a dinner food). By the same token, it is simplistic to think of every item as having one ''real'' categorization, as other categories are readily accessed and used.

Thus, the results suggest a kind of nonhierarchical network of category relations, in which items are connected to all the categories they exemplify. Items are related to other items by shared category membership (e.g., eggs and toast are both breakfast foods) and/or by shared properties (e.g., toast and muffins both are made from wheat, have carbohydrates, are baked, etc.). How does one decide in such a system how to categorize an item? Because this sort of network is less highly structured than a pure taxonomic one (see Hampton, 1982; Murphy & Lassaline, 1997), there is no single entry point—instead, category access must be determined in part through goals and contexts. So, if coffee and cereal are seen together, breakfast foods quickly come to mind, but if coffee and beer are seen together, beverages come to mind. Thus, the contexts of other foods, the time of day, the setting, or other cultural indicators can all determine which category is activated for a given food. Furthermore, as discussed below, the goal of the activity that one is involved in may activate a way of conceiving of the food. Our results suggest that the script categories are represented in memory to a large degree, but that they may also be constructed in some circumstances. For example, one may not have a well-established category of foods eaten at the movies, but one can easily construct such a category post hoc, including popcorn, soda, certain candies, and ice cream. If one often eats at the movies, this information may become more and more saliently represented for these items, until it can be as important a way of representing them as their taxonomic categories (see Barsalou, 1991).

In short, the long-term representation of cross-classified categories may not be particularly neat. This is not to deny that some parts of the categorical system could be organized hierarchically. However, there are a number of other ways of identifying objects that simultaneously exist, with varying strengths. Although such a representation may lose some of the advantages of hierarchical systems (e.g., Markman & Callanan, 1983), it may facilitate some nonclassification functions of categories, as discussed next.

Category Functions beyond Classification

Most research on categories focuses on classification, in which items are presented, and the question is whether they are members of a particular category. Although the item-to-category path is a crucial part of the representation of categories, the focus on it may leave other interesting questions about categories relatively unexplored. Some research has examined how categories are involved in cognitive uses after classification. Much of this has focused on how categories can be used for induction (e.g., Gelman & Markman, 1986; Osherson et al., 1990) even when the classification is uncertain (e.g., Anderson, 1991; Malt et al., 1995; Murphy & Ross, 1994; Ross & Murphy, 1996). Other research has examined how such classifications are crucial for a given task, such as problem solving (e.g., Blessing & Ross, 1996; Chi et al., 1981).

Against this background, it is interesting to speculate about script categories and how they might be used. Our intuitions are that the primary purpose of such categories is not for classifying particular food items. Rather, such categories may often serve to generate particular food items, such as ‘‘What would you like for breakfast?’’, ‘‘What can I make in half an hour?’’, or ‘‘What can I have as a snack at the movies?’’ Thus, one often begins with the script category and generates exemplars in the service of some goal.

We consider this case to be an illustration of a general function of categories, which is to allow the generation of category instances that might help accomplish a plan or meet a goal. Barsalou’s (1991) work on goal-derived categories (e.g., foods to eat on a diet, places to go on vacation) demonstrates this use in categories that were spontaneously generated or at least not very well established in memory. However, many script categories are very well established categories that serve these purposes over and over again (foods to eat at breakfast time, foods to have for a snack). Thus the importance of categories in planning is not confined to novel categories.

In contrast, food taxonomic categories are probably less directly related to a particular goal or activity. Categories such as vegetable, meat, and breads and grains relate to a number of aspects of a food: its origin, its nutritive content, and so on. This kind of information may be relevant in a wide variety of situations, though it is probably also more useful in some settings than in others (e.g., it is more important to know food types in planning a balanced meal than in deciding what to eat as a snack between classes). In particular, dietary recommendations are often made by prescribing food types, for example, the ‘‘four basic food groups’’ of the 1950s and 1960s, or the ‘‘food pyramid’’ of the 1990s, which give recommendations in terms of ‘‘meats’’ and ‘‘grains’’ rather than ‘‘breakfast foods’’ (e.g., Hertzler & Anderson, 1974; USDA, 1992). Because food types are so directly related to the nutritive content of foods, that is one primary way in which these taxonomic categories are useful. However, it is also the case that many popular food categories do not map directly onto nutritive categories, and so recommendations are sometimes made for less familiar categories or subcategories of foods, such as legumes or cruciferous vegetables. It is not surprising that intuitive categories of foods do not directly reflect the foods' biochemical makeups or health consequences (which is another sign that these categories are not natural kinds in the Putnam sense).

