1999

Artifact categories : evaluating theories of linguistic categorization

Molly J. Stanton
Lehigh University

Stanton, Molly J.

Artifact categories; evaluating theories of linguistic categorization

May 31, 1999

Artifact Categories: Evaluating Theories

of Linguistic Categorization

by

Molly J. Stanton

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Psychology

Lehigh University

April 22, 1999

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

April 15, 1999
(Date)

(Advisor)

(Co-advisor)

(Co-advisor)

(Department Chair)

I wish to thank Christine Matarese and Stephanie Ojeda, undergraduate assistants who assisted me in data collection for this thesis. I would also like to thank my advisor Barbara Malt and committee members Sue Barrett and Martin Richter for their guidance and patience throughout the duration of this project. Finally, I wish to extend my thanks to my parents Michael and Rebecca Stanton for steadfastly listening to all my unrelenting complaining about the various hurdles encountered during this project.

Table of Contents

List of Tables

List of Figures

Abstract

Artifact categories prove to be an interesting challenge for researchers interested in categorization as it may be impossible to list necessary and sufficient features for them, contrary to defining features theories. On the other hand, theories employing "fuzzy" concepts are thought to be too unconstrained to account for categorization (Murphy & Medin, 1985). An alternative view detailed by Malt, Sloman, Gennari, Shi, and Wang (1999) is presented. The Malt et al. view suggests that there may be two processes which can be thought of as categorization: naming and recognition. Recognition is driven mainly by similarity, while the communication process which governs naming has its own processes which may be separate from recognition. This research addresses a fundamental question raised by the view. Specifically, what sorts of features separate one category from another: functional, physical, or some combination of multiple features? In my study I attempt to separate the names from the recognition categories and begin to examine the extent to which similarity accounts for naming. Core theories dictate that functional features will be the only features important in categorizing artifacts. If function is indeed the core then consideration of functional features will fully separate linguistic categories. If, however multiple types of features are needed to account for linguistic categorization, support will be lent to the Malt et al. view. This research demonstrates that functional features are not the only features important in categorization and sheds doubt on core theories.

Artifact Categories: Evaluating Theories of Linguistic Categorization

Artifact categories have proved to be an interesting challenge for researchers interested in categorization. This is due in part to the difficulty in listing necessary and sufficient features for artifact categories. Whereas with natural kinds the argument can be made that there may be some sort of defining feature that all members share, for artifacts, this may not be the case. For instance, it might be said that ostriches are members of their species by virtue of the fact that all ostriches share a genetic code which is the necessary "feature" (Keil, 1989). Artifacts arguably do not possess such hidden properties like genetic code that make them members of a kind. There is no hidden property that all chairs share that makes them chairs. Also, artifact categories are more variable within the category than members of natural kind categories are (Malt, Sloman, Gennari, Shi, & Wang, 1999). For instance, the category "robin" is a natural kind category; members of the category do not show much variability. All robins have similar body structures, organs, and genetic structures. The artifact category " cereal bowl", on the other hand, shows a lot of variability in its members, and new shapes and styles of cereal bowls are created regularly. Despite variability and the ever changing nature of artifact categories, we are still able to distinguish objects called "bowl" from those called "dish" and objects called "dish" from ones called "plate". Extant theories of categorization are unable to account for how we do this. Below, I discuss major theories of categorization and how they require revision to be able to account for various data.

Existing theories of categorization tend to cluster around two poles. On the one side are theories postulating cores of defining features which uniquely determine a category, and on the other side are theories which postulate "fuzzy" concepts which probabilistically specify a category. I will discuss how both types of theories relate to artifacts, beginning with core theories.

Core Theories

Classical view

The classical view of categorization is a good example of a core theory (Smith & Medin, 1981). The classical view postulates a set of features which define a category and only that category. This set of features *must* be possessed by each member, and not by members of other categories; that is, these features are necessary and sufficient to determine category membership. Because each member must possess the category's necessary and sufficient, or defining features, every instance of the category should be an equally "good" instance. Pillow or beanbag chairs should be no worse an example of a chair than a familiar dining room chair. Thus, under this theory, there is no way to have an atypical member of a category and still have the member belong to the category (Lakoff, 1987; Murphy & Medin, 1985; Smith & Medin, 1981). One cannot say that the pillow chair is a chair but not a typical one even though it intuitively makes sense that a pillow chair is still a chair, but a rather atypical one.

The classical theory soon runs into additional problems when we consider whether categories do have a set of invariant features associated with them. Let's return to the example of the category "chair"; a list of necessary and sufficient features for the category might include, for sitting on, has three or four legs, and is made of some sort of hard material, like wood or metal. Thus, in order for something to be called a chair, it must have all three of these characteristics, and no other object may have them. But pillow chairs are categorized as chairs and they neither have three or four legs nor are they made of hard material. Furthermore, what of a stool, which shares many properties of "chairness" (e.g. it is for sitting on and is made of hard material), but is not explicitly called a chair? As this example illustrates, it becomes exceedingly difficult to come up with a set of features which is true for all members of one category and that members of no contrasting category share (Lakoff, 1987). In fact, despite many years of searching, psychologists and philosophers have been unable to come up with any necessary and sufficient features for categories (Murphy & Medin, 1985; Smith & Medin, 1981). This, by itself, does not spell the end for the classical theory because one could argue that the defining feature will be obscure or extremely abstract and we simply have not discovered it. This is precisely the line of reasoning the essentialist view takes. However, I will first discuss the research against the classical view, then I will return to essentialist claims.

Empirical evidence against the classical view

There are several sets of empirical studies which do seem to suggest that the classical theory is untenable as a theory of artifact categories; I will discuss two of them. First, Rosch and Mervis (1975) demonstrated that variations in the typicality of category members are correlated with variations in the distribution of the features associated with categories. Specifically, items rated by participants as more typical also tended to have more features that were associated with the category. Conversely, objects rated as less typical had fewer of the categories' associated features. This is counter to the classical view because it logically necessitates that all category members have the category's features, and it follows that they should all then be equally typical. But the results also counter the classical view on another front; namely, the correlation implies that there is something important to categorization in those features associated with the category with less than 100 percent probability. The second set of studies I mentioned address this possibility more fully.

Hampton (1979, 1981) presents experimental findings suggesting that people use what proponents of the classical view would call unnecessary features in making category decisions. Hampton (1981) asked one group of participants to list features of objects and rate the extent to which the feature was predictive of category membership. Next, the list was used to predict categorization reaction times for a second group of participants. Several critical results came from this study. First, some of the features listed were non-defining features. Second, these non-defining features were seen as important for categorization and were correlated with the reaction times. Reaction times provide an index of categorization performance. The presence of a common, yet non-defining feature made categorization times quicker. For example, subjects might say that the fact that a chair is made of wood is important for categorizing it as a chair, and participants may be faster at making a category decision about a given chair that is indeed made of wood based on the wood feature. But obviously, not all chairs are made of wood, hence wood is not a necessary feature of chairs. If the defining features of the objects participants were asked to judge were really the only important features for categorization, then these features should have been the only ones that facilitated the reaction times of the category decisions. In contrast, many non-defining features facilitated reaction times in category decisions. Clearly the classical view does not allow non-necessary features to be important in making categorization decisions because it is precisely the necessary features and only the necessary features which define the category.

As a corollary, the necessary and sufficient features assumption in the classical theory implies that the boundaries between different categories will be discrete. If all instances of a category and no members of contrasting categories have these features, then the boundaries of the categories should be all or nothing. Either a given object has the defining features and belongs to a given category or it does not and belongs to another category. The research mentioned above by Rosch and Mervis suggests that this prediction of the classical theory also does not hold. Category boundaries appear to be fuzzy.

<u>Essentialism</u>

As mentioned previously, there is a second kind of core theory which purports to solve some of the problems of the classical view. This view is called psychological essentialism. Psychological essentialism claims that while it may not be possible to come up with a *set* of features objects in a category share, one may argue that objects within a category may share some hidden property that makes them all members of a particular category. This hidden property could be a genetic structure or chemical composition, that is not obvious to the categorizer. It is the non-obviousness of the essence which explains why people are not easily able to point to a defining feature when asked to do so in categorization tasks (Smith & Samuelson, 1998).

Essentialism is similar to the defining features view in that possession of the "essence" determines category membership. However, it differs from classical defining features views in that the feature that makes objects category members is a hidden property and average people may know little or nothing about it. It should be noted that most essentialist theories do not make the strong claim that these essences reflect objective truths about the world, rather, it is enough that people behave as though they believe that there are essences which unambiguously define a category (Malt, 1994).

Indeed, for at least some natural kinds the essentialist argument may account for how people categorize. Keil (1989) asked participants whether a raccoon painted up to look exactly like a skunk was still a raccoon or had become, instead, a skunk. The participants stated that the animal was still a raccoon even though outwardly it was indistinguishable from skunks. Keil proposed that the participants believed something inside the animal, like genes, made it a raccoon despite changes in outward appearance.

