Journal of Memory and Language **40,** 230–262 (1999)
Article ID jmla.1998.2593, available online at http://www.idealibrary.com on IDEAL®

# Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts

Barbara C. Malt

*Lehigh University*

Steven A. Sloman and Silvia Gennari

*Brown University*

and

Meiyi Shi and Yuan Wang

*Lehigh University*

We argue that it is important to distinguish between categorization as object recognition and as naming because the relation between the two may not be as straightforward as has often been assumed. We present data from speakers of English, Chinese, and Spanish that support this contention. Speakers of the three languages show substantially different patterns of naming for a set of 60 common containers, but they see the similarities among the objects in much the same way. The observed patterns of naming therefore cannot arise only from the similarities that speakers of the three languages see among the objects. We also offer suggestions about how complexity in naming may arise, and the data provide some evidence consistent with these suggestions. Exploring how artifacts are named vs "known" may provide new insights into artifact categorization. © 1999 Academic Press

What does it mean to categorize? In the real world, at least two different acts are appropriately called categorization. First, people recognize objects as having properties in common with entities stored in memory,[1] and this recognition results in an encoding in an internal representation system. Second, people connect objects with words, both in producing a name for an object and in understanding an object name used by someone else. These two acts are surely closely connected: Objects that have important features in common tend to be given the same name. Indeed, many prominent models of categorization (e.g., Estes, 1986, Gluck & Bower, 1988; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, Palmeri, & McKinley, 1994) as well as several informal theories (e.g., Gelman & Wellman, 1991; Keil, 1989; Rosch

Barbara Malt and Yuan Wang, Department of Psychology, Lehigh University. Steven Sloman and Silvia Gennari, Department of Cognitive and Linguistic Sciences, Brown University. Meiyi Shi, Department of Electrical Engineering and Computer Science, Lehigh University. This work was supported by NIMH Grant MH51271 to Barbara Malt and Steven Sloman. We thank Douglas Medin for suggesting use of the Cultural Consensus Model, Larry Hubert and Martin Richter for statistical advice, and Robert Goldstone, Kenneth Livingston, and Gregory Murphy for helpful comments on an earlier draft of this paper. We also thank the following for permission to reproduce images of their products: Consumer Value Stores, Disney Enterprises, Inc., International Home Foods, Inc., Johnson & Johnson, Lehigh Valley Farms, Mott's Consumer Services, Neutrogena Corporation, Playtex Products Inc., The Procter & Gamble Company, Rite Aid Corporation, Rubber Maid Incorporated, Spring Tree Corporation, and Unilever United States, Inc. Address correspondence and reprint requests to either Barbara Malt, Department of Psychology, 17 Memorial Drive East, Lehigh University, Bethlehem, PA 18015 (e-mail: bcmø@lehigh.edu) or Steven Sloman, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912 (e-mail: Steven_Sloman@brown.edu).

[1] Whether those properties are physical or not, readily perceived or more hidden.

0749-596X/99 $30.00

& Mervis 1975; for discussion see Sloman & Rips, 1998) have taken the primary job for a theory of categorization to be to give an account of what sort(s) of commonalities among objects lead them to be grouped together. They assume that given such an account, name selection is straightforward.

Despite their close connection, though, the act of naming differs critically from the act of recognition. Naming is part of a communication process, whereas recognition is not. The name selected for an object may reflect requirements for successful communication, whereas the representation of commonalities presumably is influenced primarily by constraints such as storage efficiency and the ability to support inference. Because of this fundamental difference in the nature of the two acts, the coupling between recognition and naming may be less than perfect. In particular, we posit that names used for objects reflect influences that are independent of the process of internal representation. The goal of the work presented here is to explore the nature of the relation between recognition and naming of common artifacts.

We begin with the observation that the boundaries for linguistic categories (that is, groups of objects called by the same name) may differ from language to language. We ask whether speakers of languages that have different linguistic category boundaries for a set of objects show differences in their perception of the similarity among the objects, and whether any such differences parallel the differences in how they name the objects. If naming of objects is tightly coupled to their encoding relative to other objects, the answer to both these questions should be yes. If naming and encoding are completely independent, then differences in linguistic category boundaries will not be paralleled at all by differences in perceived similarity among the objects. If the two are partially independent, then we should expect some parallels as well as some systematic differences.

To illustrate this idea, consider the observation that in English, a large stuffed seat for one person is given the same name as a wooden chair, but in Chinese, it is given the same name as a large stuffed seat for two or more people (things that in English would be called "sofa;" Gao, personal communication). Similarly, in English, paper and plastic drinking vessels are called by the same name as coffee cups ("paper cup" and "plastic cup") but in Hebrew, they are called by the same name as things that English speakers would call a "glass" (Kronenfeld, Armstrong, & Wilmoth, 1985). These observations show that the linguistic boundary between "chair" and "sofa" is not the same in Chinese as in English, nor the linguistic boundary between "cup" and "glass" in Hebrew as opposed to English. Would speakers of Chinese perceive the large stuffed seats for one person and the large stuffed seat for multiple people as more similar to each other than speakers of English, who call them by different names? Would a speaker of Hebrew see the paper drinking vessel and the glass one as more similar than speakers of English do? The data to be presented address contrasts of this sort using a large set of containers as stimuli.

### *Related Research*

Research from several different traditions bears on the coupling of similarity and naming. However, none of these enterprises has provided unequivocal answers to the questions we address.

Several studies have shown an apparent dissociation between similarity judgments and preferred category labels for novel objects. For hypothetical objects such as something described as having a diameter halfway between a quarter and a pizza, Rips (1989) found that in many cases, the novel objects were judged more similar to members of one named category (e.g., "pizza" or "quarter") but more likely to be a member of the other (see also Rips & Collins, 1993). In a related vein, Keil (1989) and Rips (1989) presented participants with artifacts described as physically resembling one type of object but having been made to be used as another, or with animals looking very much like one type of animal but said to have internal parts of a different species. They both found that although objects were rated as more similar to the former, they tended to be categorized as the latter. In these last two types of studies, though, "similarity" is apparently interpreted by participants to refer specifically to similarity of perceptual properties. It is not clear that a comparable dissociation would be seen if the similarity judgment instructions led participants to consider less visible as well as perceptual features (see also Smith & Heise, 1992; Jones & Smith, 1993). As for Rips' quarter-pizza results, Smith and Sloman (1994) found that when richer descriptions of objects were used and participants did not have to explain their choices out loud, categorization judgments tended to parallel similarity judgments. In these paradigms, then, the observation of a dissociation between naming and similarity may be dependent on experimental details.

Other studies have looked at how perceived similarity varies with labeled category boundaries. In the literature on categorical perception, it is well established that for both phoneme and color categories, ability to discriminate between stimuli that cross a category boundary is better than ability to discriminate between stimuli that fall within the same category (e.g., Bornstein, 1987; Eimas, Miller, & Jusczyk, 1987; Pastore, 1987; Repp, 1984). Recent evidence suggests that the same phenomenon may occur for familiar faces (Beale & Keil, 1995), for pitch differences perceived by expert musicians (Burns & Ward, 1978), and for members of social categories (Eiser, 1996; McGarty & Penny, 1988; Tajfel, 1957; 1959). Learning of artificial categories also appears to affect the ability to discriminate among stimuli. Goldstone (1994a) found that participants who learned stimuli as members of contrasting categories perceived them as more distinct than participants who did not learn the categories (an "expansion" effect), and Livingston, Andrews, and Harnad (in press) found that participants who learned stimuli as members of the same category perceived them as more similar than participants who did not learn them as members of a shared category (a "compression" effect; see also Kurtz, 1995). All these results suggest that the representation of a stimulus is tightly coupled to its category membership, with perception of the similarity among stimuli paralleling the division of the stimuli into categories. However, this body of research has for the most part not been concerned with categories labeled by familiar common nouns for which category membership has evolved over time and been learned by individuals over an extended period. As we explain below, familiar objects may acquire names that are only loosely related to their intrinsic properties through mechanisms that are unlikely to affect either perceptual categories such as colors and phonemes, or artificial stimuli learned under controlled circumstances.

Studies most directly relevant to our concerns are those that look at well-established lexical categories and make comparisons across speakers of different languages. A number of such studies stem from tests of the Whorfian hypothesis, the proposal that language influences thought (e.g., Whorf, 1956; see Lucy, 1992). One domain in which this hypothesis has been tested is that of named color categories. As just noted, these categories are more perceptual in nature than the object categories of immediate interest to us. However, whereas the work on categorical perception of color has focused more on the universal and possibly innate aspects of color perception (including perception by prelinguistic infants and nonhuman animals; see Bornstein, 1981, 1987), tests of the Whorfian hypothesis have focused on the perception of color by adults with well-established named color categories, and they have looked for evidence of differences in color perception by speakers of languages differing in their color vocabulary. Early research on this topic argued for an influence of color vocabulary on color perception (Brown & Lenneberg, 1954; Lantz & Stefflre, 1964; Lenneberg & Roberts, 1956), but later, more sophisticated work suggested that the opposite of the Whorfian hypothesis was true: Color perception may be universal even in linguistically mature adults and not susceptible to influences of differing color vocabularies (Berlin & Kay, 1969; Heider, 1972; Kay & McDaniel, 1978; see also Kay & Kempton, 1984). The latter conclusion remains somewhat controversial, however (Lucy, 1992). The lessons to be drawn from this body of work about the tightness of the similarity-naming coupling are somewhat murky, then, despite the long history of interest in the topic.

Finally, in a study more closely in line with our own work than any other, Kronenfeld et al. (1985) looked at the names given to various drinking vessels and the similarity among them judged by speakers of several languages. They found distinctive differences in how American, Japanese, and Israeli participants grouped the objects into linguistic categories but relatively small differences in their perceived similarity among the objects. Kronenfeld et al. concluded that there is no simple language/thought isomorphism, and they argued in favor of a view of word use in which focal uses are extended to other objects in a variety of ways. These suggestions are compatible with our approach, and the empirical result suggests that our proposal of a dissociation between naming and recognition may be correct. However, their sample of objects was small, their analysis was largely informal, and they did not attempt to assess whether the observed differences in naming paralleled the differences that did exist in perceived similarity. The study we report here goes substantially beyond Kronenfeld et al.'s study in evaluating the relation of perceived similarity among objects to the names they are given.

In sum, the message from previous work related to our question is mixed. Some of the work suggests that linguistic categories may be closely tied to recognition categories as reflected in perceived similarity among objects, while other results suggest that they may show some dissociation. The bulk of the existing work, in fact, points to a close parallel, but the one study most closely tapping the issues we wish to explore suggests a dissociation. Our data will provide a large-scale evaluation of the issue in a more direct way than has been done previously.

### *Theoretical Framework*

*Basic assumptions.* Drawing on a variety of past research along with the distinction between linguistic categories and recognition categories, we assume a simple working model of object recognition and naming.

Specifically, we assume that objects can be represented as points in a similarity space and that objects tend to form clusters in this space (Rosch & Mervis, 1975; see also Malt, 1995). (We make no a priori assumption about relative feature weights in this space or their stability across contexts). Recognition categories correspond to clusters of points, and linguistic categories are associated with regions of similarity space. In general, the strength of a name for an object will vary in proportion to the similarity of that object to other objects in a cluster, but nearest neighbors are given the most weight in determining name strength. Categories and classification exist at two different levels in this view: Two objects are members of the same recognition category if they are represented within the same object cluster on a particular occasion, and they are members of the same linguistic category if they are given the same name on a particular occasion.

*The relation of recognition to naming.* The framework implies that, in general, the probability that an object is called by a given name will reflect the centrality of the object in a cluster of objects associated with that name; that is, the centrality with respect to a recognition category. However, complexity in how names are chosen for objects can arise in several ways. Some sources of complexity fall out of the assumption that activation of a name is based on the similarity of an object to other objects, not only similarity to a category prototype. Other sources, in contrast, arise as a result of experience with names for objects and pressures from communication about objects (e.g., Clark & Marshall, 1981; see also Sloman & Ahn, in press). It is the hypothesized existence of these sources of complexity that leads to our suggestion that naming and recognition, although closely linked, also will show partial independence. We now briefly describe three such possible sources of complexity.

*Chaining.* Lakoff (1987) proposed that words might be applied to entities quite unlike central exemplars of the word because intermediate uses form a chain from those words into the category. Lakoff's arguments, however, were based on prepositions and noun classifiers (markers preceding nouns in certain contexts in

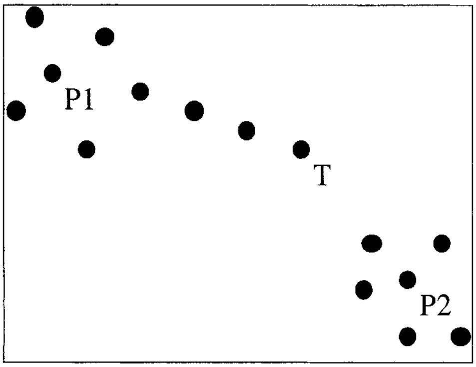


**FIG. 1.** A chain of exemplars leading from T to P1.

many languages). We propose that the same mechanism applies to the case of common nouns. For example, in Fig. 1, the object T may share the name of the cluster of objects with prototype P1 because of a chain of intermediate objects, despite being closer to the cluster with prototype P2 (see also Heit, 1992). In the case of the word “bottle,” for instance, this name is applied to objects as diverse as those in Fig. 2. Given the great diversity, why do people call the objects in the upper panel of Fig. 3 “juice box” instead of “bottle”? The answer may be that they developed from, and are linked to, the more familiar cardboard juice box in the lower panel of Fig. 3, which itself was called a box by virtue of sharing the features of shape and cardboard material with other objects in the cluster of boxes. Presumably the plastic bear in Fig. 3 is called a juice box because of an additional link in the chain: It developed from the plastic juice box that developed from the cardboard juice box.