There are three points that relate to these considerations. First, while we have focused on a particular alternative organization for foods, it is important to realize that different uses may lead to a number of different organizations. For example, if one wanted a diet high in protein, one might consider a macronutrient organization of foods including proteins, carbohydrates, and fats. Even within our script categories, it is possible that the healthy foods/junk foods categories represent an alternative organization in which foods are characterized by their effects on health (e.g,. which might include fattening, stomach-upsetting, etc.). A consistent use of a set of items may lead to establishing new category representations (see Barsalou & Ross, 1986).

Second, compared to the mutual exclusivity of taxonomic categories, there is considerable overlap of items in script categories: As seen in Experiments 1 and 2, many items belong to multiple script categories—and at the same time belong to taxonomic categories. Speculatively, this difference with taxonomic categories may be partially related to the classification-generation distinction as well. The difference in overlap between the categories may reflect different ways in which the categories are used. In classification of an item (with no additional context), one might often want to access information that is likely to be useful across a wide variety of situations, and the taxonomic categories may provide just that kind of information. Thus it may be most useful, in the absence of a constraining context, to have the nearly mutually exclusive taxonomic categories dominate. (Recall that foods were only rated as very good members of one taxonomic category in Experiment 2.) However, if the goal is to use the category to generate exemplars during planning, then it may be more useful to have many exemplars in each category, which would lead to items being represented in different categories. Such a representation might lead to some interference for accessing category names of an exemplar, but these may be cases in which the taxonomic categories dominate, or the context might help to select the appropriate one. It is clear (e.g., Experiments 4 and 5) that people can access script categories from the food names, but not as easily as from the taxonomic categories.

Third, this examination of script categories questions the focus in the field on classification. In many situations, the goal is not simply to classify. For example, math problems are classified because the classification helps to access relevant knowledge on how to solve the problems. A situation or person may be classified in order to make some prediction about what is likely to happen (e.g., Murphy & Ross, 1994; Malt et al., 1995). We are not saying that classification is not important, but rather that researchers need to consider other functions of categories (e.g., prediction, problem solving, planning, explanation, communication) in order to understand why concepts have the structure they do and that we might think about classification as one part of a larger goal-oriented task (see Markman et al., 1997; Ross, 1996a, b, 1997, in press-a, b; Yamauchi & Markman, 1998).

Directions for Future Research on Conceptual Structures

The prototypical category-learning experiment is one in which subjects learn two categories at the same level of abstraction and of the same type. Although these studies are useful in a number of respects, the present results suggest that the image of category structure that comes from such experiments may not be very representative of conceptual structures in real, complex domains. The present results have provided some understanding of how multiple categorizations operate in a rich domain, but the results also suggest a number of new issues that need to be addressed by further research.

One issue has to do with levels of categorization. Our investigation treated basic level categories as items and studied their organization into higher level categories more like superordinate categories in the object domain (Murphy & Brownell, 1985; Rosch et al., 1976; Murphy & Lassaline, 1997). For example, apple is a basic level category, and we used apple as a food item and examined its classification as a fruit and snack food. One question this raises is to what degree cross-classification exists at lower levels of categorization. Clearly, foods can be identified at a number of lower levels (e.g., as apple, cooking apple, Granny Smith apple, etc.), but it is not known whether such alternative categories are as salient as the script and taxonomic categories we investigated. Similarly, it is not known whether the basic categorization can be overruled by goal-directed categories (e.g., whether one could see an apple as a snack food more readily than as an apple in some situations). Work on more usual object categories suggests that the basic level advantage is very difficult to eliminate by context and task demands (Lin, Murphy, & Shoben, 1997).