Evidence against essentialism

The essentialist view may sound as though it is an improvement over the classical view. After all, it does seem to provide a plausible explanation of why a set of "obvious" necessary and sufficient features defining a single category was never found. However, the essentialist view does not perform as well as a cogent theory of artifact categories. Function is generally believed to be the essence of artifact categories. Similarly, in recent years function has also been proposed as the core property of artifacts for classical theories as well. Malt and Johnson (1992) demonstrated that essentialism is not a valid theory for artifact categories. The researchers performed a set of studies aimed at determining whether artifact categories have cores. Specifically, they tested whether having the function normally associated with an object is sufficient for determining category membership. They also tested whether having a particular function is also necessary for category membership.

With respect to the former, Malt and Johnson created a set of common objects, some with their usual functions but with new physical features instead of the physical features normally associated with the object and some objects with their usual functions and usual physical features. If function were sufficient for category membership, then the new physical features should have no effect on participants' judgments of category membership. Put another way, consideration of physical features would add nothing to category decisions; they should be accounted for purely by determining the object's function. This was not the case. The objects with the unusual physical features added had fewer participants place them in their usual category than did the unaltered objects, demonstrating that physical features did make a difference in category judgments and hence, function was not sufficient for category membership.

To test whether possession of a particular function is necessary for category membership, Malt and Johnson compiled a set of object descriptions where some objects had normal physical descriptions paired with normal function, almost normal or related function, or bizarre function. If having a particular function were necessary for category membership, then objects with a related but different or abnormal function should not be given the category label. Indeed those objects with normal functions were almost unanimously stated to be members of the category. However, more than half of the objects, both those with related functions and bizarre functions, were also rated as members of the normal category. These results suggest function in artifacts is also not necessary for determining category membership.

Fuzzy Concept Theories

Over time, evidence from studies such as the one mentioned above has mounted which suggests that categories do not have a discrete all or none structure as implied by defining features views, or unchangeable internal essences that define categories. Rather, they have an analog or graded structure (Rosch & Mervis, 1975). Put another way, there are no features or essences common to every single category member, but there are features associated with the category as a whole. Category members typically possess a subset of the associated features. With some kinds of fuzzy concept theories, certain members of a given category may possess many of the features associated with the category; these members often become category prototypes.

These properties suggest that, at the other extreme, those theories postulating fuzzy concepts may account for more of the categorization data. There are two main "fuzzy" concept theories, family resemblance theories (Rosch & Mervis, 1975) and prototype theories (Reed, 1972). Both these theories maintain that categories have structures of probabilistically associated features as described above.

Family resemblance and prototype views

In a family resemblance structure, categories are viewed, not as discrete bounded entities, but as networks of overlapping attributes. Objects will be placed in a given category to the extent that they have features which overlap with other members of that category. Most importantly, some objects within a category may have many features in common with other members of the same category, but because they possess some features not shared by many other category members, they are removed to varying degrees from the objects at the center of the category which do possess many of the category's associated features. Going back to our "chair" example, pillow chairs are in the chair category even though they share fewer of the features associated with chairs than a familiar dining room chair does. By this example, we get the impression that perhaps actual category boundaries are not as sharp as the classical or essentialist view dictates.

The prototype theory is similar to the family resemblance theory except that the prototype model has a central tendency which gets labeled as the category prototype. This is a key difference. In both prototype and family resemblance models there are multiple features shared by members of the category, and not every instance will have every feature, but with the prototype model, people abstract a prototype from repeated exposure to individual exemplars. Prototypical instances of a category will be precisely those members which have many features in common with their own category and few features in common with other categories (Rosch & Mervis, 1975). Category membership is a function of the overlap of a possible specific instance of the category with the prototype. Let's take an example for clarification. Dining room chairs are highly typical chairs because they have many features in common with the chair prototype and few with other categories like "appliances" or "tools". Thus how likely a given object is in the category would depend on the degree to which it overlaps with the category prototype. The more features a given object has in common with the category prototype, the more likely it is to be considered a member of that category.

To sum up, in a family resemblance framework no abstraction process for creating a prototype is postulated. Instead, there are multiple members with a pool of shared features. Both prototype and family resemblance models predict graded category membership.

Evidence against family resemblance and prototype views

Family resemblance and prototype theories can account for graded category boundaries; however, it has been argued that the family resemblance view is too unconstrained to be a candidate for how humans categorize. This is because little evidence exists to tell us what features are relevant to a category. Here the use of the term unconstrained means that by themselves, family resemblance views do not tell us which sets of features form possible categories and which form incoherent ones (Murphy & Medin, 1985). For instance, the formulation of the family resemblance view cannot tell us whether one collection of features forms a more coherent category than another. How are we to decide which collections of features are the correct ones?

The family resemblance and prototype theories also fall prey to an argument related to one of the arguments leveled against the core theories. With core theories years of investigation yielded not a single unambiguous example of a defining set of features. With prototype and family resemblance theories no one has yet come up with a clear accounting of the internal structure of a single category (Murphy & Medin, 1985; Smith & Samuelson, 1997).

Hybrid views

Both fuzzy concept theories and the core or essentialist theories can account for some, but not all, of categorization, and both have serious flaws as mentioned above. In an attempt to account for more aspects of categorization and to overcome the flaws in each theory, researchers have created an amalgamation of the two extremes which is often referred to as the hybrid view of categorization (Medin & Wattenmaker, 1987; Keil, 1987). The hybrid theory was proposed to overcome the difficulties with supposing that there must be a list of necessary and sufficient features that every instance of a category possesses. The hybrid view still maintains a core but that core is no longer an entire set of necessary and sufficient features; it can be a single important trait, such as genetic codes in natural kinds. This idea is similar to the essentialist view but the hybrid view goes one step farther (Malt, 1990; Keil, 1989; Murphy & Medin, 1985).

In addition to the core, there are also some features possessed by members of the category which are probabilistically associated with the category and are not present in every member. The key aspects of both classical or essentialist and family resemblance models are preserved and combined. For example, under the hybrid theory, the core of the concept "chair" could, in theory, consist of the notion "for sitting on", whereas features like "has four legs" and "has a back" apply to most chairs but not all and aid in the categorization of potential chairs. As another example, take counter chairs at a diner which sometimes have but a single "leg", or beanbag chairs which have no legs or back. These "chairs" do not appear to share many properties in common with the majority of chairs, but they still have the core function of being "for sitting on" and we would still be able to identify an object as a chair.

On the surface it sounds as though the core feature and associated probabilistic features proposed by the hybrid theory addressed many of the problems associated with previous theories. However, the hybrid theory soon encounters the same difficulty other core theories have. For the hybrid theory as well, pointing out what the core feature is for any concept proves to be extremely difficult in practice. Therefore, the hybrid theory, with its core feature assumption, also falls prey to the same arguments leveled against the classical theory, namely how to determine which features define the category (Lakoff, 1987).

The hybrid theory also falls prey to the same criticisms of family resemblance and prototype views. As with these theories, there is no principled way to determine what counts as one of the probabilistically associated features of a hybrid category (Smith & Samuelson, 1997).

A New Framework

Naming versus recognition view

Research in categorization over the past years has swung back and forth between the poles of fuzzy concepts and core theories, and no major progress had been made (Malt & Johnson, 1992). Malt et al. (1999) have presented a theory of categorization which is an alternative to both defining features and fuzzy concept theories. Malt et al. maintain that "categorization" may really consist of two separate processes: recognition and naming.

Treating these two processes as identical may have lead to the apparently contradictory views categorization research has taken. When we begin to treat the processes as separate we may be able to resolve the cycling and tension between core and probabilistic theories.

Most importantly, the theory makes explicit the assumption that what has been called "categorization" is really two different processes at work on a single representation. The first of these is that people recognize objects as members of a kind; recognition is based on similarity and consists of encoding into a representational system. The model assumes that objects can be represented as points in a multi-dimensional feature space. The clusters objects form in the feature space correspond to the recognition categories. These clusters of objects have no fixed or discrete boundaries, and the objects form these clusters on the basis of similarity. Similarity can come from a number of sources: physical similarity, functional similarity, or overall similarity to name a few important ones. According to Malt et al., different laboratory tasks draw attention to different kinds of features and thereby appear to produce contradictory results. In other words, the dimensions influencing a category decision may be differentially weighted depending on the task and characteristics of the category.