Note that in terms of recognition, the plastic juice boxes may be conceived of as more like bottles than they are like boxes. In fact, our analysis assumes that their position in similarity space may be closer to the prototype of a bottle than to that of a box, based on their physical and functional features. However, naming, by hypothesis, is influenced by something other than perceived similarity to central examples: It is influenced by similarity to a near neighbor that may be at some distance from central examples, which in turn may have been influenced by another near neighbor.

*Convention.* Another potential source of complexity in naming is ordinary experience. Being told that an object has a particular name


**FIG. 2.** Some bottles. Disney character © Disney Enterprises, Inc. Used by permission from Disney Enterprises, Inc.

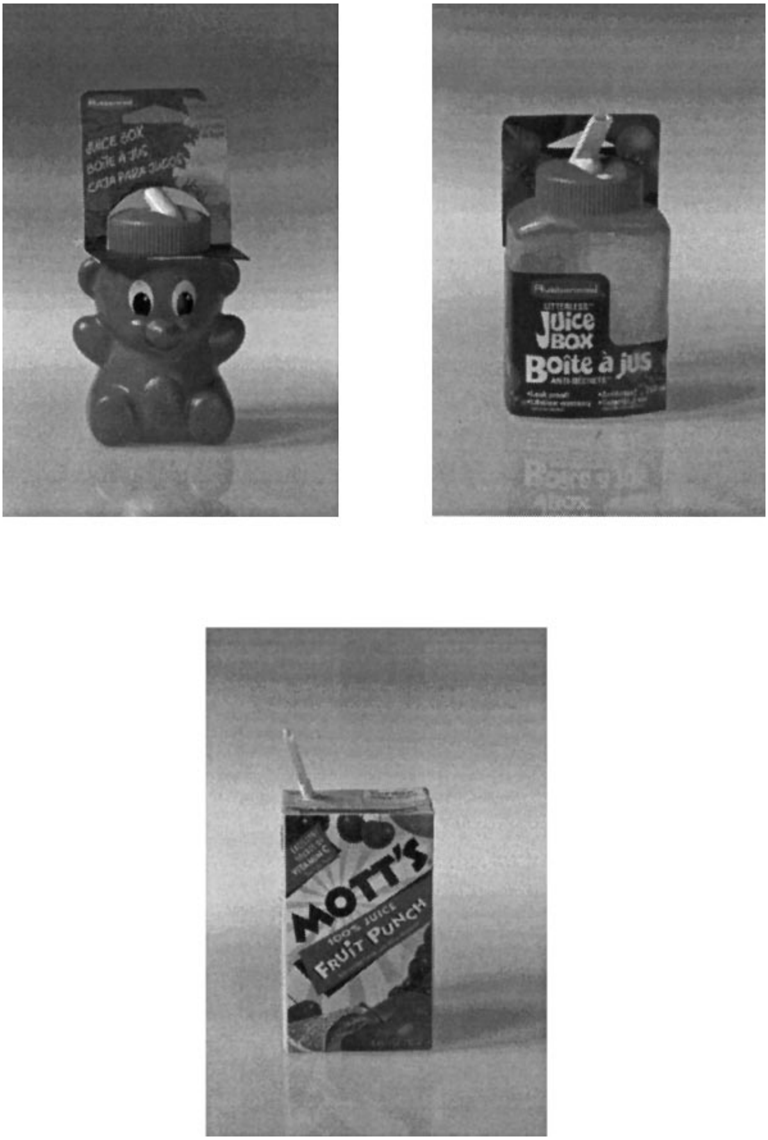


**FIG. 3.** Upper panel: some recent versions of juice boxes. Lower panel: a traditional juice box.

can alter the strength of association of the name to the object, independent of the similarity of the object to other objects associated with the name. Thus these objects may come to have a particular category name as a linguistic convention (cf. Clark, 1993; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Lehrer, 1990) rather than because of specific similarity relations. If the odd plastic objects in Fig. 3 can be called "box," why is the rectangular cardboard object with a lid in Figure 4, which shares many features with other objects usually called "box," rarely called an ice cream "box" but instead an ice cream "carton" or "container"? Similarly, the other object in Fig. 4 is rarely a Chinese food "box" but also a "carton" or "container." These names may well have no psychological explanation other than convention. People may recognize the place of these objects in similarity space among objects that get called boxes, but grew

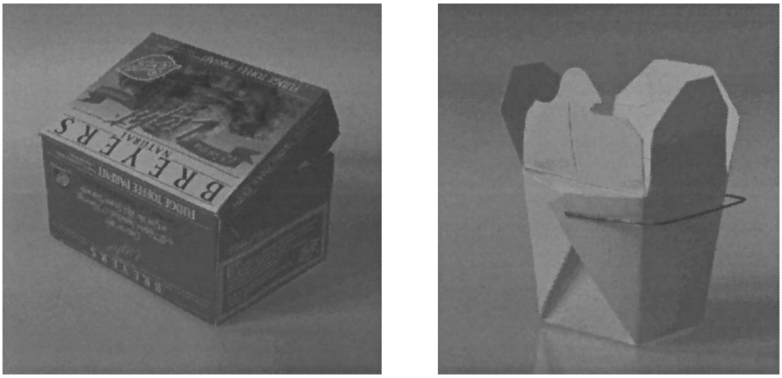


**FIG. 4.** Containers or cartons, not boxes.

up hearing them called cartons or containers, and so they call them by those names. (Why conventions arise that defy similarity is itself an issue. In some cases, members of a chain that originally motivated use of a word may become obsolete, but the use is passed on as a convention. In other cases, a manufacturer or other source of a name may want to be creative or whimsical, or may want to set a product apart from other similar products).

*Pre-emption.* In some cases, people may avoid calling an object by a particular category name because using that name would lead to ambiguity or confusion with another object. For instance, a soup tureen (for serving soups) may be named "tureen" even though its features fall within the range of objects called "bowl" or "pot" because calling it a soup bowl or soup pot would create referential confusion with vessels for eating or cooking soup. The use of "bowl" or "pot" for the serving container may therefore be "pre-empted" (cf. E. V. Clark, 1988, 1993; H. H. Clark, 1991; Lehrer, 1990; Markman & Wachtel, 1988; McCawley, 1978) by the other uses of these names. Again, in terms of recognition, people no doubt perceive the resemblance to bowls and pots. However, this resemblance alone does not determine the name normally used for the object.

## *The Cross-Linguistic Approach*

The possibility of such sources of complexity in naming suggests that although objects that are recognized as similar to other objects in a cluster will tend to have the same name as members of that cluster and vice versa, there will be occasions when this close relation breaks down. An object perceived as similar to others in a cluster may come to have a different name, and an object that shares the same name as members of a cluster may sometimes be perceived as more similar to members of a cluster named differently. It is not the primary purpose of the present paper to find support for the specific mechanisms just discussed. Rather, the mechanisms are described here to provide concrete illustrations of why we hypothesize that the relation between recognition and naming is not always straightforward. For the case of the cross-linguistic comparisons we focus on, what is important to note is that such mechanisms have the potential to create different patterns of naming for the same objects across languages, even if perception of similarities is constant across speakers of languages. For instance, an American manufacturer may establish a convention in English for calling an object by a name other than the one dictated by similarity, but a parallel convention may not exist in other languages. Similarly, one culture may develop a chain that influences the naming of a particular object in the language of that culture, while another culture with a slightly different assortment of objects in use may develop a different chain of naming or no chain at all for that object. In addition, the mechanisms may build on each other. If one culture adopts a convention for naming a particular object in a distinctive way, that object may influence naming of the next new variant to appear (i.e., a chain begins to develop), whereas a different culture that names the original object based only on similarity to a cluster will have no basis for naming the new variant differently.[2]

We have already noted that speakers of different languages do, in fact, seem to have different linguistic boundaries among objects, for at least some domains. Our goal in this work, then, is to assess whether such differences in linguistic boundaries correspond to differences in perceived similarity among the objects, or whether there is some degree of independence between the linguistic categories and the perception of the similarities among the objects.

## STUDY

We collected names and similarity judgments for a large set of objects, for speakers of English, Chinese, and Spanish. We can ask three specific questions: (1) Does the division of objects into linguistic categories differ across the three languages for this set of objects? (2) Does the perception of similarity among the objects differ across speakers of the three languages? (3) If at least some differences in linguistic categorization and perceived similarity are found, do these differences parallel one another?

Because there has been significant debate in the literature over what sort of features may be most relevant to artifact categorization, we collected information about three distinct variants of perceived similarity: physical, functional, and overall similarity. By looking at the correspondence of each type of similarity judgment to naming, we can evaluate whether similarity on one particular type of feature (or combination, in the case of overall similarity) shows a closer coupling to naming than others. Manipulating the type of similarity participants judge also helps address the observation that perceived similarity can vary depending on context or task demands (e.g., Barsalou, 1983, 1993; Medin, Goldstone, & Gentner, 1993; Murphy & Medin, 1985; Smith & Samuelson, 1997). By explicitly instructing participants about what aspect of similarity they should focus on, construals of the task other than the one we intend should be reduced, and we can have more confidence that we know what dimensions participants are using in their judgments. In addition, we can assess whether similarity would parallel naming under contexts that would favor any of these three aspects of similarity.[3]

As already noted, although it has been argued in the past that categorization is not based on perceived similarity (e.g., Gelman, 1988; Keil, 1989; Rips, 1989; see also Carey, 1985), this argument has been made using "similarity" to mean similarity of what are referred to as "su-

[2] Lehrer (1990) provides several interesting examples of verbs that share a prototype but show differences in their extensions across cultures. French and Spanish have verbs with the same central meaning as English "run" (i.e., fast locomotion on foot), but whereas in English the operation of machines is referred to as running, in French and Spanish the verb for "walk" is used to talk about machines. Similarly, although the prototypical meaning of "run" is the same in American and British English, Americans extend the word to refer to a politician seeking election to office but the British do not, using the word "stand" for this activity. Such cases suggest language- (or culture-) specific processes at work in determining the particular peripheral activities these verbs are extended to, in agreement with our suggestions about nouns.

[3] Although it might be suggested that context can affect similarity judgments not just in the sense of causing heavier weighting of different dimensions (or sets of dimensions) at different times, but also in the sense of causing stretching or shrinking of psychological distance on a dimension, we do not see this possibility as a major concern for our investigation. A cardboard container for drinking juice may be judged more similar to a traditional glass juice bottle if presented in a triad with a chair than if presented with a plastic juice bottle. However, in the case of mentally representing a container in everyday encounters, the context of encoding is presumably most often that of other containers—that is, the object is functioning as a container and therefore is represented in terms of its properties shared with previously encoded containers, not in terms of properties shared with chairs (or dogs, etc.). As Goldstone (1994b) notes, stable contexts can provide stable grounding for perceived similarity.

Of course, in some real word situations, a container may be used as a paperweight or a decoration, etc., and this change from its normal use may influence how it is mentally represented. However, our present concern is with cases where an object is used in accordance with its intended use and it is represented relative to the domain it is ordinarily considered to be part of.

perficial" features, primarily physical attributes such as size, color, and texture. These theorists have suggested that categorization is based on "deeper," more hidden properties. For artifacts, it has been suggested that an artifact's function may be the "deeper" property that determines category membership (e.g., Keil, 1989; Medin & Ortony, 1989). By including functional as well as physical and overall sorts, we will be able to evaluate not only the relation between physical and overall similarity among objects and the names they receive, but also the relation between perceived similarity on "deep" functional properties and the names that the objects receive.

### *Method*

*Participants.* Participants for the primary tasks were 76 native speakers of English, all students at Lehigh University; 50 native speakers of Chinese, 10 of whom were students at Lehigh and 40 of whom were students at Shanghai University, China; and 53 native speakers of Spanish, all students at Comahue National University, Argentina. An additional 15 Lehigh University students gave familiarity judgments described below. The 10 Chinese students at Lehigh University used English regularly in their academic work but Chinese as their primary language for all other purposes. The remaining Chinese participants and all the Argentinean participants exclusively used their native language in their daily activities, although some had had training in English. (Two of the Argentinean students considered themselves fluent in English; the rest did not). The American students received course credit for their participation; the Chinese and Argentinean students either were paid or participated as unpaid volunteers.

*Materials.* The stimuli were a set of 60 common containers that were mostly a mixture of objects likely to be called "bottle" or "jar" in English, along with some additional ones not likely to be called either "bottle" or "jar" but sharing one or more salient properties with bottles and jars. The objects were chosen to represent a wide range of bottles, jars, and other similar containers. The large size of the stimulus set, and the large range of objects in it, allows a sensitive comparison of the linguistic category boundaries for speakers of the different languages and a thorough evaluation of the relation between naming and perceived similarity.

The objects were photographed against a neutral background with a constant camera distance to preserve relative size. A 12-inch ruler was also included in front of each object to provide size information. Figure 5 displays black-and-white images of some of the objects, taken from the color photographs used in the experiment.