Second, a number of issues are raised by considering the access of categories in multiple organizations. Are multiple categories simultaneously accessed when an item is experienced? It might seem useful to have access to multiple category representations, but the utility may depend on whether one is able to effectively use information from multiple categories. At least some work on stereotyping in social cognition suggests that classifying such cross-classifiable items does not lead to all the information becoming available. For example, Macrae, Bodenhausen, and Milne (1995) found that priming one categorization (e.g., Chinese for a Chinese woman) appears to inhibit the other categorization (e.g., woman). Our experiments showed that either script or taxonomic categories can be accessed and used for induction, but we did not examine whether such categories are simultaneously evoked and used. Perhaps one type of category (e.g., taxonomic) is accessed earlier than other types (e.g., script), which is similar to a suggestion by Barsalou (1991). Clearly, much remains to be learned about how such knowledge is accessed.

Third, how are multiple category organizations used to make various types of inferences? Medin et al. (1997) suggest that extensive experience in a domain can lead to utilitarian representations that may be used to respond to inferences related to this use (e.g., Heit & Rubenstein, 1994). Nelson and Miller (1995) suggest that the distinctiveness of the different categories determines which is used in making inferences. For instance, they found that the inferences about multiply categorized people, such as a dog owner who is also a skydiver, is largely determined by information about other skydivers rather than information about other dog owners. In the food categories considered here, it is not clear which organization is the more distinctive. In Experiment 7, neither type dominated in an inference task; the preferred category depended on the content of the inference, though the taxonomic categories clearly dominated the default sorting data. We are currently examining how people use information from both category types in making inferences (Murphy & Ross, in press). For example, if someone has information about some properties of breads and grains and breakfast foods, how will these be combined in making an inference about a bagel?

In short, our results have suggested a number of new questions on how multiple categories are organized and used. If these questions are to be answered, it will be necessary to investigate richer domains with more possibilities for cross-classification than most psychological research on concepts has generally done.

Conclusions

The seven studies reported here investigate a rich real-world domain, foods. The results suggest that people have multiple organizations of this category, and the studies focused on the taxonomic and script organizations. The script categories are accessed from the item, though not as strongly or consistently as the taxonomic categories. In addition, script categories are used for making inferences, though Experiment 7 suggests that the taxonomic and script categories may each be applied most readily to different types of inference situations. We speculate that such script categories may be particularly useful for nonclassification functions of categories and that such functions comprise an important, but relatively neglected, area of research in the cognition of categories.