The second process that can be thought of as "categorization" is naming. People attach names to objects when they discuss them. Names in the model are associated with objects with varying degrees of strength. Name activation is generally determined by an object's similarity to category exemplars. Hence there is an explicit exemplar assumption in the model in line with work by Nosofsky (1992, 1984) and Medin and Schaffer (1978). Exemplar views are a relatively recent way to look at categories. More specifically, they share with prototype views (as well as most others) the idea that categorization involves judgments of similarity to stored representations. However, instead of comparing a novel object to a single prototype, it is compared to specific previously encountered instances of the category. These "instances" are the category exemplars. Furthermore, the specification of a category is implicitly defined by its instances (Medin & Schaffer, 1978). Thus, with the exemplar assumption coupled with the assumption that different processes affect naming and recognition, the model predicts that similarity should account for names given to objects most of the time.

However, the framework leaves open the possibility that factors other than similarity may enter into naming. Because communication has its own demands, there are complexities that enter into the process of naming that do not affect recognition. One of these factors is the novel names manufacturers give to slight variations on familiar items. For example, Malt et al. (1999) discuss a small object with a flip up straw that, except for the small size, closely resembles other plastic objects with flip up straws that are usually labeled "sports bottles". However, this smaller object is called a "juice box" by the manufacturer because it has a functional relationship to the cardboard juice boxes widely available to go into children's lunch boxes. So the object would be closer in similarity space to other types of plastic squeeze bottles, but is instead called a "box" on its package label. This sort of process is known as historical chaining (Lakoff, 1987).

Convention is also a possible source of complexity in naming (Malt et al., 1999). If people grow up hearing a given object called by a certain name, that name will have a strong association with the object independent of the object's similarity relations with other objects. Thus, in Malt et al.'s theory, naming is strongly affected by processes such as chaining and historical convention. These processes are separate from recognition processes.

Current Study

The present study aims to examine how the Malt et al. theory may prove to be a viable alternative to both core and fuzzy concept views of categorization. In this study I attempt to separate consideration of the similarity driven recognition categories from the linguistic categories used to communicate about objects. I can then examine some of the issues which follow from the naming versus recognition view.

A subset of participants in the experiments that follow were asked to rate pairs of stimulus objects on the basis of various types of similarity: physical similarity, functional similarity, and overall similarity. The objects participants were asked to rate were all objects used in the preparation and serving of food. These data were used to generate a similarity space. This space corresponds to the recognition categories discussed in Malt et al. (1999). In my analysis, rather than generating multi-dimensional feature space, I used additive tree analysis to generate the graphic depiction of the similarity relations between the objects.

The rest of the participants were asked to provide names for the objects. This is my measure of linguistic category membership. With both recognition and linguistic data collected I can examine issues of the dimensions which best separate the linguistic categories, and then to begin to identify objects where similarity does not predict the name given to the object and offer some possible reasons why this may be so.

The theory outlined above and the studies described lead to a set of questions about features and category boundaries. What kinds of features separate linguistic category boundaries where the linguistic categories are determined by the names given to objects? We will look at three possibilities. First, does an object's function determine its category membership? Under core theories, an object's function is the best candidate for a core (Keil, 1989). If function is truly the defining feature for artifact categories, one would expect that function would delineate the boundaries between linguistic categories the best. If function does not do a good job at separating categories, further doubt will be cast on core views.

Second, do physical features determine linguistic category membership? Physical features include an object's shape, size, color, etc. If physical features separate categories best, this outcome would be contrary to core theories but in line with prototype views (Rosch & Mervis, 1975). Most prototype theories take physical features as the central features for category decisions, a point reflected in much of the literature on prototype views. However, with respect to the physical properties of artifacts, physical properties which define artifact categories have not been discovered (Rosch & Mervis, 1975).

Finally, does overall similarity account for linguistic category membership? The term overall similarity refers to some combination of features including but not exclusively, functional and physical features. This is more consistent with the Malt et al. view which assumes multiple dimensions underlying categorization, but it is also in line with some other exemplar theories (e.g. Nosofsky, 1992; Estes, 1986).

I predict that overall similarity will do the best at delineating linguistic category boundaries because, under the Malt et al. framework, no single dimension or combination of physical or functional feature types is expected to separate the linguistic categories fully. And by fully I mean with absolute completeness, where all and only objects called "bowl" would be confined clearly to a cluster, and so on. More features and kinds of information than is present even in overall similarity will be required for any accuracy approaching this level. This is suggested by the fact that no single feature type or combination of feature types have been discovered that does clearly delineate boundaries for any given category. Aside from these considerations, if we hold to the idea that a single feature will separate categories fully we are coming dangerously close to endorsing the classical theory again with its single or very small sets of defining features.

Also, I believe overall similarity will best separate category boundaries because, in addition to functional features I believe that physical features are also important to category decisions because many features we attend to in identifying and classifying objects are the perceptually based ones like shape or material of construction. Shape, for instance, may determine the difference between a bowl and a plate. And past research by Malt and Johnson (1992) that was discussed earlier, suggested that physical features were attended to and used by participants making category judgments.

Some criticisms of the use of overall similarity must be discussed, namely one could argue that overall similarity should separate categories the best, not because Malt et al.'s assumptions are correct, but simply because with the consideration of both physical and functional features more information is going into the category judgment. Therefore, because overall similarity entails more information we would expect better separation of categories regardless of the underlying theory. This is not necessarily the case, however. When I ask participants to make function and physical similarity judgments I do not stipulate how many functional or physical features the participant should attend to. I simply specify the type of features he or she should look at. The same argument holds for the overall similarity judgment. A participant could attend to the same number of functional features as he or she could overall features. Even if there are more bits of information going into the overall judgment, the function as core theory maintains that physical information should not add anything to the power of function to separate the categories. This is because the function as core theory states that function is the sole determinant of category membership. Therefore, overall should not, under core theory, be expected to do better than function.

A final issue regarding similarity types must be addressed. It could be argued that differences in shape may be a byproduct of an object's function, and, therefore, even if physical features are superficially important, function really is the primary consideration in categorizing artifacts. Differences in shape do seem to have consequences for how the object is used. Experience tells us it would be difficult at best to serve soup on a plate. Notice, however, that functional and physical features are dependent on each other in this case, but, at the same time, they are not always the same thing. In many instances physical features will be diagnostic of functional features, but, in many other cases those physical features will not be diagnostic, yet will still be important for categorization. This point was demonstrated in the Malt and Johnson (1992) research discussed earlier. In any event, this issue rather than supporting function-as-core views, illustrates the need for multiple types of similarity in categorization, which goes directly against core theories.

The analyses I use on the data will focus mainly on determining which similarity type is best at separating linguistic categories. I will also give a qualitative account of what kinds of specific features participants focused on within each type of similarity. It is hoped that this will provide a closer look at the recognition and linguistic categories by allowing some of the specific features which were important for each type of similarity to be examined. Should many different types of features be found in overall similarity, more support will be given to the Malt et al.'s contention that multiple dimensions, or in this case, types of features will be needed to account for naming.

Study 1: Naming

Naming data were collected from a group of participants. The names participants gave were used to determine the linguistic category membership of each object in the stimulus set.

Method

Stimuli. The stimuli consisted of 60 color photographs of various bowls, dishes, and plates. I chose those categories because they were closely related to each other in that they are all used in the preparation and serving of food, yet they are not the same. I attempted, with the stimulus set, to include as much variability in size, shape, material, etc. as possible for these categories. Variability was emphasized to reveal the boundaries of the categories. With many examples of closely related items and items which run the gamut of exemplars, I can get a clearer look at where the category boundaries lie. The objects used in the experiments were found in the homes of members of the Psychology Department, the homes of friends of the researcher, and in thrift stores.

The initial set of 79 stimulus objects was pilot tested by having 10 Lehigh University undergraduates provide names for all stimulus items considered. Those objects which received the labels "bowl", "dish", or "plate" more frequently than any other label were retained for the final stimulus set of 60 objects.