For the Chinese and Argentinean participants, each picture was marked at the bottom, in Chinese or Spanish respectively, with the nature of the contents of the object (e.g., "milk," "medicine"). That information would otherwise be less obvious to these participants than to Americans since the labels on the containers were in English. The information given did not in any way indicate a name for the type of container.

*Procedure.* Because of the large number of objects involved, presentation of all possible pairs of objects for pairwise similarity judgments would have been prohibitively time consuming, especially for data collected outside the U.S. where facilities and access to participants was limited. We therefore asked participants to sort the objects into piles. The sorting data can then be used to derive a measure of the similarity between each pair of objects (see analysis below). This method for obtaining similarity judgments for large object sets has been widely used (Rosenberg & Kim, 1975).

Sorts were based on either physical, functional, or overall similarity. Each participant carried out two of the three sorts because doing all three would have been too time-consuming. Sort type combinations were rotated across participants so that each pair of two sorts was carried out in each possible order approximately equally often. For English, 52 participants carried out physical sorts, 51 did functional sorts, and 49 did overall sorts. For Chinese, 34 participants did physical sorts, 33 did functional sorts, and 33 did overall sorts. For Spanish,

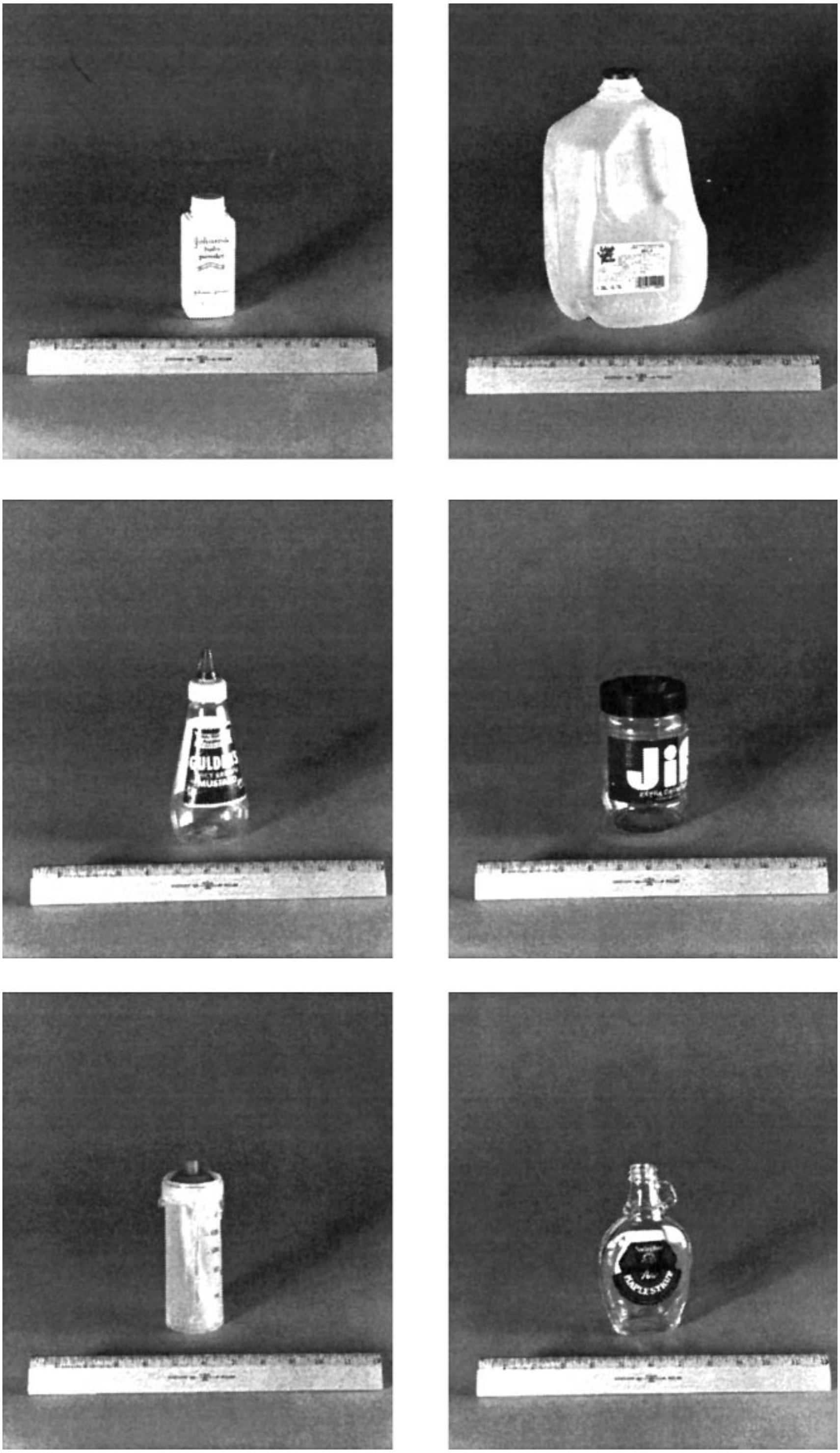


**FIG. 5.** Black-and-white versions of some of the color photographs used in the experiment.

there were 35 participants for the physical and functional sorts and 36 for the overall sort.

Participants were given the general information that they would be asked to look at pictures of familiar objects and indicate how similar they are to one another. They were then asked to look through the set of 60 photos to familiarize themselves with the range of objects. When they were through looking at the photos, they were told that they would be asked to sort the pictures twice, each time in a slightly different way. They were given instructions for the first sort and were given as much time as they wanted to complete it. After the first sort was finished, the piles formed were recorded, the photos were shuffled, and instructions were given for the second sort. Participants again were allowed as much time as they wanted to complete the sort.

Instructions for the English-speaking participants for the physical sort were as follows: "This time, I would like you to focus on the PHYSICAL QUALITIES of each container; that is, what it looks like, what it's made of, and so on. I would like you to put together into piles all the containers that you think are very similar to each other PHYSICALLY."

"Note that we are interested in how physically similar the CONTAINERS themselves are, not what comes in the containers. Only put two pictures together if the containers are like each other. Do NOT put pictures together just because the containers hold things that tend to be found together. For instance, if one container contains coffee and another contains milk (or one contains peanut butter and another contains jelly), or if several containers contain cleaning products (or health products), DON'T put them together unless you really think the containers themselves are physically alike."

"Please use at least two different piles, but not more than 15. You can have as many or as few pictures in each pile as you want, and you can take as long as you want to decide on your sorting."

Instructions for the functional sort were similar except that they stated that "This time, I would like you to focus on the FUNCTION or USE of each container, that is, how it contains the substance that is in it (in a stack; in separate pieces, as a single solid; as a liquid; with pouring capability, etc.). I would like you to put together into piles all the containers that you think are very similar to each other in how they FUNCTION." The warning about not putting objects together just because they are often found together was modified to ask that they not be put together "unless you really think that the containers themselves are holding similar substances in similar ways."

Instructions for the overall sort stated that "This time, I would like you to focus on the OVERALL QUALITIES of each container. This means you can focus on any feature of the container including what it looks like, what it's made of, how it contains the substance that is in it (in a stack; in separate pieces; as a single solid; as a liquid; with pouring capability, etc.) or any other aspect of the container that seems important or natural to you. I would like you to put together into piles all the containers that you think are very similar to each other OVERALL." The warning was also modified to ask that containers not be put together "unless you really think the containers themselves are similar in an overall way."

Data from Argentinean participants were collected in Spanish by the third author (a native speaker of Argentinean Spanish), and data from Chinese participants (both in the US and in China) were collected in Chinese by the fourth author (a native speaker of Chinese from the Shanghai area). These authors, who are also fluent in English, translated the English instructions into their native language taking care to convey the meaning of the English versions as accurately as possible. The examples of substances that tend to be found together were modified for each language to be ones familiar to that group of participants.

After participants completed the sorting portion of the experiment, the pictures were again shuffled. Participants were then asked to give a name for each object. They were asked to give whatever name seemed the best or most natural to them, and they were told that it could be one word or more than one word. The instructions emphasized that participants should name the object itself, not what it contained. Twenty-eight of the American participants provided the naming data reported here (the remaining American participants gave other sorts of name judgments that will not be discussed here). Fifty-one participants gave names for Spanish, and all fifty of the Chinese participants gave names.

Following the naming task, a subset of the Chinese and Argentinean participants also gave typicality ratings for the objects for use in analyses not reported here. Nine of the Argentinean participants also rated each object for its familiarity on a scale of 0 to 7 (where 0 represented not at all familiar and 7 represented high familiarity). Completion of all the tasks took 1 to 2 h. A separate group of 15 American participants carried out the familiarity rating task in English in a short session in which they also completed another rating task not relevant to this study.

*Results and Discussion*

We describe the results in three parts. First, we compare the three languages' linguistic categories. Second, we compare the perceived similarity among the objects as reflected in the sortings done by speakers of the three languages. Finally, we examine whether differences in linguistic categorization are paralleled by differences in perceived similarity among the objects.

*Comparison of linguistic category boundaries.* The names produced for each object were tallied separately for the three languages. Tallies were based on the head noun of each response (e.g., "bottle," "brown bottle," and "small bottle" all counted as instances of the "bottle" name). In Spanish, names given were sometimes single-word diminutive forms of other names (e.g., "frasquito" is a diminutive of "frasco"). In the analyses reported below, the diminutive form was combined with the nondiminutives and treated as a single category since phrases such as "small bottle" in English were counted as instances of "bottle." However, the pattern of results remains the same if the diminutives are treated as separate categories.

Some names were produced by only one or a small number of subjects across all 60 objects. We eliminated these idiosyncratic or otherwise infrequent responses by restricting analyses to category names that were the dominant (most frequent) name for at least one object. There were 7 such names in English, 5 in Chinese, and 15 in Spanish.

Our first analysis considers only the dominant name for each object. Table 1 shows the number of objects out of 60 for which each name was most frequent. (Pilot data from Costa Rican, Puerto Rican, etc. participants revealed large regional differences in the Spanish names used for these objects. The names reported here do not necessarily reflect the names that would be dominant for other dialects of Spanish). To help indicate the amount of overlap among the categories of the various languages, the Chinese and Spanish categories are described in terms of their English composition.

For American students, the objects fell mainly into three categories, "bottle," "jar," and "container," that were roughly equal in size, with a few objects being given other names. For Chinese students, most of the objects fell into one large category that encompassed all the English jars, most of the English bottles, and some of the English containers. The remainder were distributed across four other categories. Argentinean students used the greatest number of name categories, a total of 15. The most frequent by far was "frasco" and its diminutive "frasquito." The other names tended to be restricted to a small number of objects. As is evident by examination of the English composition of the Chinese and Spanish categories, there is some correspondence in how the languages divide the objects into linguistic categories; for instance, all the objects called "jar" in English are put into a single category in both Chinese and Spanish, and the three objects called "jug" in English are put into a single category in Spanish. At the same time, though, the differences in the way the three languages group the objects into linguistic categories are striking.

This analysis does not take all of the naming data into account, however. Very few of the objects are well represented by a single dominant name. Only two of the 60 objects were

TABLE 1

English, Chinese, and Spanish Linguistic Categories

| English | *N* |
|---|---|
| Jar | 19 |
| Bottle | 16 |
| Container | 15 |
| Can | 5 |
| Jug | 3 |
| Tube | 1 |
| Box | 1 |

| Chinese | *N* | English Composition |
|---|---|---|
| 瓶 | 40 | 13 bottles, 8 containers, 19 jars |
| 罐 | 10 | 3 containers, 2 bottles, 5 cans |
| 桶 | 5 | 3 jugs, 1 bottle, 1 container |
| 盒 | 4 | 3 containers, 1 box |
| 管 | 1 | 1 tube |

| Spanish | *N* | English composition |
|---|---|---|
| frasco/frasquito | 28 | 6 bottles, 3 containers, 19 jars |
| envase | 6 | 4 containers, 2 bottles |
| bidon | 6 | 3 jugs, 1 bottle, 2 containers |
| aerosol | 3 | 3 cans |
| botella | 3 | 3 bottles |
| pote/potecito | 2 | 2 containers |
| lata | 2 | 2 cans |
| tarro | 2 | 2 containers |
| mamadera | 2 | 2 bottles |
| gotero | 1 | 1 bottle |
| caja | 1 | 1 box |
| talquera | 1 | 1 container |
| taper | 1 | 1 container |
| roceador | 1 | 1 bottle |
| pomo | 1 | 1 tube |

called by the same name by every English speaker; only seven by Spanish speakers; and twelve by Chinese. Our second analysis of the naming data therefore takes into account, not only the dominant name of each object, but its entire name distribution; that is, the number of times each name was assigned to each object. The idea is to compare the linguistic categories of the three languages by comparing the naming distributions across the languages. We cannot compare the names of objects directly across languages because the languages, of course, have different sets of names. Instead, we compare the similarity of each object's name distribution to every other object's using a Pearson correlation: For each pair of objects within a language, we calculated the correlation, across all the names, between the name frequencies tallied for those objects. (Other measures of name similarity, such as the Euclidean distance between them, gave comparable results). For each language, this measure gives us 1770 correlations representing the name distribution similarity for each possible pairing of the 60 objects ($60 \times 59 / 2$ because the matrix is symmetric). The 1770 name similarity values for one language can then be correlated with the values for name similarity in each of the other two languages. This second-order measure reflects the extent to which two languages correspond in the pairs of objects that have similar name distributions. Table 2 gives the correlations between each pair of languages' name similarity measures.