APPENDIX A

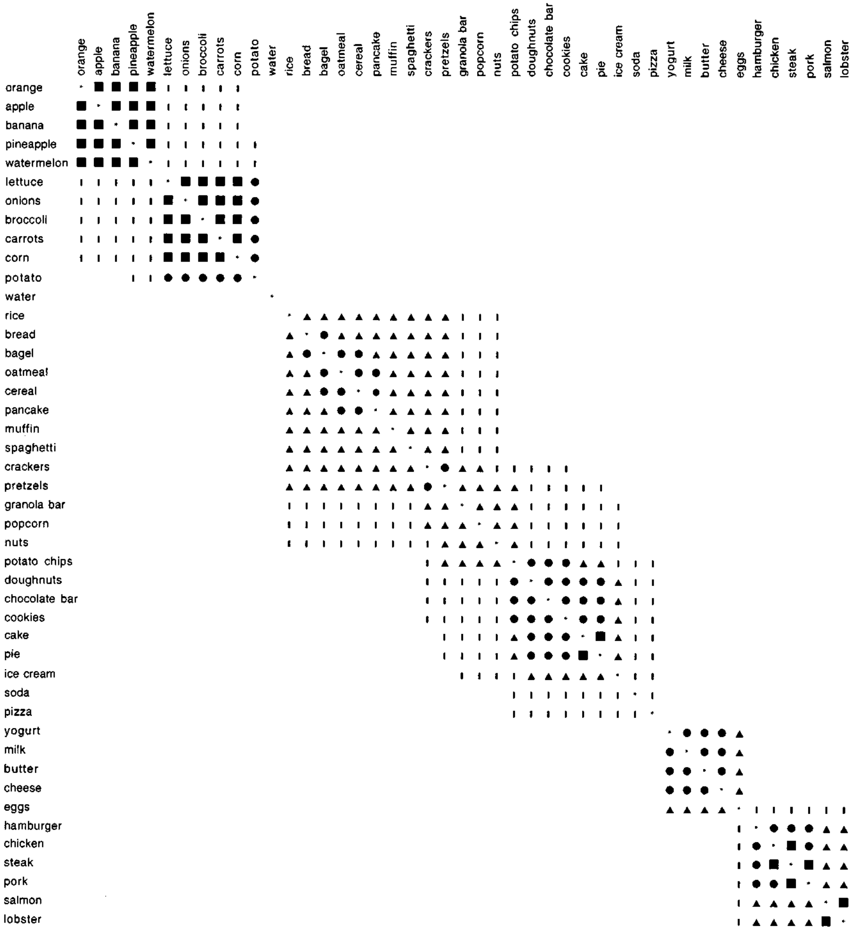

FIG. 3. Primary Robinson matrix of food items for Taxonomic sorting instructions in Experiment 3.

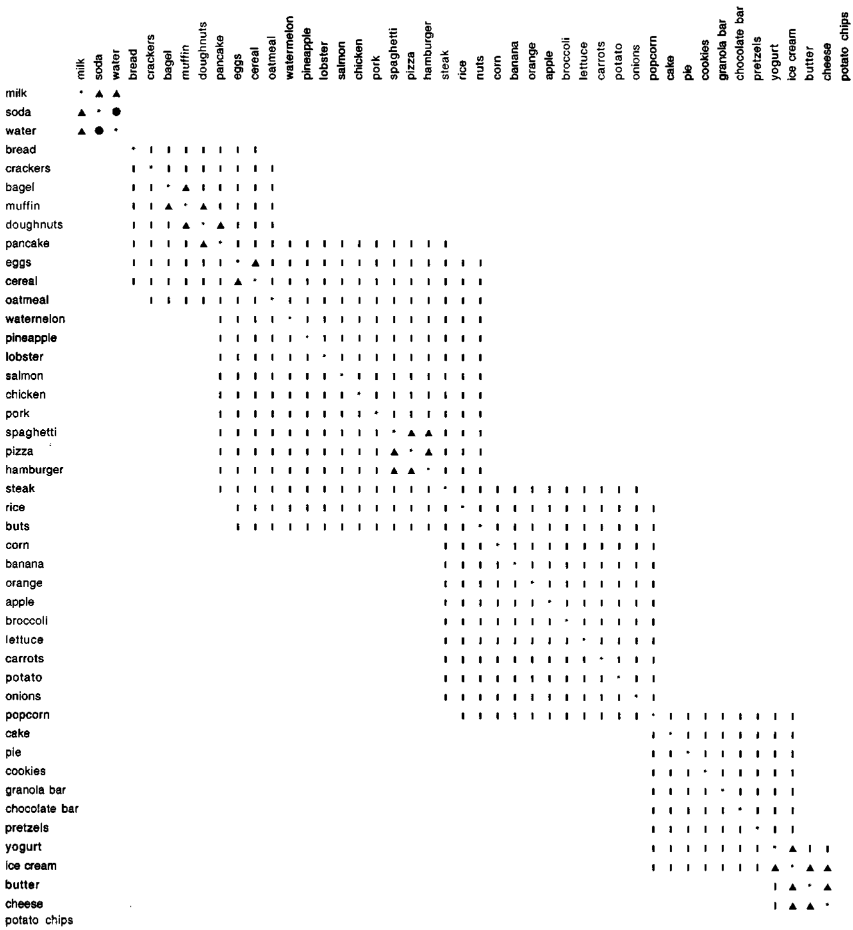

FIG. 4. Secondary Robinson matrix of food items for Taxonomic sorting instructions in Experiment 3.