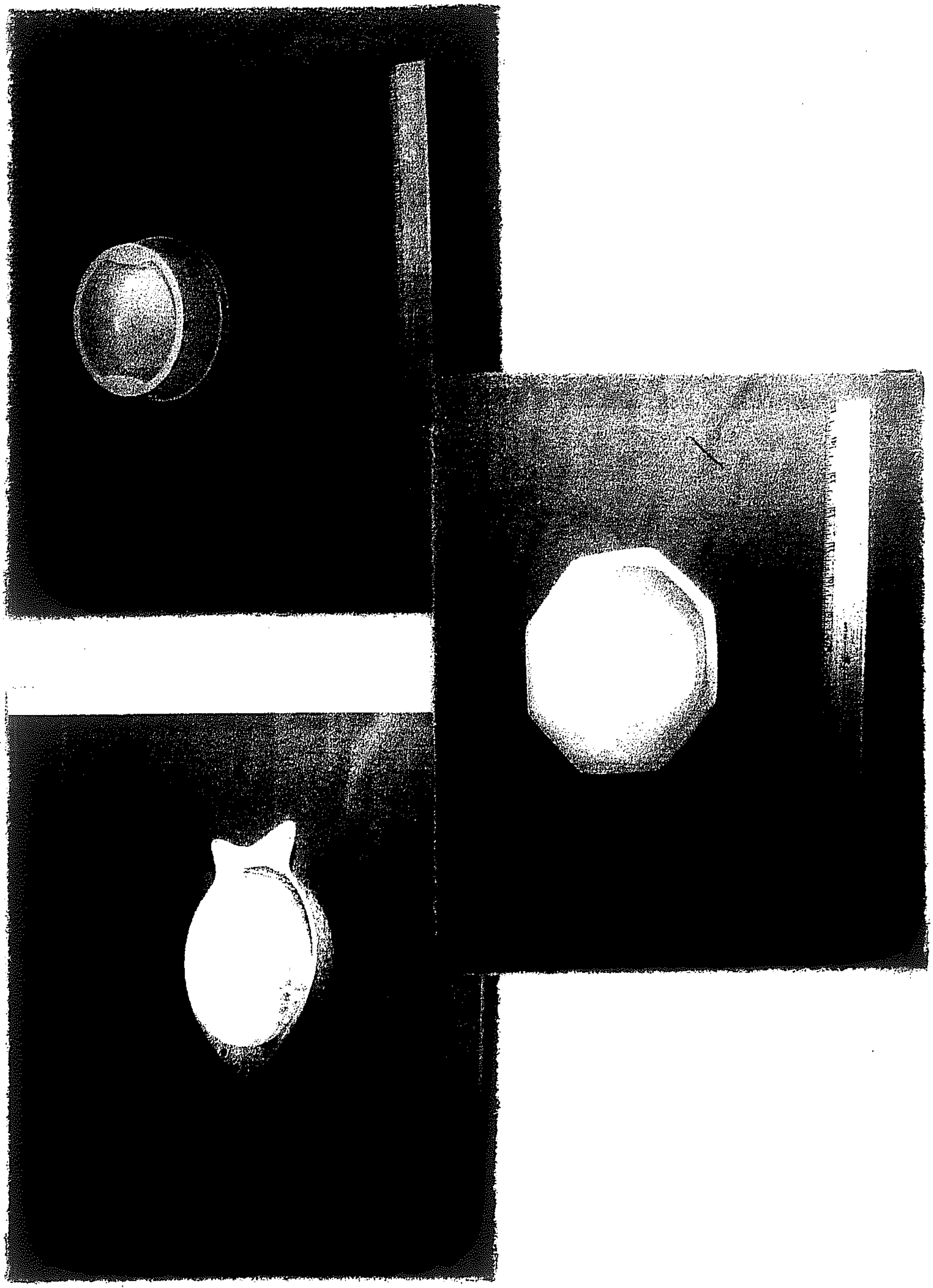
The stimulus set compiled after the pilot test included items such as large divided plates, small decorative bowls, crystal candy dishes, and so on. All final stimulus pictures were taken by the researcher against a grey drop cloth. The distance from the camera to the object was also kept constant to preserve size information. The same 15 inch wooden ruler appeared in each picture to maintain scale. Some of the pictures used are shown in Figures 1-3.

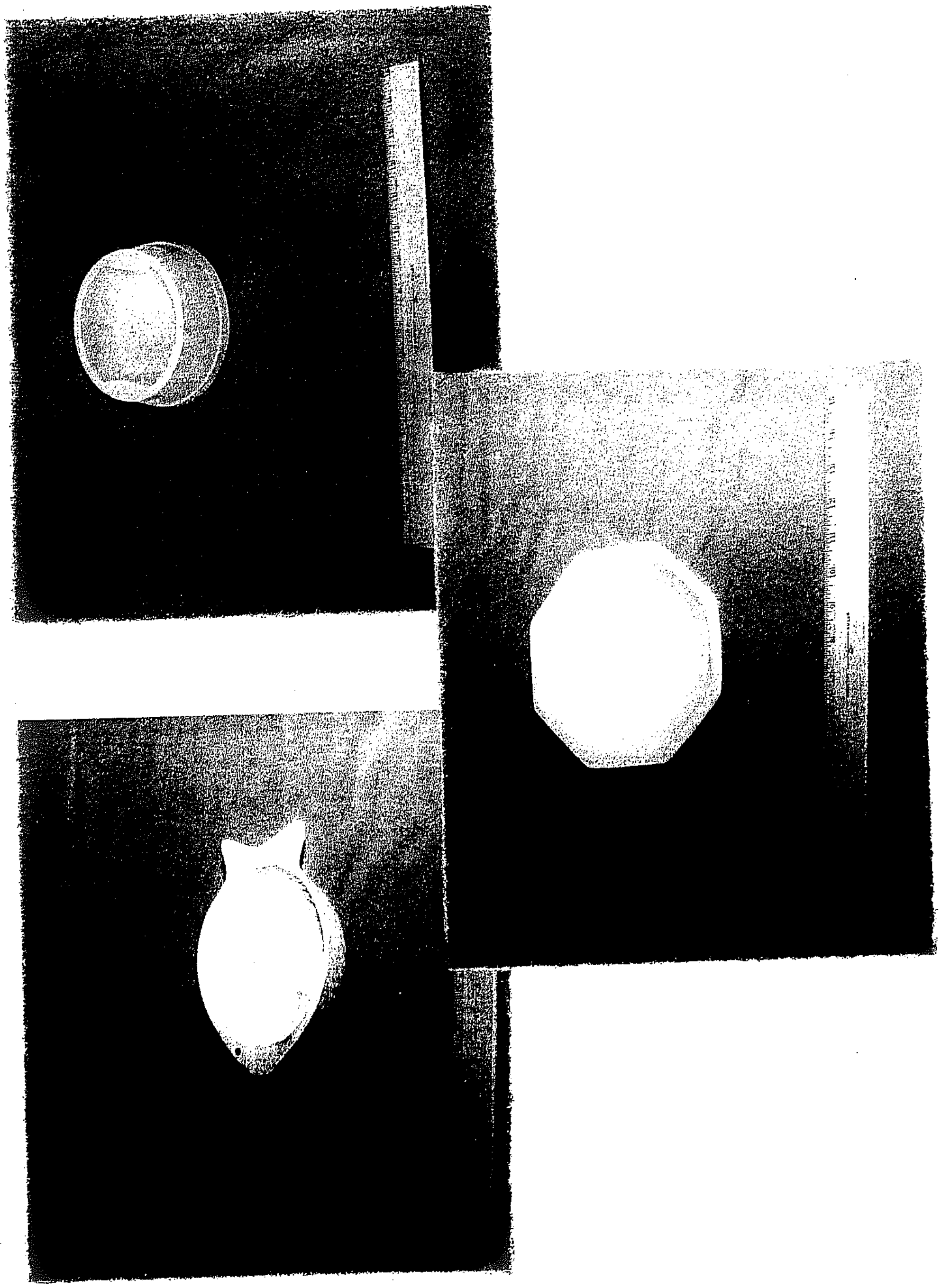
Participants. Twenty-four Lehigh University undergraduates enrolled in Introductory Psychology and Introductory Sociology provided names for the stimulus pictures. Students received course credit for their participation.

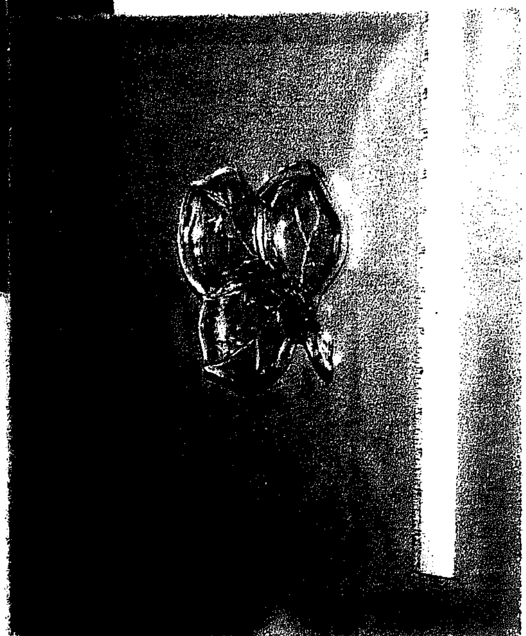
Procedure

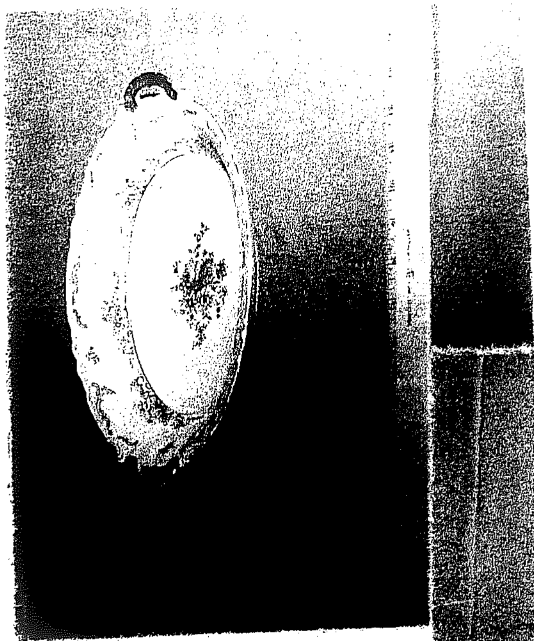
An initial set of 79 objects, which were mentioned above, was presented to 10 Lehigh University undergraduates, who received course credit for their participation. The participants were asked to provide names for the 79 objects. Those objects which were labeled a most frequent name other than "bowl", "dish", or "plate" were not retained for the first stimulus set. So objects which received names like "ashtray" more frequently than the labels "bowl", "dish", or "plate" were rejected. There were 60 objects in this set.