The correlations are all positive (though their statistical significance cannot be properly evaluated because cell entries are not independent; each object contributed to 59 observations). The positive correlations indicate that the three languages do not divide up the set of objects in completely independent ways, consistent with our earlier observation. The relatively high correlations between English and Spanish and between Chinese and Spanish are the result of a cluster of objects that have low name similarities in both languages and another cluster with

TABLE 2

Correlations among Languages between Measures of Name Similarity

| | Chinese | Spanish |
|---|---|---|
| English | .35 | .54 |
| Chinese | | .55 |

high similarities. Nevertheless, the correlations for all the language pairs are appreciably less than 1; the languages show substantial differences as well as some agreement. English and Chinese show particularly little correspondence: Each language explains only 10–12% of the other's variance. None of the correlations indicate that one language can account for more than 40% of any other language's variance.[4] This result supports the pattern that was revealed by looking at the dominant names: The linguistic categories of the three languages are not entirely independent of each other, but nevertheless they do diverge substantially.

To further evaluate the differences among the groups, we used the cultural consensus model (CCM) of Romney, Weller, and Batchelder (1986). The general idea of the model is to represent the relations among the responses of all participants regardless of group. The representation is then analyzed to see if its underlying structure embodies group differences. More specifically, a measure of association was computed for every pair of participants in the experiment. This measure represents the proportion of object pairs that both participants individually gave the same name to. Let $X_{ikl}$ be 1 if participant *i* gave the same name to objects *k* and *l* and O if the participant did not. Then the measure of association between participants *i* and *j* is

$$M_{ij} = \Sigma_{k,l>k}(X_{ikl} \times X_{jkl})/1770.$$

[4] The modesty of the correlations for English with Spanish cannot be explained by lesser familiarity with the objects by one group leading to different naming strategies. Eliminating nine objects that fell below the midpoint of the familiarity scale for Argentineans and Americans improved the correlation only very slightly (to $r = .56$). Familiarity ratings were not obtained from Chinese participants so the parallel analysis cannot be carried out for correlations with Chinese. It is possible that lesser familiarity with some of the objects did affect Chinese patterns of naming. However, this possibility is entirely compatible with our view of naming. If Chinese participants have had less exposure to some of the objects, they are less likely to have acquired naming conventions for the objects that violate similarity relations, for example, and so their linguistic category boundaries are likely to differ from those of people with more direct exposure to names for the objects.

The resulting matrix of associations among the participants was then subjected to principal components analysis to determine its underlying factor structure. The hypothesis of no group differences implies that only a single factor should emerge and that all participants should load equally on it. To estimate the principal components and obtain factor loadings, we used the principal axis factoring option in the FACTOR command of SPSS (SPSS, 1990).

Three factors emerged that accounted for significant variance (i.e., that had an eigenvalue greater than 1). The first factor accounted for 19% of the total variance of the association matrix, the second for 2%, and the third for only .9%. All three factors distinguished the groups effectively. For the first factor, mean factor loadings for the American, Chinese, and Argentineans were .31, .55, and .34, respectively, $F(2,126) = 236$; MSe = .003. For the second factor, means were .09, −.13, and .16, $F(2,126) = 482$; MSe = .0022. For the third factor, means were .16, −.016, and −.053, respectively, $F(2,126) = 166$; MSe = .0025. For all three factors, $p < .0001$. In short, this analysis confirms that the groups named the objects differently.

*Comparison of perceived similarity.* The sorting data were used to derive a measure of the similarity between each pair of objects. Pairwise similarity was determined by counting the number of times that a pair of objects was placed in the same pile across the members of a group (separately for each sort type). Objects that were placed into the same pile by a large number of members of the group are considered highly similar for that group, and objects that are rarely or never put into the same pile are considered low in similarity for the group. As in the measurement of name similarity, this procedure yields 1770 pairwise judgments, one for each possible pair of objects.

Our initial comparison of perceived similarity looks at how the three groups sorted the objects in the aggregate. To compare groups, we correlated each pair of groups' judgments across all pairs of objects. The three groups sorted the objects remarkably similarly. Table 3 shows the correlations between the groups for

TABLE 3

Correlations among Groups in Sorting

| | Chinese | Argentineans |
|---|---|---|
| **Overall similarity** | | |
| Americans | .91 | .94 |
| Chinese | | .91 |
| **Physical similarity** | | |
| Americans | .89 | .88 |
| Chinese | | .82 |
| **Functional Similarity** | | |
| Americans | .77 | .79 |
| Chinese | | .55 |

the sorts by overall, physical, and functional similarity. The groups' sorts were more highly correlated with each other than their naming similarities were, for all three types of sorts. The correspondence in perceived similarity is most striking for sorting by overall similarity—all correlations between groups are above .9—and it is comparable to the agreement between halves of the same group (with mean split-half reliabilities of .89). The correspondence between groups is also high for physical similarity—all correlations are above .8 and close to the mean split-half reliability of .91. The groups disagreed more on their functional sorts, particularly the Chinese and Argentineans, and the agreement between groups was lower than within the groups (with a mean split-half correlation of .93). This lower correspondence for functional sorts is perhaps to be expected since the Chinese and Argentinean students may have had less direct experience with the objects than the American students.[5]

The fact that overall sorting correlations are highest of all is surprising in that the participants were free to use whatever features of the objects they wanted in this sort, and yet their agreement was higher than for the more constrained sorts. If members of different cultures focus preferentially on different aspects of the objects in comparing them, then the less constrained sort should show less correspondence across groups. The fact that it does not suggests that the groups spontaneously used similar weightings in making comparisons.

To evaluate the differences among the groups in a way that is parallel to the analysis of differences for naming, we applied the CCM to these data. The analysis is the same as that for naming except that instead of considering whether an individual gave two objects the same name, we consider whether the individual sorted them into the same pile. Again, if there are no group differences, only a single factor should emerge and all participants should load equally on it.

For the overall similarity sorts, only a single factor had an eigenvalue greater than 1. That factor accounted for 8% of the total variance of the matrix. Mean factor loadings for the Americans, Chinese, and Argentineans were 0.25, 0.28, and 0.27, and the difference among these loadings was not significant, $F < 1$. Thus, the CCM indicates no group differences for overall sorts.

For the physical similarity sorts, two factors accounted for significant variance. The first accounted for 8.1% of the variance and the second for only 1.3%. Mean factor loadings for the three groups on the first factor were .26, .31, and .24, $F(2,118) = 4.75$; MSe $= .0090$; $p < .05$, and on the second factor were .012, $-.029$, and .078, respectively, $F(2,118) = 9.34$; MSe $=$ .011; $p < .001$. So the factors emerging from

[5] The lower correspondence for functional sorts is not due to a difference in the number of piles used by participants in the different cultures (which might suggest differing interpretations of the instructions): Means for the function sorts were 10.06 for Americans, 10.57 for Chinese, and 10.71 for Argentineans, $F(2,108) < 1$. There was also no difference in the number of piles used in the overall sorts, with means of 11.10, 11.44, and 10.64, respectively, $F < 1$. Although the mean number of piles in the physical sorts did vary slightly more, with means of 10.67, 9.63, and 11.91, respectively, $F(2,110) = 3.33$, MSe $= 12.22$, $p < .05$, the correlations for the physical sorts are high enough and the difference among the piles small enough that it is unlikely the instructions were interpreted differently for physical sorts either.

the analysis of physical sorts do distinguish the groups. However, the difference among the loadings is substantially smaller than the difference was for naming, indicating that these factors do not distinguish the groups nearly as well as the naming factors did. The smaller difference for sorting is shown by analyses of variance on the factor loadings, comparing task (physical sorting versus naming) for the three groups. On both the first and the second factor, the interaction between group and task was highly significant, $F(2,244) = 32.78$; MSe = .0060 on the first factor and $F(2,244) = 32.61$; MSe = .0064 on the second factor. Thus, the CCM shows much smaller group differences for physical sorting than for naming.

For the functional similarity sorts, again two factors emerged. The first accounted for 7.7% of the variance and the second for 1.6%. Again, the factors did distinguish the groups: For factor 1, groups means were .25, .29, and .27, respectively, $F(2,116) = 4.13$; MSe = .0060; $p < .05$, and for factor 2, means were .0037, −.10, and .12, respectively, $F(2,116) = 44.65$; MSe = .0092; $p < .0001$. However, again the factors did not distinguish the groups nearly as well as the naming factors did. The interaction between task and group was significant both for the first factor, $F(2,242) = 51.18$; MSe = .0045, and for the second, $F(2,242) = 10.60$; MSe = .0056. Thus, the CCM shows group differences for functional sorts, but smaller differences than observed for naming. (It should be noted that the ANOVAs assessing this and the interaction for physical sorts are not optimally appropriate because the factors scores for naming and sorting have different distributions. Nevertheless, the size of the interactions strongly suggests that the groups differ more on the naming task than on the sorting task. Moreover, differences between the distributions for the two tasks cannot easily explain the observed interactions).

In sum, relative to the rather striking differences in linguistic categorization among the three languages, speakers of those languages show negligible differences in their perception of the overall similarity among the objects and significant but only small differences in their perception of physical and functional similarity. This observation leads to the suggestion that the coupling between linguistic categories and perception of similarities among the objects, while clearly present, cannot be very tight. The final set of analyses explores their relation in more detail.

*Relation of naming to perceived similarity.* To provide a general sense of the relation between naming and the perception of similarity for each group of subjects, we obtained multidimensional scaling solutions (e.g., Shepard, 1974) of each set of sorting data using the KYST algorithm, and we labeled the objects in each solution with their dominant linguistic category. The drop in stress values was larger between one- and two-dimensional solutions than between two- and three-dimensional solutions in all cases, and the stress values for the two-dimensional solutions were all below .20, indicating a good fit of the solution to the data. We present the two-dimensional solutions since they are the most interpretable and little advantage is gained from the three-dimensional version. These solutions are given in Figs. 6–14. As the figures show, many members of each linguistic category cluster together in similarity space, but some occur closer to members of other linguistic categories. The failure of the similarity space to fully separate the linguistic categories is as clear for the case of the functional sorts as for the other two types, a fact that is not consistent with views of artifact categorization in which shared function is the crucial determinant of artifact category membership. Across all three sort types, the distribution of names shown in these MDS solutions suggests that the linguistic categories are complex and do not map directly onto the similarity clusters.

Our next analyses assess the relation between perceived similarity and linguistic categorization in more detail: What is their overall statistical relation, and, further, do the small differences between the groups' sorting performance parallel any of the differences observed in their naming? As our primary analysis, we correlated each language's measure of name similarity (the correlation mea-

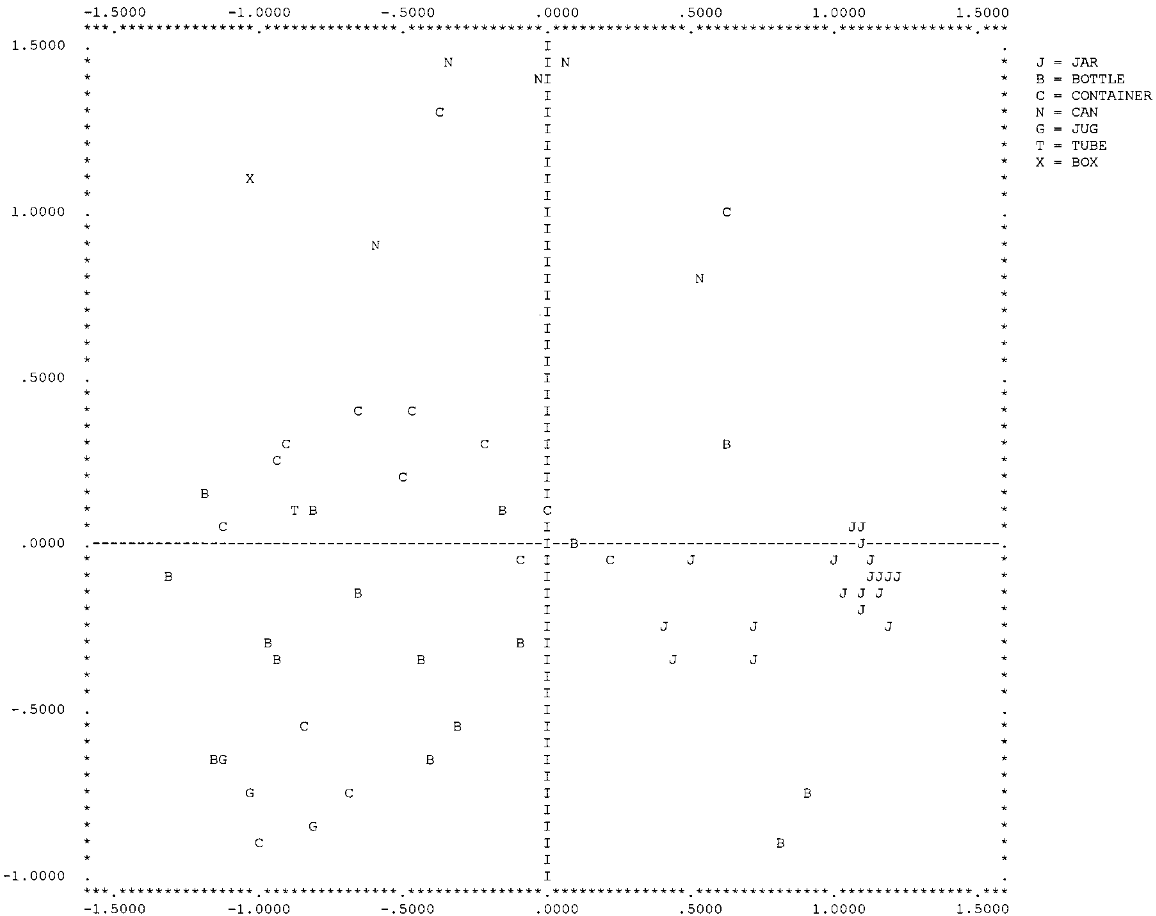


**FIG. 6.** MDS solution for American physical sorting data. Note: For all the solutions, when two or more objects fall at the same point in space, they are shown side by side instead so that all objects are represented in the solution.