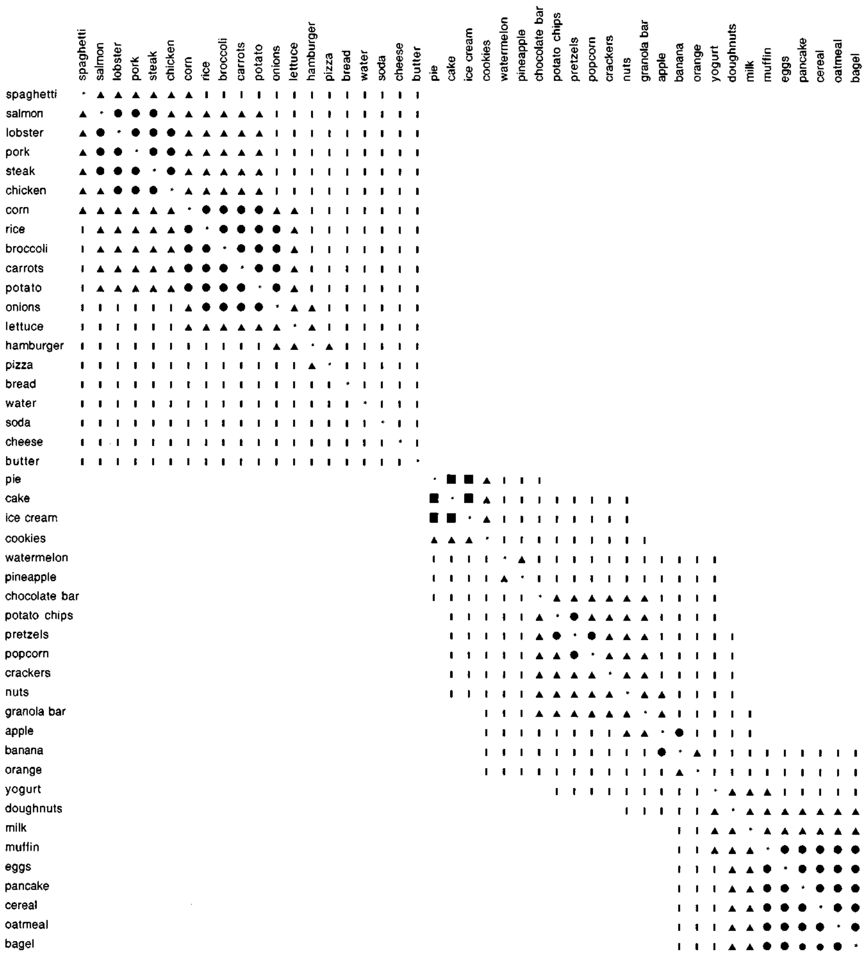

FIG. 5. Primary Robinson matrix of food items for Script sorting instructions in Experiment 3.