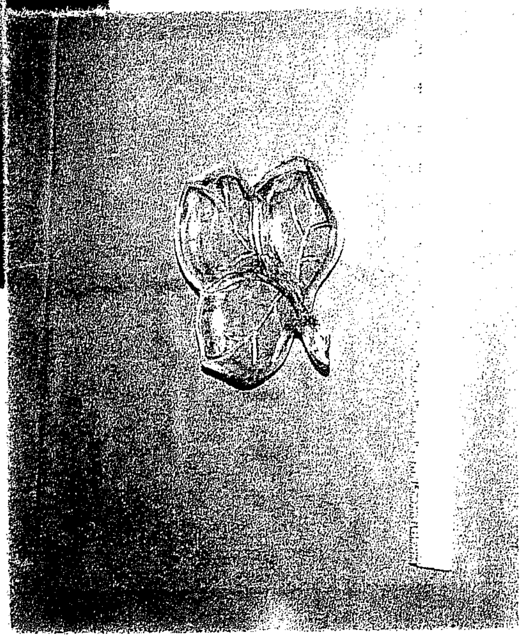
A first naming study was performed on this set of 60 objects. Eighteen Lehigh University undergraduates provided names for these objects for course

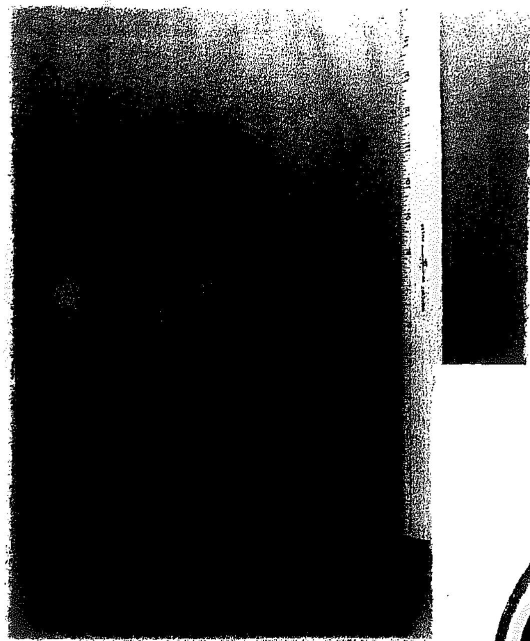


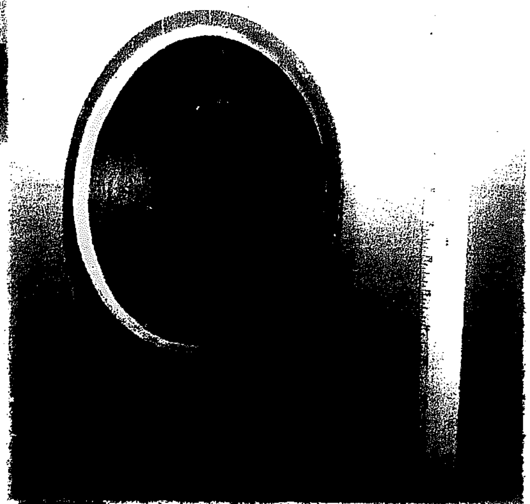


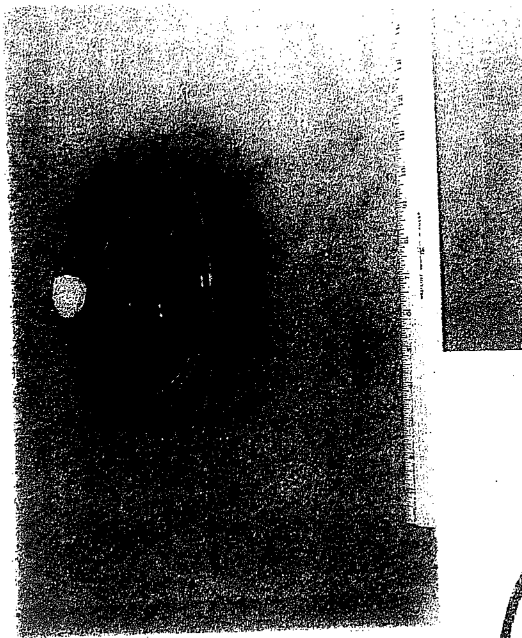


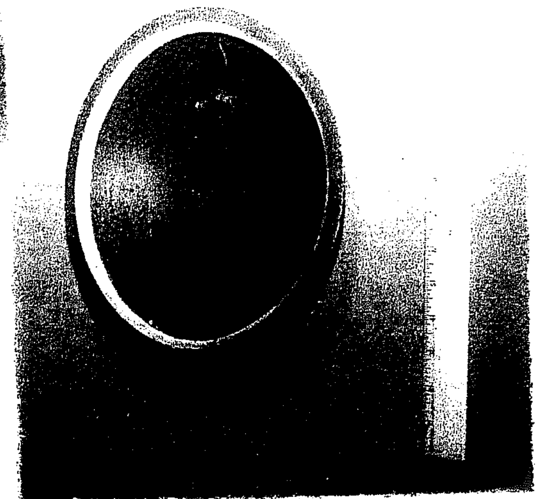


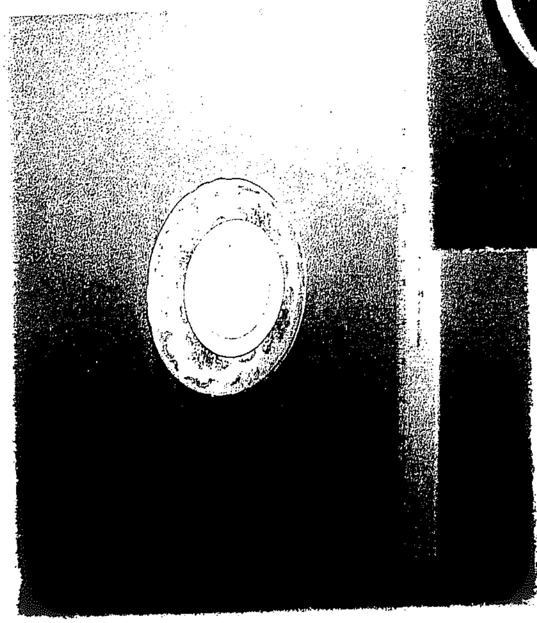












credit. Eight of these objects were labeled with dominant names other than bowl, dish, or plate and were removed from this stimulus set, leaving 52 objects.

Twenty-three new objects were added to the 52 remaining objects from the above stimulus set for a total of 75 objects. For all of these objects naming and typicality data were collected. Sixty of these objects which were called bowl, dish, or plate more frequently than any other label were kept for all analyses discussed in this paper.

After the participants entered the testing room, they were asked to glance through the stimulus set to familiarize themselves with the objects contained in the set. Each participant was then asked to provide a name for each object (a complete set of instructions is included in Appendix A). They were told that they were to provide the name for the object that they would use in ordinary conversation. The instructions stipulated that the names could be one word or more than one word, and that, though some of the objects may be difficult to name, they should try to provide a name for each one. After the participant was familiar with the naming instructions, the experimenter presented the stimulus pictures to the participant one at a time.

Results

To determine the preferred name, all the names given by the participants were tabulated along with the frequency of their occurrence. For each stimulus object, the most frequently given name is the linguistic category membership assumed for the remaining analyses. The names and their frequencies are presented in Table 1.

There were instances where there were two names with equal frequencies for a few objects. I define an object with more than one dominant name as an object with two names that were given with identical frequency or two names no more than two apart in frequency. The most frequent name given to the object is marked with a star in Table 1. For objects with two dominant names the name most frequently given is marked with a star and the second dominant name is marked with a karat. All objects received dominant names of either "bowl", "dish", or "plate".