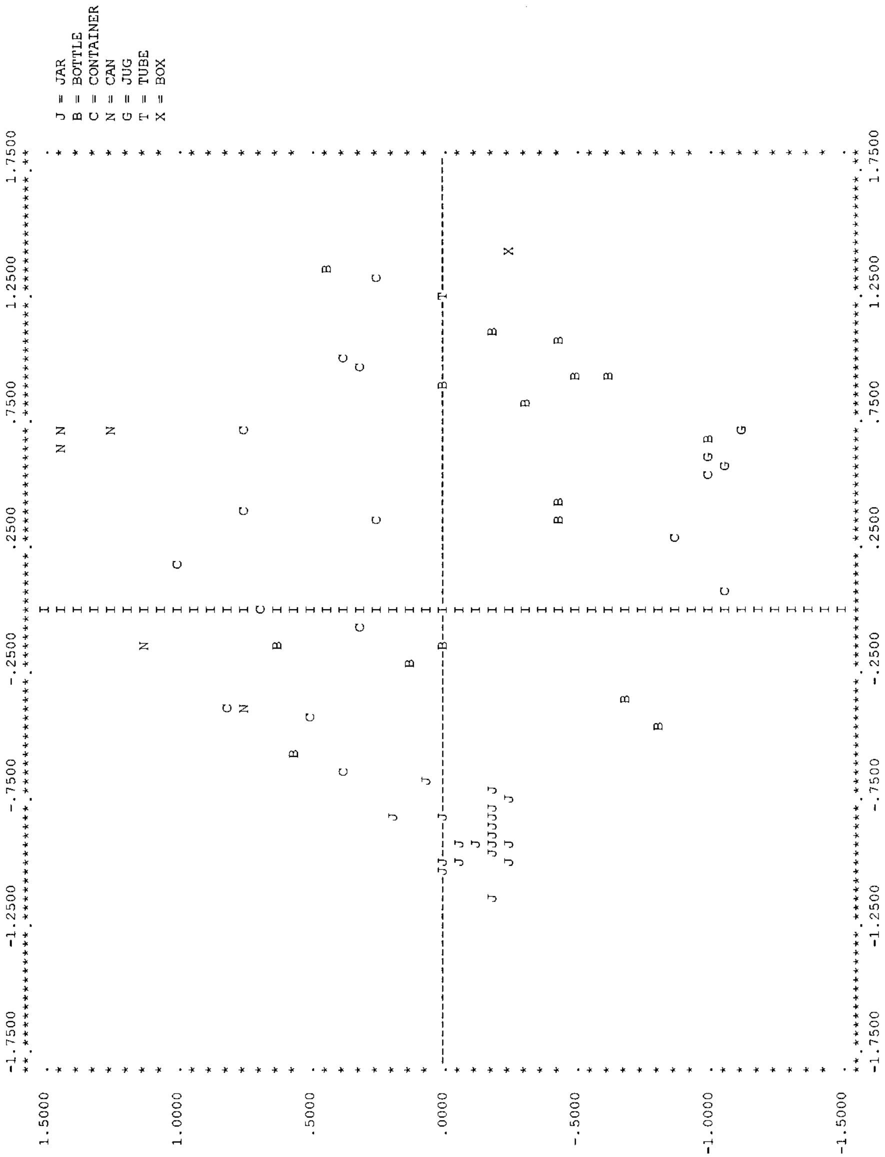


**FIG. 7.** MDS solution for American overall sorting data.

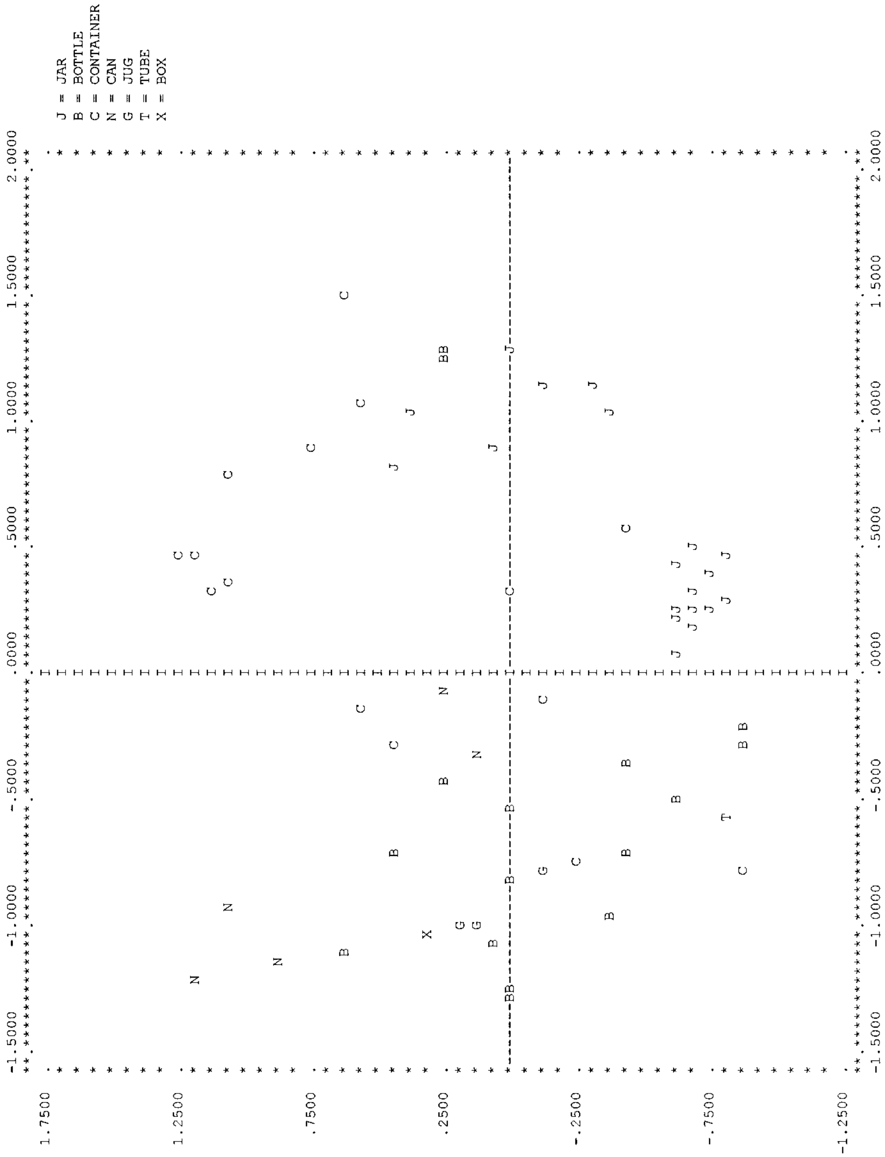


**FIG. 8.** MDS solution for American functional sorting data.

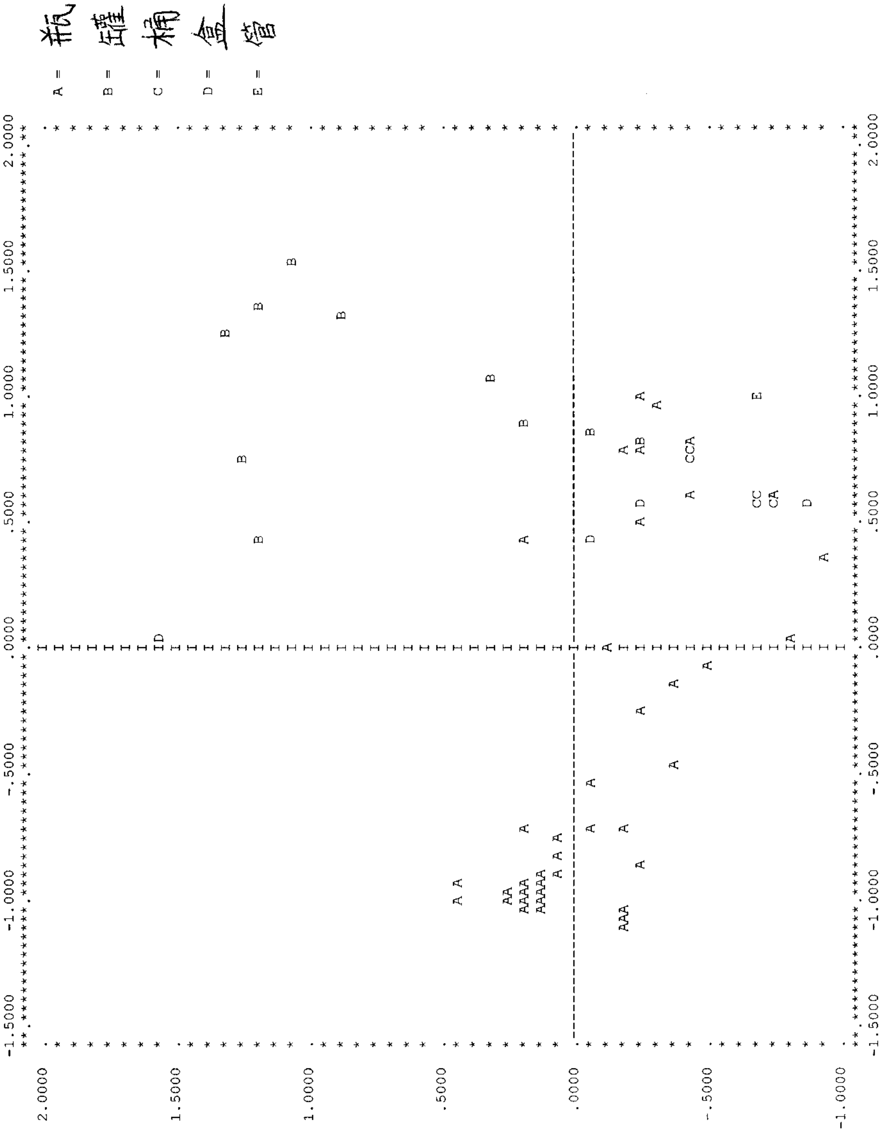


**FIG. 9.** MDS solution for Chinese physical sorting data.

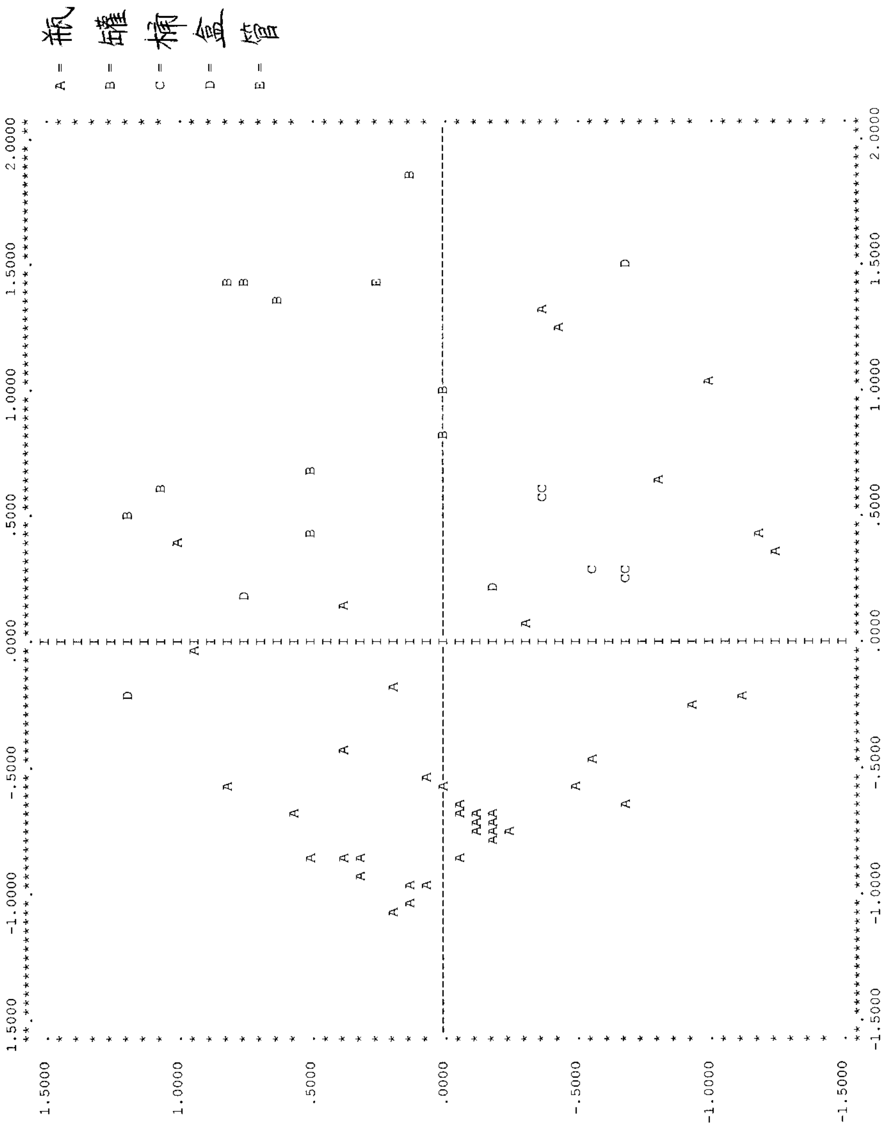

**FIG. 10.** MDS solution for Chinese overall sorting data.

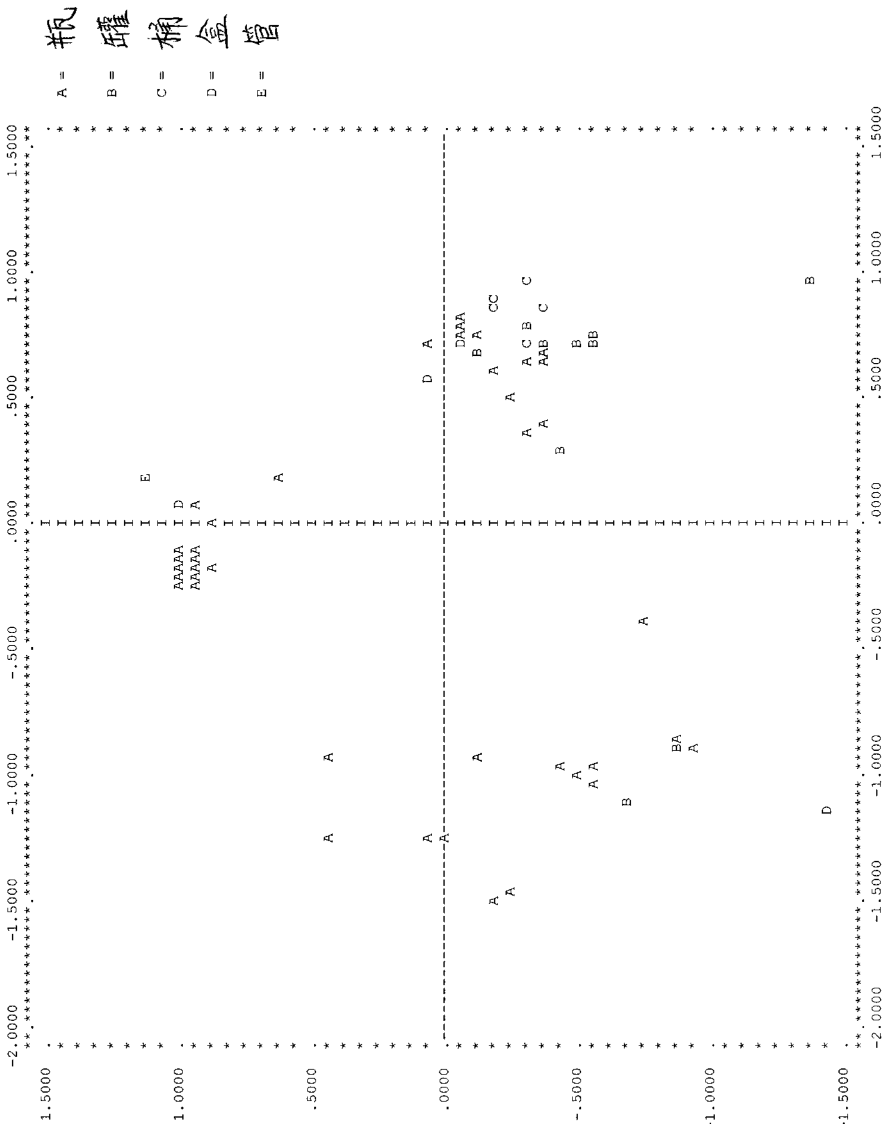