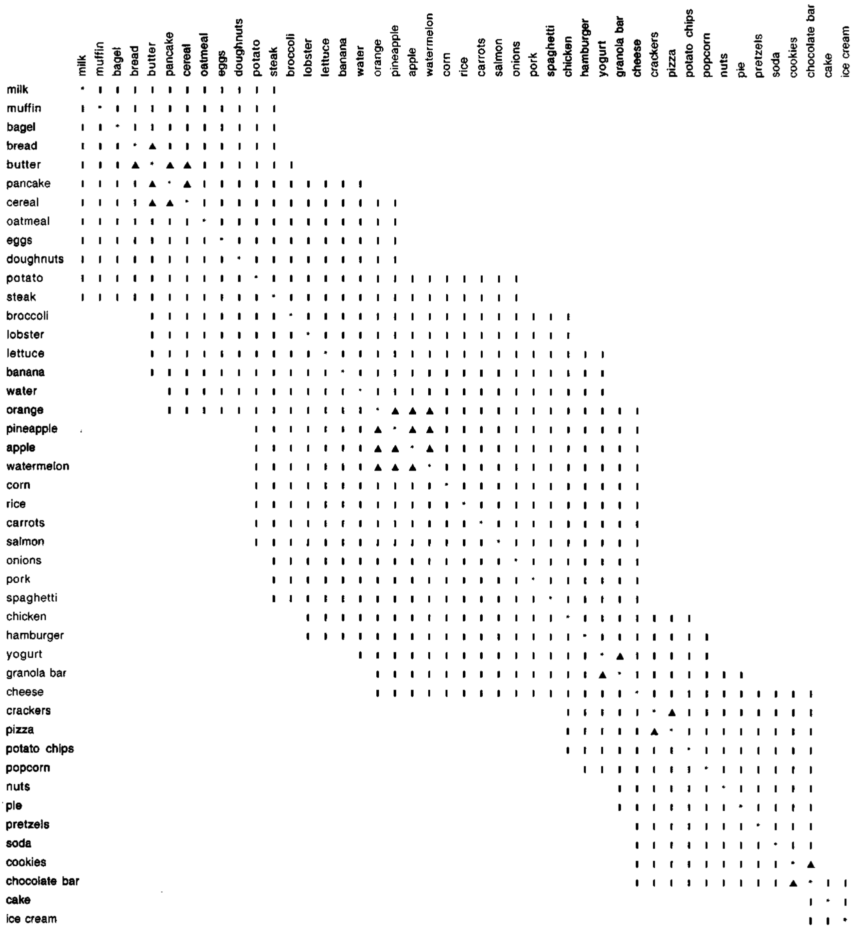

FIG. 6. Secondary Robinson matrix of food items for Script sorting instructions in Experiment 3.

APPENDIX B
Categories and Food Item Pairs Used in Experiment 4

Category	Food pairs	
Taxonomic		
beverages	soda	tea
beverages	water	milk
breads and grains	rice	bagel
breads and grains	cracker	cereal
dairy foods	yogurt	cheese
dairy foods	milk	butter
fruits	pineapple	cantaloupe
fruits	watermelon	strawberry
meats	chicken	bacon
meats	pork	salmon
vegetables	onion	carrot
vegetables	lettuce	potato
Script		
breakfast foods	egg	waffle
breakfast foods	bagel	bacon
desserts	ice cream	cookie
desserts	pudding	pie
healthy foods	banana	chicken
healthy foods	apple	broccoli
junk foods	pie	chocolate bar
junk foods	ice cream	potato chip
lunch foods	hamburger	soup
lunch foods	sandwich	pizza
snack foods	apple	pretzel
snack foods	nuts	cookie
Ad hoc		
foods that are often cooked in water	spaghetti	broccoli
foods that are often cooked in water	corn	oatmeal
foods that go bad quickly if unrefrigerated	milk	pork
foods that go bad quickly if unrefrigerated	yogurt	fish
foods that squash easily	tomato	marshmallow
foods that squash easily	pie	banana
foods that you can carry in a paper bag	nuts	apple
foods that you can carry in a paper bag	popcorn	orange
foods you can throw far	apple	egg
foods you can throw far	potato	orange
foods you eat with a spoon	cereal	pudding
foods you eat with a spoon	soup	yogurt

REFERENCES

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review,* **98,** 409–429.

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive Psychology,* **31,** 219–247.

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

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition,* **11,** 211–227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and* Cognition, **11,** 629–649.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1–64). New York: Academic Press.

Barsalou, L. W., & Ross, B. H. (1986). The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **12,** 116–134.

Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **22,** 792–810.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General,* **120,** 278–287.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science,* **5,** 121–152.

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

Cohen, J., MacWhinney, B., Flatt , M., & Provost, J. (1993). Psyscope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers,* **25,** 257–271.

Coley, J. D., Medin, D. L., & Atran, S. (1997). Does rank have its privilege? Inductive inference within folkbiological taxonomies. *Cognition,* **64,** 73–112.

Corter, J. E. (1982). ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation,* **14,** 353–354.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology,* **20,** 65–95.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in children. *Cognition,* **23,** 183–209.

Greenspan, S. L. (1986). Semantic flexibility and referential specificity of concrete nouns. *Journal of Memory and Language,* **25,** 539–557.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior,* **18,** 441–461.

Hampton, J. A. (1982). A demonstration of intransitivity in natural categories. *Cognition,* **12,** 151–164.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory,and Cognition,* **20,** 411–422.

Hertzler, A. A., & Anderson, H. I. (1974). Food guides in the United States. *Journal of the American Dietetic Association,* **64,** 19–28.

Hubert, L., & Arabie, P. (1994). The analysis of proximity matrices through sums of matrices having (anti-)Robinson forms. *British Journal of Mathematical and Statistical Psychology,* **47,** 1–40.

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General,* **126,** 248–277.

Johnson, K. E., & Mervis, C. B. (1998). Impact of intuitive theories on feature recruitment throughout the continuum of expertise. *Memory & Cognition,* **26,** 382–401.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development,* **63,** 1536–1557.