Study 2: Similarity

Similarity data were collected for use in additive tree analysis.

Method

Stimuli. The stimuli for this study were the same final set of 60 items used in Study 1 except that the pictures were scanned into a computer file and were presented to the subjects on a computer screen. The ruler was cropped out of the photos for presentation to the subjects because of restrictions in the file size of the photos that could be presented with the program we used for stimulus presentation. The program used for stimulus presentation is called RSVP, version 3.0. The pictures of the stimulus objects were still in color for the computer presentation.

Participants. Participants were 42 Lehigh University undergraduates enrolled in introductory psychology or sociology who did not take part in Study 1, and 15 students at Brown University who had also not given any naming and typicality judgments. The Brown students volunteered to participate for pay. Each Lehigh

	Frequency of label										
No.	Bowl	Dish	Plate	Holder	Container	Saucer	Coaster	Tpprwre	Platter	Tray	Cup
1	22*	2	0	0	0	0	0	0	0	0	0
2	10*	7	1	3	1	1	0	0	0	0	0
3	3	2	18*	0	0	1	0	0	0	0	0
4	14*	7	2	0	0	0	1	0	0	0	0
5	17*	1	0	0	1	0	0	0	0	0	0
6	16*	7	0	0	0	0	0	0	0	0	0
7	20*	2	0	0	1	0	0	0	0	0	0
8	3	12*	5	7	2	0	0	0	0	4	0
9	14*	0	0	0	6	0	0	4	0	0	0
10	14*	9	1	0	0	0	0	0	0	0	0
11	2	10*	9^	1	0	2	0	0	0	0	0
12	0	12*	7	0	0	0	0	0	1	2	0
13	17*	7	0	0	0	0	0	0	0	0	0
14	20*	2	0	1	0	0	0	0	0	0	1
15	23*	0	0	0	0	0	0	1	0	0	0
16	9^	11*	0	0	1	0	0	0	0	0	1
17	23*	1	0	0	0	0	0	0	0	0	0

Note. A "*" indicates the most frequent name, while a "^" indicates a competing name.

Table 1

Frequency of names given by participants for Study 1.

	Frequency of label										
No.	Bowl	Dish	Plate	Holder	Container	Saucer	Coaster	Tpprwre	Platter	Tray	Cup
18	22*	0	0	0	0	0	0	0	0	0	2
19	0	9*	4	1	0	2	1	0	1	2	0
20	11*	10^	0	0	0	0	0	0	0	0	0
21	20*	4	0	0	0	0	0	0	0	0	0
22	14*	8	0	0	0	0	0	0	0	1	0
23	11*	5	0	1	4	0	0	0	0	0	0
24	11*	8	0	0	1	0	0	0	0	0	0
25	19*	2	3	0	0	0	0	0	0	0	0
26	3	6	9*	0	2	0	0	2	0	0	0
27	0	1	22*	0	0	1	0	0	0	0	0
28	0	15*	0	5	2	0	0	0	0	0	0
29	0	14*	1	0	0	0	0	0	1	2	0
30	10*	3	0	3	3	0	0	0	0	0	2
31	8^	10*	0	0	2	0	0	1	0	0	0
32	8^	10*	0	4	1	0	0	0	0	0	0
33	12*	7	0	0	0	0	0	1	0	0	0
34	14*	1	0	3	2	0	0	0	0	0	2

Table 1 continued

Frequency of label

No.	Bowl	Dish	Plate	Holder	Container	Saucer	Coaster	Tpprwre	Platter	Tray	Cup
35	0	15*	0	3	2	0	0	1	0	3	0
36	1	18*	0	0	1	0	0	0	1	0	0
37	1	16*	1	1	1	0	0	0	0	0	0
38	1	15*	0	0	0	0	0	0	0	4	0
39	4	9*	0	3	3	0	0	0	0	1	0
40	1	15*	0	0	1	0	0	1	0	0	0
41	3	12*	0	0	2	0	0	0	0	0	0
42	20*	2	0	0	0	0	0	0	0	0	0
43	9	3	12*	0	0	0	0	0	0	0	0
44	5	8*	4	0	0	0	0	0	0	5	2
45	0	12*	7	0	0	0	0	0	1	3	0
46	0	8*	5	0	0	1	0	0	4	5	0
47	1	9*	5	0	0	0	0	0	0	5	0
48	0	1	15*	0	0	7	1	0	0	0	0
49	0	6	16*	0	0	0	0	2	0	0	0
50	0	16*	1	1	1	0	0	0	0	5	0

Table 1 continued

						Frequency of label					
No.	Bowl	Dish	Plate	Holder	Container	Saucer	Coaster	Tpprwre	Platter	Tray	Cup
51	0	2	21*	0	0	1	0	0	0	0	0
52	9	1	13*	0	0	1	0	0	0	0	0
53	2	5	14*	0	0	0	0	0	1	1	0
54	0	1	22*	0	0	0	0	0	0	0	0
55	0	2	22*	0	0	0	0	0	0	0	0
56	0	11*	3	4	0	0	0	0	1	2	0
57	0	7	11*	0	0	1	0	0	1	3	0
58	0	0	24*	0	0	0	0	0	0	0	0
59	0	9*	9*	0	0	0	0	0	2	4	0
60	0	6	15*	0	0	0	0	0	1	1	0

Table 1 continued

student received course credit for participation.

Procedure

Participants were asked to rate the similarity of each object in the stimulus set to every other object in the stimulus set. The pairs to be rated were presented on a computer screen and the participant responded via the keyboard.

First, the participants were asked to perform a set of practice trials. They were instructed that two pictures would appear on the screen following a brief exposure to a cross in the center of the screen. Participants were asked to judge how similar the objects presented were to one another. The scale for the judgments ranged from one to eight. One was for very dissimilar objects and eight was for highly similar objects. Participants were to respond on the keyboard, an eight point numeric scale was clearly imposed over the middle row of the keyboard. The "a" key was designated "1", and the ";" key was designated "8". The numeric scale was clearly imposed over the keys. At this point the participants were given no specific directions on how to judge the objects' similarity but they were informed that more specific directions would be given to them later. The pictures for the practice set were not any of the pictures used in the actual experimental trials.

Following the practice trials came a brief familiarization phase. Familiarization trials were important because we wanted the participants to be familiar with the range and variability of the objects they were rating. This way they were less likely to judge one pair of objects as highly similar only to find that another object was even more similar to one of the objects in the pairing than the one in the original pairing was. For this part of the experiment, participants were asked to simply look at the set of 60 pictures to see what the stimulus set looked like. Again, the presentation was on the computer. Each time the participant pressed the space bar, six of the pictures were presented on the screen. Participants could look at the six items as long as they wanted before pressing the space bar again to see the next six objects. The items were grouped in numerical order so the first screen contained objects one through six, the second items seven through 12, and so on.

Once the participant completed the familiarization phase, the experimental trials began. For the experimental trials, participants were informed that they would be indicating their judgments with the eight response keys just like they did in the practice trials, except that for these trials they would be asked to judge the similarity of the objects in a specific way. They were also told that these trials would progress in the same manner as the practice trials.

The specific manner in which they were to make their judgments was based on a particular instruction given to them by the experimenter. There were three different versions of rating the similarity of the objects: physical, functional, and overall similarity. Each participant did only one type of rating due to the length of the experiment.

For the physical instruction, 19 participants were asked to focus on the physical properties of each object, such as what it looked like, what it was made of, etc. They were told to rate objects which were very similar to each other in their physical appearance (e.g., size, shape, and color) as highly similar and those which differ in their physical appearance as less similar (for complete instructions see Appendix B). For the Function instruction, a different 19 participants were asked to rate as highly similar those objects which were used in a similar way to prepare, serve, or hold food and other items and to judge as less similar those objects which were used in dissimilar ways (for complete instructions see Appendix B). Finally, another 19 participants rated the objects on Overall similarity. These participants were asked to rate as highly similar those objects which were similar overall, that is similar on features such as what the objects looked like, how they are used, or any other aspects of the objects that seemed important or natural to the participant (for complete instructions see Appendix C). The experimenter stressed to the participant that they were to rate objects as similar only if they felt they were similar in the relevant way, not because the objects were used together. For example, participants were informed that I did not want them to rate a salad plate and a bowl for preparing salad as similar because the objects were both used for salad and the objects, themselves, would not necessarily be similar in function, physical, or overall features.

Each of the possible 1 770 pairwise comparisons of the 60 stimuli were presented to the participants. Due to the amount of time it takes to make such a large number of comparisons, the experiment was done in three separate sessions for each participant. Each session lasted approximately one hour. The only difference between the first session and the second and third was that for the latter two there were no practice trials or familiarization phase. The pairs presented in each session were randomly selected by the computer program used to control presentation. At the beginning of each of the last two sessions, participants were asked to read the instructions once more to refresh their memories.