**FIG. 11.** MDS solution for Chinese functional sorting data.

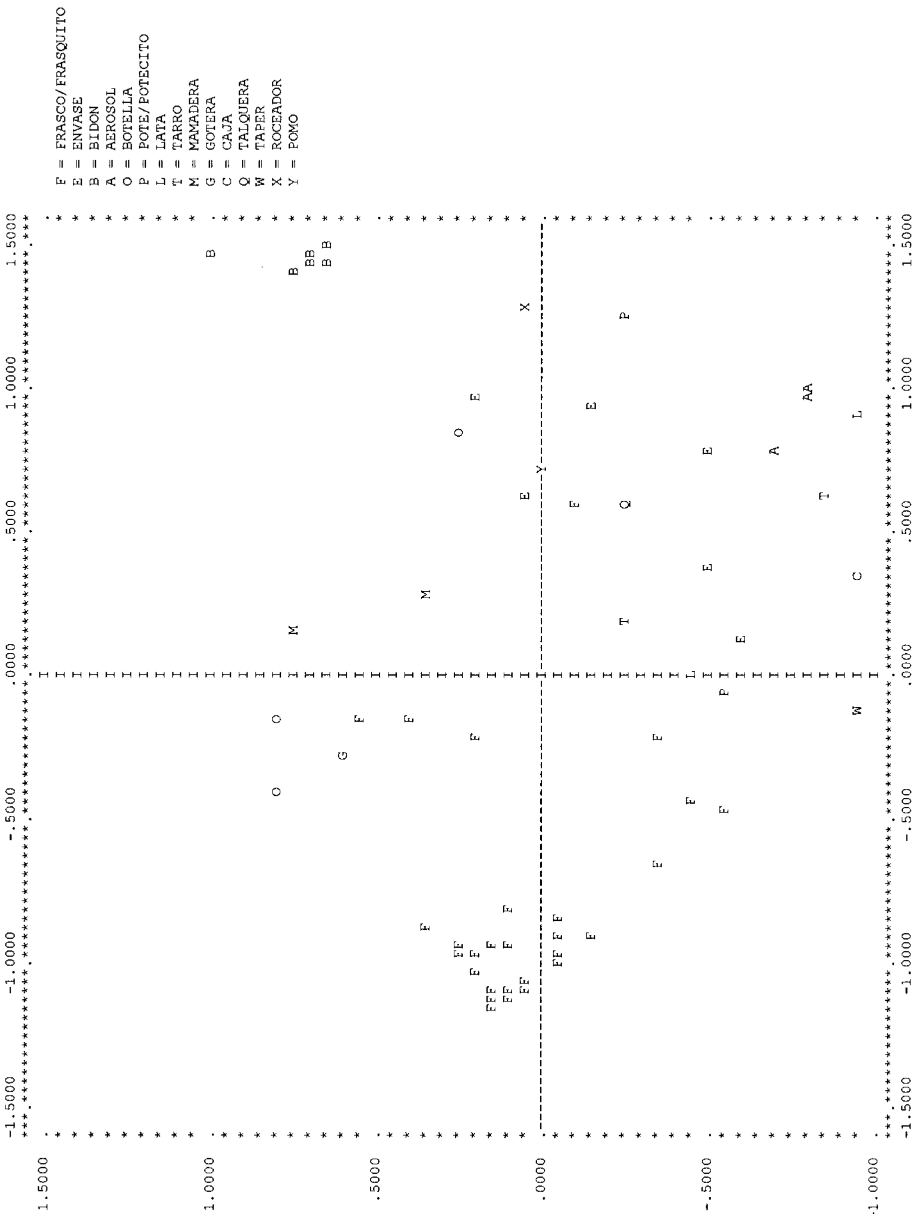


**FIG. 12.** MDS solution for Spanish physical sorting data.

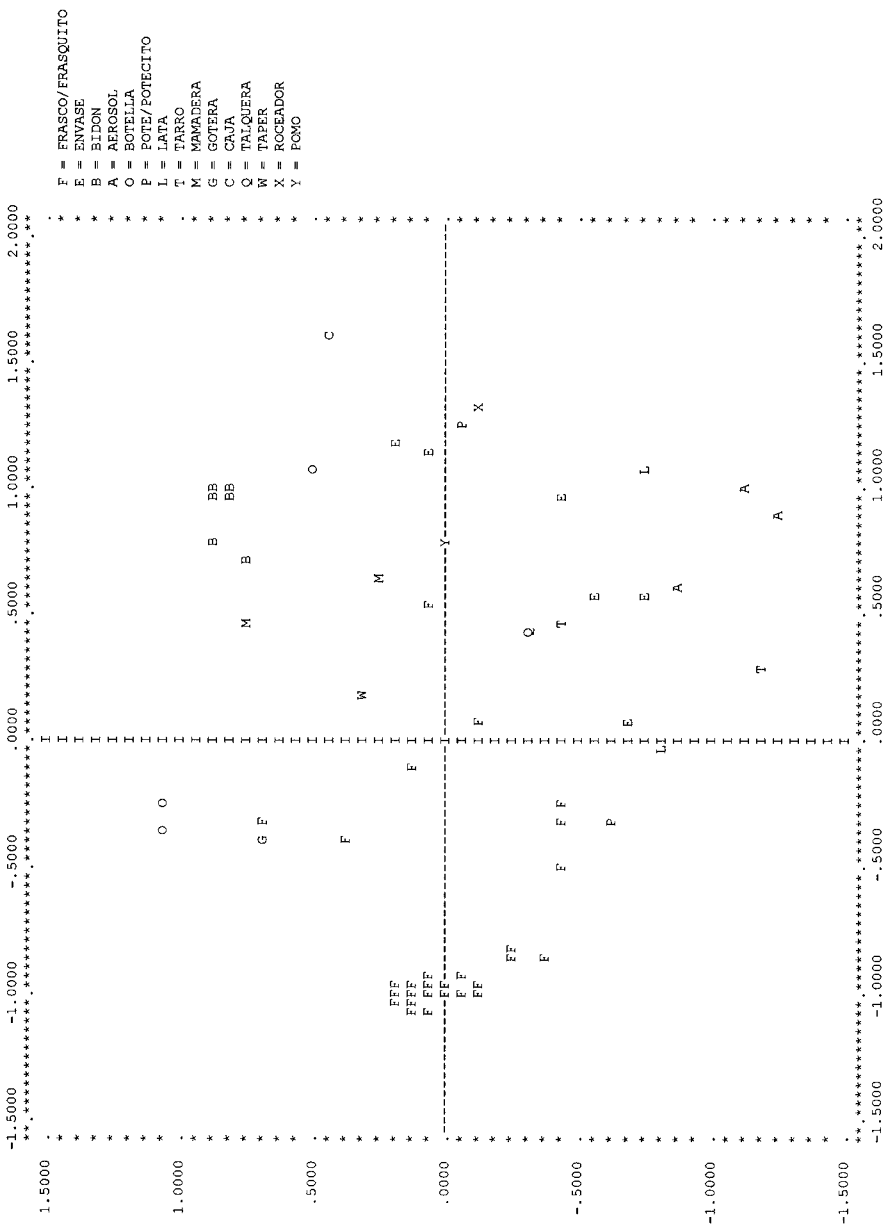


**FIG. 13.** MDS solution for Spanish overall sorting data.

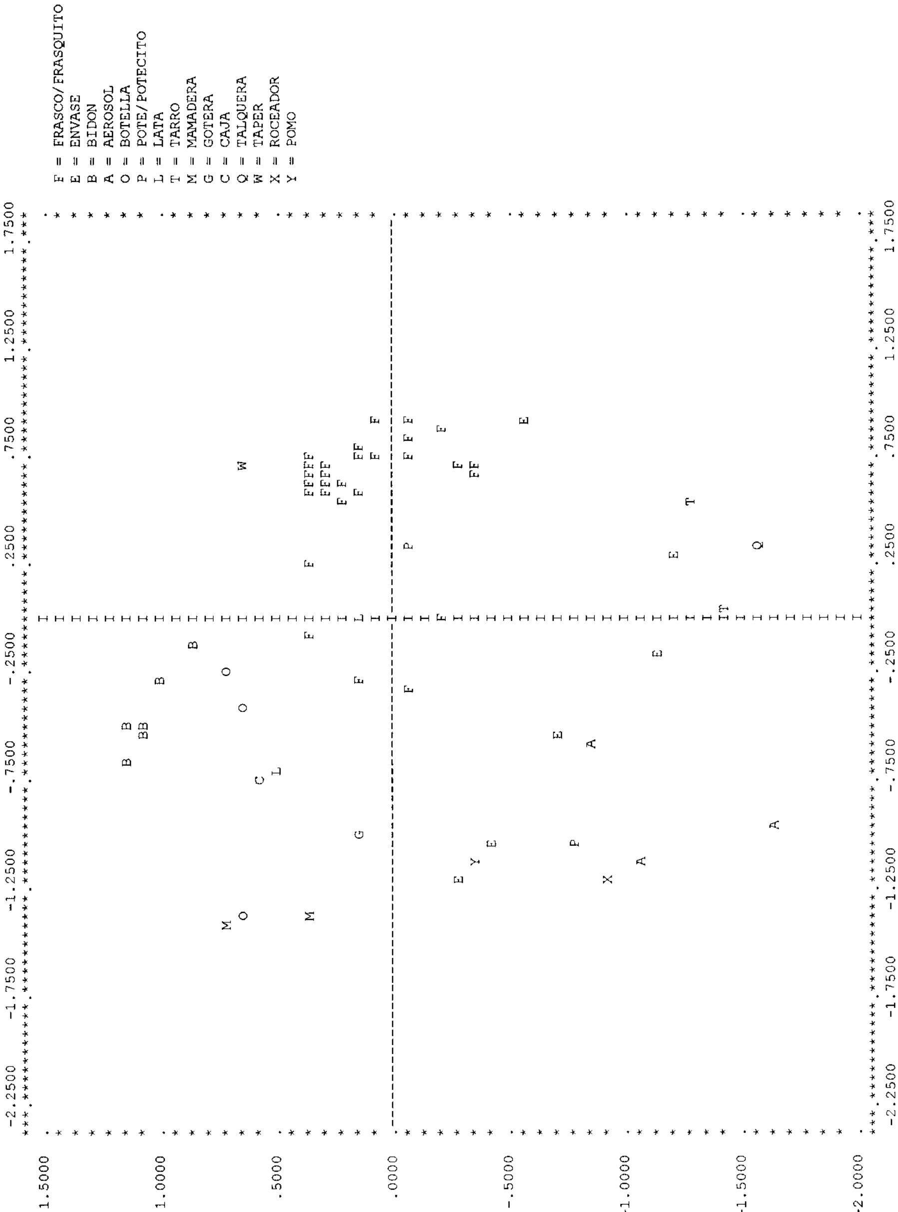


**FIG. 14.** MDS solution for Spanish functional sorting data.

TABLE 4

Correlations between Name Similarity and Sorting for Americans, Chinese, and Argentineans