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

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review,* **99,** 22–44.

Lassaline, M. E., Wisniewski, E. J., & Medin, D. L. (1992). Basic levels in artificial and natural categories: Are all basic levels created equal? In B. Burns (Ed.), *Percepts, concepts and categories* (pp. 327–378). Amsterdam: Elsevier.

Lin, E. L. (1996). *Thematic relations in adults' concepts and categorization.* Univ. of Illinois, Urbana-Champaign. [unpublished doctoral dissertation]

Lin, E. L., Murphy, G. L., & Shoben, E. J. (1997). The effects of prior processing episodes on basic-level superiority. *The Quarterly Journal of Experimental Psychology,* **50A,** 25–48.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology,* **32,** 251–295.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology,* **69,** 397–407.

Malt, B. C. (1994). Water is not H_2O. *Cognitive Psychology,* **27,** 41–70.

Malt, B. C. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology,* **29,** 85–148.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language,* **31,** 195–217.

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Making predictions using uncertain natural categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 646–661.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior,* **23,** 250–269.

Markman, A. B., Yamauchi, T., & Makin, V. S. (1997). The creation of new concepts: A multifaceted approach to category learning. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes.* Washington, DC: Am. Psychol. Assoc.

Markman, E. M., & Callanan, M. A. (1983). An analysis of hierarchical classification. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2). Hillsdale, NJ: Erlbaum.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology,* **32,** 49–96.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review,* **85,** 207–238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology,* **35,** 113–138.

Murphy, G. L. (1991). Meaning and concepts. In P. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 11–35). Hillsdale, NJ: Erlbaum.

Murphy, G. L. (1993). A rational theory of concepts. In G. V. Nakamura, R. M. Taraban, & D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 29: Categorization by humans and machines* (pp. 327–359). New York: Academic Press.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **11,** 70–84.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 93–131). London: Psychology Press.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology,* **27,** 148–193.

Murphy, G. L., & Ross, B. H. (in press). Inductions with cross-classified categories. *Memory & Cognition.*

Nelson, K. (1996). *Language in cognitive development: Emergence of the mediated mind.* New York: Cambridge Univ. Press.

Nelson, L. J., & Miller, D. T. (1995). The distinctiveness effect in social categorization: You are what makes you unusual. *Psychological Science,* **6,** 246–249.

Nosofsky, R. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 54–65.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review,* **97,** 185–200.

Putnam, H. (1975). The meaning of ''meaning.'' In *Mind, language and reality, Philosophical Papers, Vol. 2* (pp. 215–271). Cambridge: Cambridge Univ. Press.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge: Cambridge Univ. Press.

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

Ross, B. H. (1996a). Category representations and the effects of interacting with instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **22,** 1249–1265.

Ross, B. H. (1996b). Category learning as problem solving. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 35, pp. 165–192). San Diego, CA: Academic Press.

Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language,* **37,** 240–267.

Ross, B. H. (in press-a). Post-classification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Ross, B. H. (in press-b). The effects of later learning on classification: Category use and subclassification. *Memory & Cognition.*

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: The influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **22,** 736–753.

Ross, B. H., & Spalding, T. L. (1994) Concepts and categories. In R. Sternberg, (Ed.), *Handbook of perception and cognition, Vol. 12. Thinking and problem solving* (pp. 119–148). San Diego, CA: Academic Press.

Rozin, P., Dow, S., Moscovitch, M., & Rajaram, S. (1998). What causes humans to begin and end a meal? A role for memory for what has been eaten, as evidenced by a study of multiple meal eating in amnesic patients. *Psychological Science,* **9,** 392–396.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **8,** 484–494.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard Univ. Press.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition,* **22,** 377–386.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology,* **71,** 888–898.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology,* **23,** 457–482.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General,* **113,** 169–193.

U.S. Department of Agriculture, Human Nutrition Information Service (August 1992). Leaflet No. 572.

Wardlaw, G. H., & Insel, P.M. (1990). *Perspectives in nutrition.* St. Louis: Times Mirror/ Mosby College Publishing.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language,* **39,** 124–148.

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition,* **8,** 161–185.

Accepted November 29, 1998