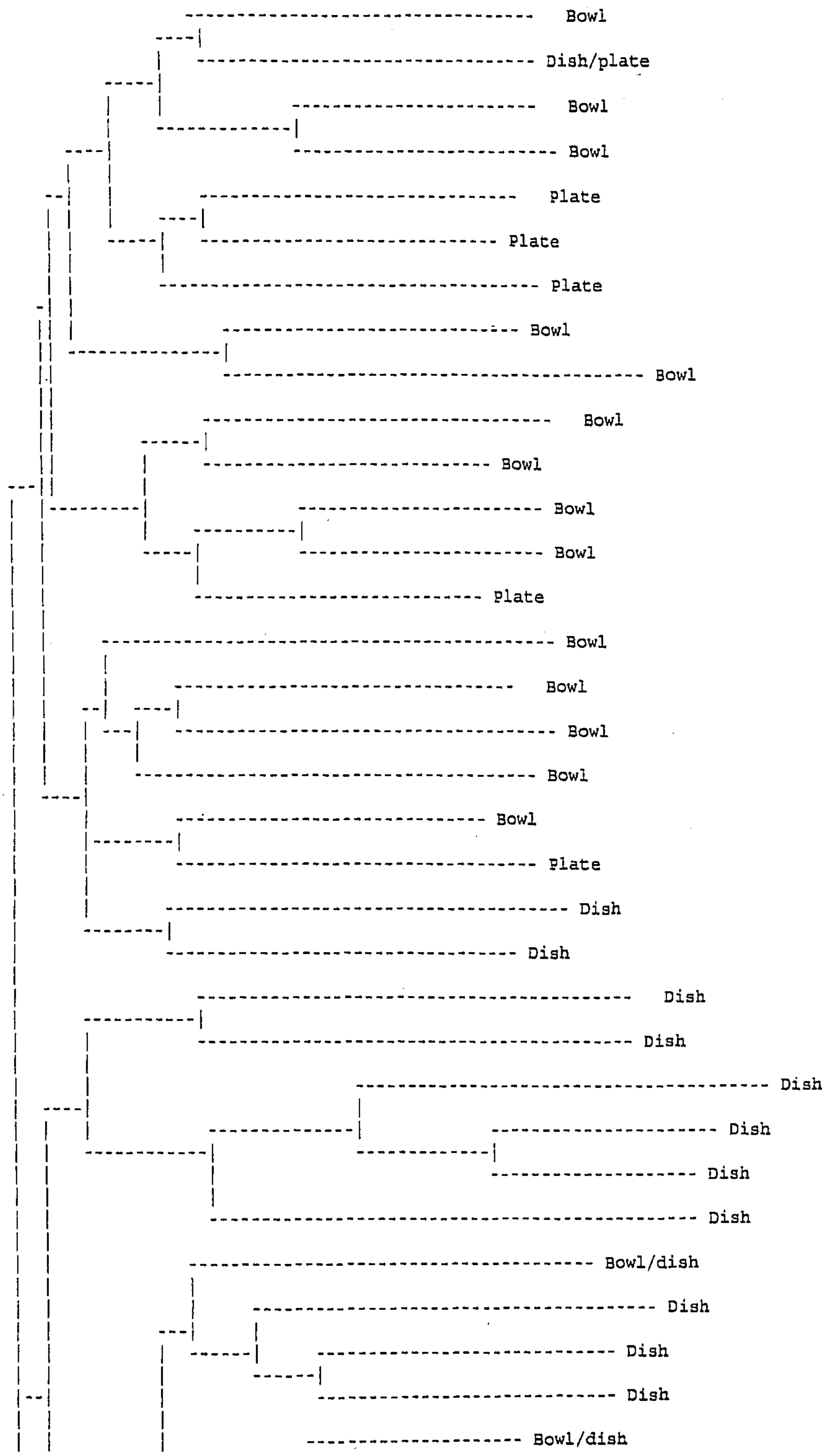
Results and discussion. Malt et al. (1999) used multi-dimensional scaling solutions to generate the similarity space. I would have liked to have used the same analyses because it would make the results comparable to the results Malt et al. generated; however, similarity space as was created by multi-dimensional scaling (MDS) solution for the present data set proved to be too complicated to interpret. Instead, additive similarity trees were used to analyze the data; one tree was generated for each type of similarity. The same similarity data used in MDS procedures can be used in and represented as an additive tree (Sattath & Tversky, 1977). MDS is a spatial model where each object is represented as a point in coordinate space. Additive trees, in contrast, are network models where each object is represented as a node in a connected graph (Sattath & Tversky, 1977). The relations of objects in the graph reflect their distance from each other. Dissimilarity in additive trees is represented by the horizontal length of the path joining the nodes. Often, with trees the stress of the fit is lower than with MDS (Sattath & Tversky, 1977). This property is highly desirable because stress is a measure of goodness of fit. Therefore, in many cases trees fit the data better than MDS. Trees are also easier to interpret than MDS when the scaling solution does not present a one or two dimensional solution.

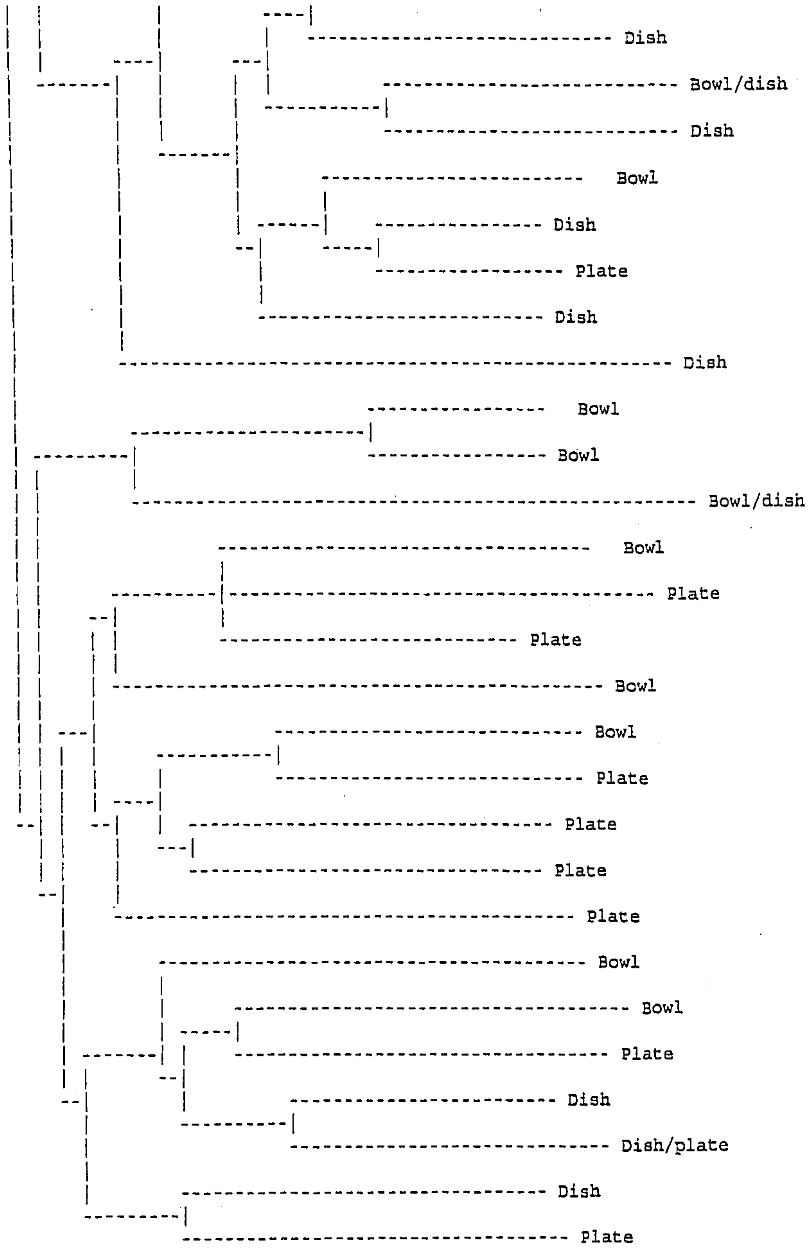
Additive similarity trees can be used to determine if the particular kinds of similarity separate category boundaries or group like named objects together. If a given similarity type is good at delineating linguistic category boundaries then those objects will be closer together in the tree. In similarity trees similarity is measured by looking at the path connecting a pair of objects. As one traces the path through the graph, the distance traveled horizontally is measured, vertical moves are not counted in the distance. These horizontal portions of the path connecting the objects will be shorter if objects are similar and longer if they are less similar. Items at the top of the tree are most dissimilar from items at the bottom of the tree. Therefore, if a particular similarity type is good at delineating category boundaries, we could expect to see something like a stratification, with one category more or less confined to the first part of the tree, a second below that, and so on. Those horizontal portions of the arcs can also be considered a measure of what features the objects connected by the path have in common (Sattath & Tversky, 1977).