| | Naming | | |
|---|---|---|---|
| | American | Chinese | Argentinean |
| Overall sorting | | | |
| American | .68 | .44 | .76 |
| Chinese | .59 | .54 | .77 |
| Argentinean | .69 | .46 | .81 |
| Physical sorting | | | |
| American | .70 | .43 | .71 |
| Chinese | .65 | .47 | .68 |
| Argentinean | .65 | .48 | .78 |
| Functional sorting | | | |
| American | .48 | .27 | .55 |
| Chinese | .34 | .13 | .32 |
| Argentinean | .57 | .42 | .79 |

sure described above; the Euclidean measure shows similar results) with the similarity measure derived from the sorts over the 1770 object pairs, for speakers of the different languages. If naming and the categories revealed by sorting are closely linked, then name similarity for a given language should correlate with sorting by speakers of that language, and it should correlate more strongly than with sorting by speakers of other languages. The relevant correlations are shown in Table 4 for the three different kinds of sorts.

Before addressing the central hypothesis, we make some general observations about these correlations. First, they are fairly substantial for overall and physical sorting, and intermittently so for functional sorting (especially for the Argentineans). This fact indicates a definite relation between naming and similarity. Second, the correlations are nevertheless far from perfect. As we can already deduce from the multidimensional scaling solutions and our earlier observation of large differences between the groups in naming but relatively small differences in sorting, the two kinds of categorization are not the same. Naming must be influenced by factors beyond simple similarity. Finally, the Argentinean sorts corresponded slightly better with naming than either American or Chinese sorts did in 7 out of 9 cases, and the Argentinean names had the strongest correspondence to sorting of the three languages in 8 out of 9 cases. The latter result may reflect the fact that there were more name categories in Spanish, which may provide a more sensitive measure (a suggestion supported by the fact that Chinese, which has the smallest number of name categories, has the weakest correspondence). Whatever the reason for this pattern, in any case, the advantage for Spanish is small.

We now ask whether name similarity for a given language correlates more highly with sorting by speakers of that language than with sorting by speakers of other languages. If so, this fact would support the existence of a parallel between naming and perceived similarity. In the case of overall similarity, Argentinean and American sorting were essentially equal predictors of American name similarity; the best predictor of Argentinean naming was Argentinean sorting; and the best predictor of Chinese naming was Chinese sorting. Similarly, for physical sorting, the best predictors for American and Argentinean naming were American and Argentinean sorting, respectively, and for Chinese naming, Chinese sorting was a close second to Argentinean. Hence, for these two sorting types, the data suggest some correspondence between sorting and naming that is tied to the specifics of a language culture. Note, however, that the correspondence is weak in that the predictive advantage gained is small relative to the overall predictive power of sorting by speakers of all the languages and it is not robust, failing to appear in two out of six comparisons. In the case of functional sorting, the advantage disappears altogether: Argentinean sorting is the best predictor of the American, Chinese, and Spanish name categories. Overall, these results indicate that there may be a link between differences in perceived similarity and differences in naming for speakers of the three languages, but any such link is small compared to the extent of the divergences in naming. The results are consistent with the preceding analyses in indicating that there is partial independence of naming and similarity.

Finally, we further examined the relation between naming and perceived similarity by looking at the relation between naming distributions for each pair of languages and perceived similarity by speakers of the two languages. The calculation of name distribution similarity for pairs of objects reported earlier provides a measure of the extent to which the members of a pair receive the same name in a given language. Subtracting the name similarity value for each pair in one language from the value for that pair in the other language, then, provides an indication of the extent to which the two languages have comparable degrees of name similarity for each pair. Likewise, the sorting values for each pair provide a measure of the extent to which members of a pair are perceived as similar by speakers of a given language. Subtracting the perceived similarity value for each pair in one language from the value for that pair in that the other language provides an indication of the extent to which speakers of the two languages perceive comparable degrees of similarity for each pair. These two derived measures can be correlated to see whether pairs for which the two languages diverge in name similarity are also pairs for which speakers of the languages diverge in perceived similarity, and the converse. If naming and perceived similarity parallel each other, then where languages diverge from each other in naming of pairs, they should also diverge from each other in perceived similarity, and where they match in the naming of pairs, they should match in perceived similarity.

The correspondences revealed by this procedure were positive but only weak to moderate. For physical sorts, the correlations between the two derived measures were .36 for American and Chinese, .14 for American and Spanish, and .32 for Chinese and Spanish. For overall sorts, the correlations were .22 for American and Chinese, .26 for American and Spanish, and .10 for Chinese and Spanish. For functional sorts, they were .01 for American and Chinese, .17 for American and Spanish, and .04 for Chinese and Spanish. This result supports the conclusion that differences in naming among the languages are only partially related to differences in perceived similarity, and naming and perceived similarity show substantial independence.

## GENERAL DISCUSSION

### *Summary*

Analyses of dominant names and of similarities among name distributions, and application of the Cultural Consensus Model, all revealed substantial differences in American, Chinese, and Argentinean linguistic categories for 60 containers. In contrast, the three groups showed only small differences in their perceptions of the similarity among the objects, as revealed by correlations across the sorts and by the CCM: No systematic differences were observed in overall similarity, and the differences in physical and functional sorts were significantly smaller than those observed in naming. MDS solutions show an imperfect relation between naming and similarity, and two correlational measures indicated that some correspondence does exist between specific differences in naming and in perceived similarity across the groups, but this correspondence is not strong. Our data therefore indicate that although perceived similarity and naming show a positive relation, this relation is far from perfect, and factors other than similarity must contribute to naming choices.

### *Accounting for Divergences between Similarity Clusters and Naming*

In predicting partial independence of naming from similarity, we proposed the existence of several mechanisms that might result in complexity in how objects are named. Our data allow us to look for examples of the mechanisms at work. We looked for cases in which (a) an object's mean similarity to members of a different linguistic category is equal to or greater than its similarity to its own, or (b) its nearest neighbor (the object with which it has been grouped most often in the sorting tasks) is from a different linguistic category than its own. We give two examples here that illustrate the three proposed mechanisms.

*Convention and pre-emption.* A possible illustration of naming influenced by both conven-

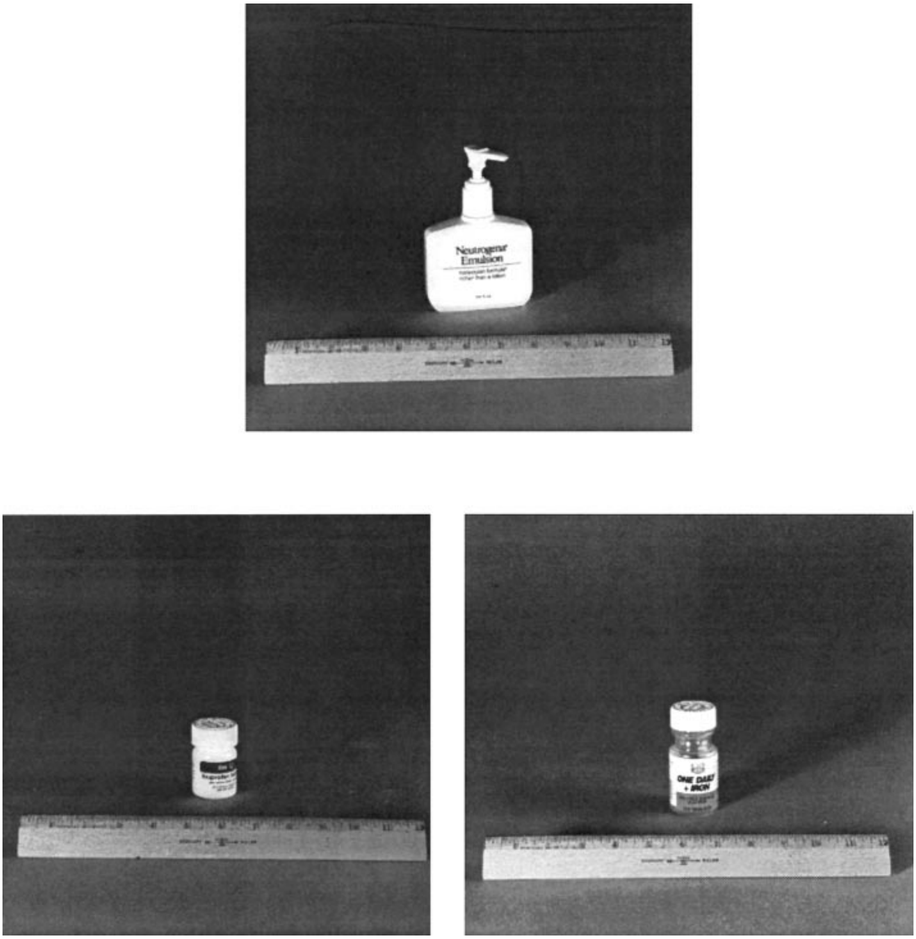


**FIG. 15.** Upper panel: an object named “container” with higher average similarity to bottles. Lower panel: two objects named “bottle” with higher average similarity to containers.

tion and pre-emption is a squarish plastic object with a pump top holding hand lotion (see top panel, Fig. 15). This object’s dominant name was “container,” but its average similarity was greater to bottles than to other containers in all three sort types, and its nearest neighbor in each sort was a bottle. This object may therefore be named “container” and not “bottle” by convention, rather than being driven by similarity. The origin of this convention may have been a pre-emption: Many hand lotion containers are normally called bottles, having forms substantially more like typical bottles. The less common form of a container for hand lotion may therefore have acquired the label “container” to distinguish it from those types of hand lotion bottles, whose form would ordinarily be inferred from the phrase “hand lotion bottle.” Included among the stimuli that *did* have “bottle” as their dominant name were a plastic pump-top holder of spray cleaner and a rectangular plastic holder of shampoo, so neither the shape nor the pump top per se is sufficient to exclude objects from the category of “bottle.” (These latter objects presumably are not pre-empted from the “bottle” name because calling them “bottle” does not cause undue confusion with other objects holding spray cleaner or shampoo).

*Chaining.* Possible cases of chaining are two small plastic objects with screw-off lids the same width as the rest of the object, one holding vitamins and the other aspirin (see bottom panel, Fig. 15). Both received the dominant name "bottle" but were more similar on average to objects labeled "container" in all three sorts. We suggest that both of these objects receive their name through links to more typical objects in the "bottle" category: Pills of all sorts, including aspirin and vitamins, have traditionally come in containers called "bottle" with more typical bottle forms. (Ideally, one would be able to identify this or other cases of links in chains in the similarity data. In practice, though, because we could not possibly include every type of object belonging to each category, the links are not necessarily represented among our stimuli). Since the phrases "aspirin bottle" and "pill bottle" are so familiar (and, to a lesser extent, "vitamin bottle" or "bottle of vitamins"), phrases such as "aspirin container" or "vitamin container" for new variants of those objects may sound odd, and the habitual name carries over through their relation to the older, typically bottle-like versions. In fact, a different object virtually identical in size and shape to the aspirin bottle but holding dried basil was called "container" and not "bottle." The difference in naming may be because there is no well-established phrase "bottle of basil" or "basil bottle" to draw this object into the "bottle" category.

### *Implications for the Whorfian Hypothesis and Related Views*

The Whorfian hypothesis would presumably predict a substantial influence of linguistic categories on a person's perception of the similarities among objects (at least as psychologists have traditionally interpreted the hypothesis, whereby the structure of a language strongly influences or even fully determines the way its native speakers perceive and reason about the world; e.g., Brown, 1976; Slobin, 1979; cf. Lucy, 1992). Our data do not allow us to establish the causal direction of the modest correspondence we did find between differences in naming and perceived similarity by speakers of different languages. However, the data do imply that linguistic categories cannot be the only determinant of perceived similarity among these objects. The magnitude of our correlations suggest that linguistic categories are not even the primary determinant of perceived similarity. Our data, if anything, suggest that perception of the similarity among objects remains relatively constant despite wide variation in linguistic category boundaries.

Similarly, the more recent work on naturally occurring categorical perception (e.g., Bornstein, 1987; Eimas, Miller, & Jusczyck, 1987; Pastore, 1987; Repp, 1984) and related laboratory-induced effects (e.g., Goldstone, 1994a; Livingston et al., in press) suggests that category boundaries should substantially influence the perception of similarity for pairs of objects within a category versus those crossing a boundary. Our data are only minimally consistent with this suggestion. However, our study differs from these bodies of work in several important ways. In studies of color and phonemes, the stimuli are simple ones requiring mainly perceptual processing, and the categories appear to arise at least in part because of nonlinearities in people's sensitivity to the physical dimensions of the stimuli (e.g., Kay & McDaniel, 1978; Rosen & Howell, 1987; Kuhl, 1987). In contrast, our stimuli are complex objects, and naming them requires conceptual processing such as integrating information about physical and functional properties. If color and phoneme effects stem mainly from relatively low-level perceptual processing, then there is no reason to expect them to occur for complex stimuli such as ours.

As for the difference between our results and those for artificial stimuli, the paradigms again differ fairly substantially. In studies with artificial categories, category boundaries are learned over a short period of time for stimuli that are difficult to discriminate and categorize (e.g., squares varying in brightness and saturation (Goldstone, 1994a); abstract patterns (Livingston et al., in press)). In these cases, learners most likely use the information about category membership to guide their search for features that discriminate the categories. In doing so, they may give more weight to differences among stimuli from different categories than to differences within categories. In contrast, because the linguistic categories we studied have members overlapping with each other on different features and sometimes overlapping more with members of other categories, learners may not be able to find fixed boundaries on any set of dimensions that would yield correct categorization. Furthermore, these categories are learned under circumstances where communication, not classification per se, is the goal. Although a child learning the meaning of “bottle” or “jar” may look for features to associate with the word, he or she is also no doubt inclined to adopt names provided by adults (e.g., Mervis, 1987, p. 227). A child may be happy enough to call a plastic bear with a straw a “juice box” simply because Mom or Dad has called it that, and calling it that results in successful reference. Adults, likewise, may tend to adopt an object label provided by an advertiser or other source just because that label is known to and used by others (e.g., Clark & Wilkes-Gibbs, 1986; Markman & Makin, in press). In these cases, learners are exposed to different uses of a category name over long periods of time rather than en masse, and they may make no effort to adjust perception of other objects whose category membership is already well learned. Finally, as our naming data show, many of our objects received more than one name across the subjects. Although our data do not directly speak to the issue of intra-person variability, a given person may use one name on one occasion and a different name on a different occasion. Variability in naming an object may make it impossible to systematically adjust perceived within- and between-category differences.

### *Relation to the Categorization of Natural Kinds*

Although our study focused on artifacts, another important domain of categorization is natural kinds. We believe that the mapping between recognition and naming may generally be more straightforward for natural kinds than for artifacts, but the distinction between the two acts of categorization may still be useful for natural kinds. In the case of artifacts, because their features can be changed and recombined freely, it is not uncommon for artifacts to fall into areas of similarity space not closely associated with any cluster of objects, or between two different clusters, or at the end of a chain of objects. Such instances provide many of the opportunities for complexity in naming to arise. In contrast, natural kinds may tend to occur in tighter clusters with relatively few cases of unclear affiliations (Berlin, 1992; Hunn, 1977; see Malt, 1995). If so, names may generally be applied to clusters of objects with many properties in common. Still, there may be some cases of natural kinds where names are influenced by factors other than similarity per se. For instance, Malt (1994) has observed that the main ingredient in many beverages such as tea, coffee, and Sprite is the same ($H_2O$) as in other liquids that are called “water” in English (e.g., tap, sea, and sewer water), yet the first three are given unique names. Differentiating the domain of beverages is more important for most Americans than differentiating other subsets of $H_2O$-based liquids, which may be the reason that contrastive names arise for beverages but not for many others. This suggestion is consistent with the notion of pre-emption: If calling several $H_2O$-based drinks “water” would lead to referential confusion, they may be given unique names instead. In line with Malt’s (1994) suggestions, then, it may be useful to consider the influence of communicative factors on how natural kinds are named.

### *Implications for Theories of Categorization*

There are two major implications of our results for theories of categorization. First, the divergences between perceived similarity and naming may help explain conflicting perceptions of category complexity. On the one hand, the intuition that categories have clear and specifiable boundaries has led to attempts to pin down the way that membership in artifact categories is determined (e.g., Medin & Otony, 1989; Keil, 1989; Bloom, 1996). On the other hand, no empirical investigation to date has identified information that might serve to fully constrain membership in these categories (e.g., Rosch & Mervis, 1975; Malt & Johnson, 1992; Hampton, 1995). These proposals and investigations have not distinguished between categorization as mentally encoding an object and categorization as choosing a name for an object. Recognition categories, which we have suggested are defined as clusters of objects in similarity space, may be relatively simple categories by their very nature: They consist of objects that share many properties. Linguistic categories, on the other hand, may be much more complex because of the various mechanisms that influence their composition, as we have argued. These differences between the natures of recognition and linguistic categories may result in situations where one has a strong intuition, for instance, that boxes are things that are wood or cardboard and rectangular, but at the same time, one is willing to call a variety of other objects by the name "box" in the course of communicating about them. Distinguishing between the two different acts of categorization, and tapping them with methods appropriate to the act of interest (i.e., with nonlinguistic measures if recognition categories are of interest, and with naming in naturalistic contexts if linguistic categories are of interest) may help resolve this conflict between the intuition that artifact categories should be definable and the empirical difficulty in doing so.

The second, and most fundamental, implication of our results is that, although we have followed tradition in using the terms "category" and "categorization" up to this point, it may actually be more appropriate to dispense with the notion of categorization of artifacts altogether. We have suggested that categorization in the sense of recognizing an object's kind is a matter of internally representing it along with similar objects. This process of representation does not, by hypothesis, involve selecting among competing natural language labels such as "bottle" and "jar" for the object. There is also no reason to think that boundary lines of a nonlinguistic sort are necessarily drawn within the representational space. Conceptual processes such as inferring unseen attributes of a new object through its similarity to other objects do not require any such boundaries (Sloman, 1993). If recognition boundaries do not exist, then recognition does not involve categorization in the usual sense of choosing among competing classes or designations. We suggest that questions about recognition may therefore be better phrased as questions about perceiving and representing objects. At the same time, at the linguistic level, a given object clearly can receive different acceptable names from different people, as our data showed both in the existence of naming distributions across subjects speaking the same language, and in the substantially different patterns of naming between languages. Given this fact, the notion of fixed linguistic categories may also be illusory. Attempts to define or describe linguistic categories per se may therefore be less useful than attempts to determine when and why use of a particular name for a particular object will be seen as acceptable.[6] Investigating what people do with artifacts less in terms of how they put objects into fixed sets, and more in terms of the processes of perceiving, representing, and communicating, may be a more fruitful approach for future research.

## *Conclusion*

We have argued that it is important to distinguish between recognizing objects and naming them because the relation between the two may not be as straightforward as has often been assumed in the past. Our data support both parts of this contention by showing that people who speak different languages may have substantially different patterns of naming for a set of objects while seeing the similarities among the objects in similar ways; the patterns of naming therefore cannot arise only from the similarities people see among the objects. We have also offered suggestions about how complexity in naming may arise, and the data provide some evidence consistent with these suggestions. Exploring how artifacts are named vs. "known" may provide new insights into artifact categorization.

## REFERENCES

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

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.). *Theories of memory* (pp. 29–101). Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] Doing so may involve reference to category prototypes, best examples, or heavily weighted features; we do not mean to suggest that there is no stable or central component of linguistic category representations. We do suggest, though, that attempts to find a fixed set of properties or an "essence" that will account for all uses of an artifact name are not likely to be successful.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition,* **57,** 217–239.

Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies.* Princeton, NJ: Princeton University Press.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition,* **60,** 1–29.

Bornstein, M. H. (1981). Psychophysical studies of color perception in human infants: Habituation, discrimination, and categorization, recognition, and conceptualization. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 1). Norwood, NJ: Ablex.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535–565). New York: Cambridge University Press.

Brown, R. W. (1976). Reference: In memorial tribute to Eric Lenneberg. *Cognition,* **4,** 125–153.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology,* **49,** 454–462.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America,* **63,** 456–468.

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

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language,* **15,** 317–335.

Clark, E. V. (1993). *The lexicon in acquisition.* Cambridge: Cambridge University Press.

Clark, H. H. (1991). Words, the world, and their possibilities. In G. R. Lockhead and J. R. Pomerantz (Eds.). *The perception of structure: Essays in honor of Wendell Garner.* Washington, D.C.: American Psychological Association.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding.* (pp. 10–63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition,* **22,** 1–39.

Eimas, P. D., Miller, J. L., & Jusczyck, P. W. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 161–195). New York: Cambridge University Press.

Eiser, J. R. (1996). Accentuation revisited. In W. P. Robinson (Ed.). *Social groups and identities: Developing the legacy of Henri Tajfel.* Oxford: Butterworth-Heinemann.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology,* **18,** 500–549.

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

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition,* **38,** 213–244.

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language,* **27,** 166–195.

Goldstone, R. L. (1994a). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General,* **123,** 178–200.

Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition,* **52,** 125–157.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language,* **21,** 125–155.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language,* **34,** 686–708.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology,* **93,** 10–20.

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology,* **24,** 341–380.

Hunn, E. (1977). *Tzeltal folk zoology: the classification of discontinuities in nature.* New York: Academic Press.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development,* **8,** 113–139.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist,* **86,** 65–79.

Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language,* **54,** 610–646.

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

Kronenfeld, D. B., Armstrong, J. D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extensionist semantic view. In J. W. D. Dougherty (Ed.). *Directions in cognitive anthropology* (pp. 91–113). Urbana: University of Illinois Press.

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

Kuhl, P. K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 355–386). New York: Cambridge University Press.

Kurtz, K. J. (1996). Category-based similarity. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (p. 790). Hillsdale, NJ: Erlbaum.

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

Lantz, D., and Stefflre, V. (1964). Language and cognition revisited. *Journal of Abnormal and Social Psychology,* **69,** 472–481.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics,* **1–2,** 207–246.

Lenneberg, E. H., & Roberts, J. M. (1956). The language of experience: A study in methodology. *International Journal of American Linguistics,* **22** (2, part 2, Memoir 13).

Livingston, K. R., Andrews, J. K., & Harnad, S. (in press). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

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.

Markman, A. B., & Makin, V. S. (in press). Referential communication and category acquisition. *Journal of Experimental Psychology: General.*

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology,* **20,** 121–157.

McCawley, J. (1978). Conversational implicature and the lexicon. In P. Cole (Ed.). *Syntax and semantics 9: Pragmatics.* New York: Academic Press.

McGarty, C., & Penny, R. E. (1988). Categorization, accentuation, and social judgment. *British Journal of Social Psychology,* **27,** 147–157.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review,* **100,** 254–278.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning.* New York: Cambridge University Press.

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

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization (pp. 201–233). Cambridge: Cambridge University Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review,* **92,** 289–316.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review,* **101,** 53–79.

Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 29–52). New York: Cambridge University Press.

Repp, B. H. (1984). Categorical perception: Issues, methods, and findings. *Speech and Language: Advances in Basic Research and Practice,* **10,** 243–335.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* New York: Cambridge University Press.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General,* **122,** 468–496.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist,* **88,** 313–338.

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

Rosen, S., & Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 113–160). New York: Cambridge University Press.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research,* **10,** 489–502.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika,* **39,** 373–421.

Slobin, D. I. (1979). *Psycholinguistics (2nd Edition).* Glenview, IL: Scott Foresman and Company.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology,* **25,** 231–280.

Sloman, S. A., & Ahn, W-k. (in press). Feature centrality: Naming versus imagining. *Memory & Cognition.*

Sloman, S. A., & Rips, L. J. (Eds.) (1998). *Cognition (Special Issue): Rules versus similarity in human thinking,* **65,** 87–302.

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

Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In B. Burns (Ed.), *Percepts, concepts, and categories: The representation and processing of information* (pp. 233–272). Amsterdam: Elsevier.

Smith, L. B., & Samuelson, L. K. (1997). Perceiving and remembering: Category stability, variability, and development. In K. Lamberts and D. Shanks (Ed.). *Knowledge, concepts, and categories* (pp. 161–195). Cambridge, MA: MIT Press.

SPSS (1990). *SPSS Reference Guide.* Chicago: SPSS Inc.

Tajfel, H. (1957). Value and the perceptual judgment of magnitude. *Psychological Review,* **64,** 192–204.

Tajfel, H. (1959). Quantitative judgement in social perception. *British Journal of Psychology,* **50,** 16–29.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf.* J. B. Carroll (Ed). Cambridge, MA: MIT Press.

(Received October 28, 1997)
(Revision May 27, 1998)