An additive tree analysis was run for each type of similarity. The results of these analyses are presented in Figures 5, 6, and 7. The names listed at the terminal nodes of the tree are the names of each object as determined in the naming study. The stress value for a given tree should be below 0.1 in order to say a good fit was achieved. The stress for functional, physical, and overall similarity trees was 0.09, 0.07, and 0.08, respectively, indicating a good fit.

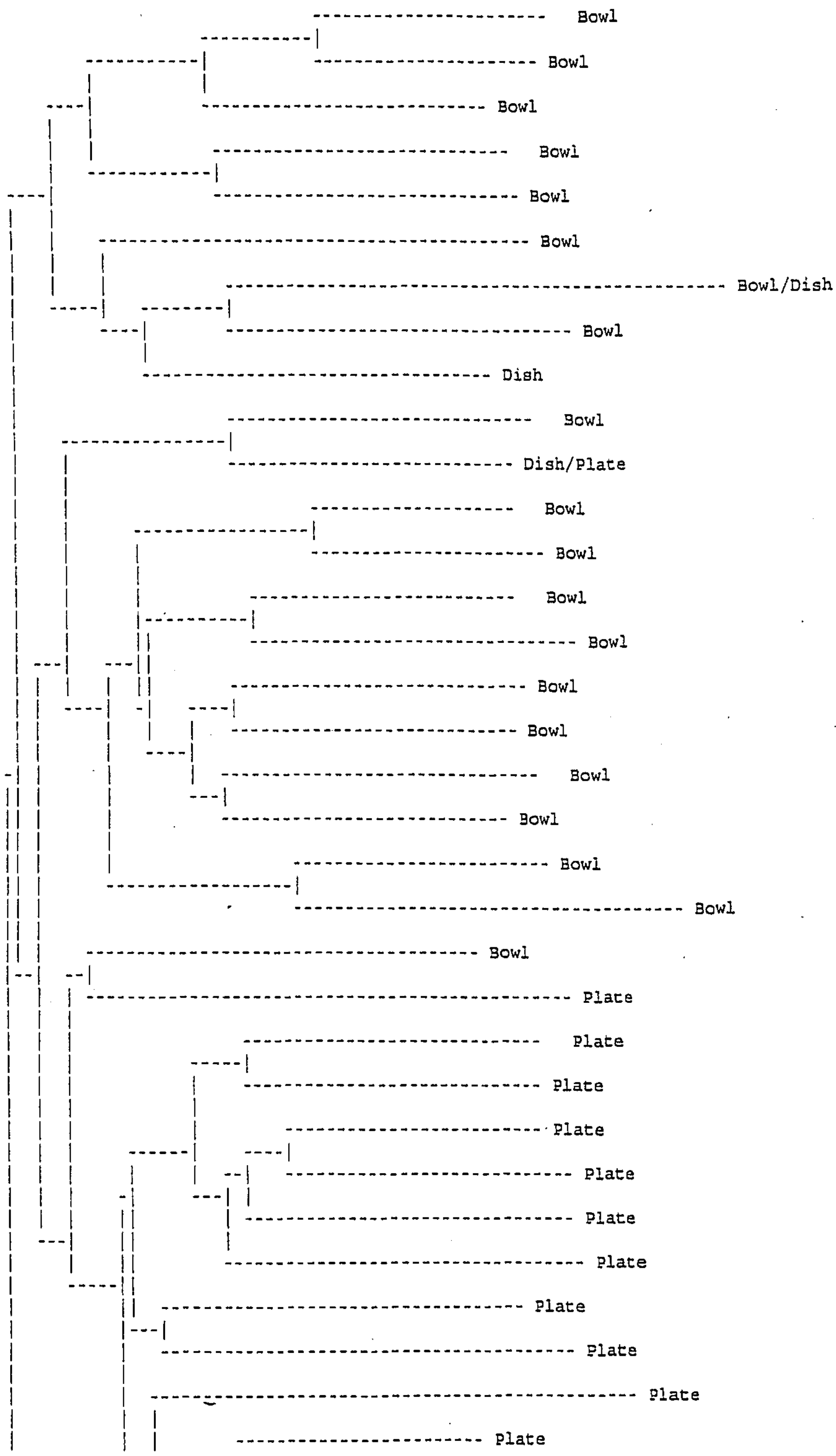
The question to be considered when looking at the trees is which type of similarity forms the most homogeneous clusters; that is, which similarity type places objects with the same labels closest together. It is also important to bear in mind that since the tree analysis represents the similarity relations between objects, this corresponds to the recognition component of the naming versus recognition model. Once again, the names which appear on the trees were generated by the naming study. I begin my analysis of the trees with a more

Physical similarity




Dish
Bowl/dish
Dish
Bowl
Dish
Plate
Dish
Dish
Bowl
Bowl
Bowl/dish
Bowl
Plate
Plate
Bowl
Bowl
Plate
Plate
Plate
Plate
Bowl
Bowl
Plate
Dish
Dish/plate
Dish
Plate

Overall similarity



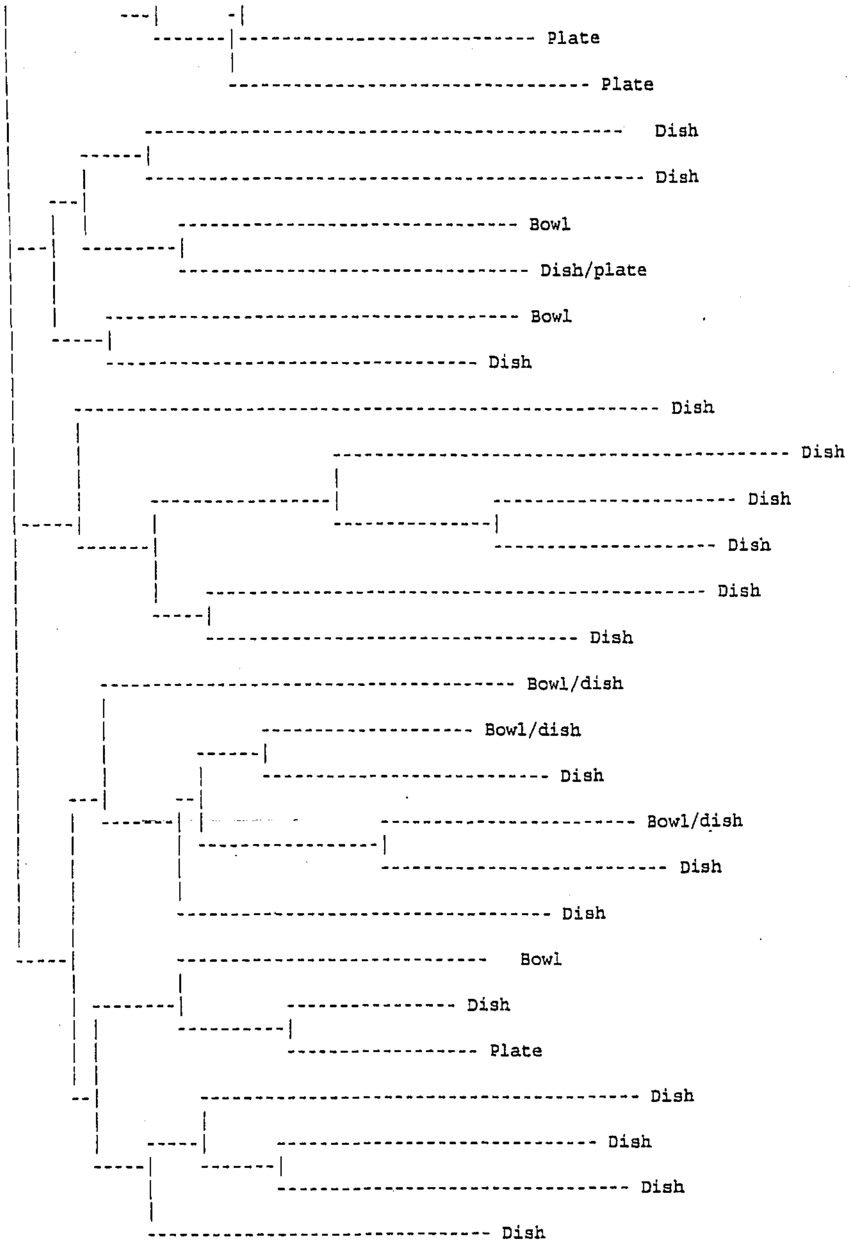
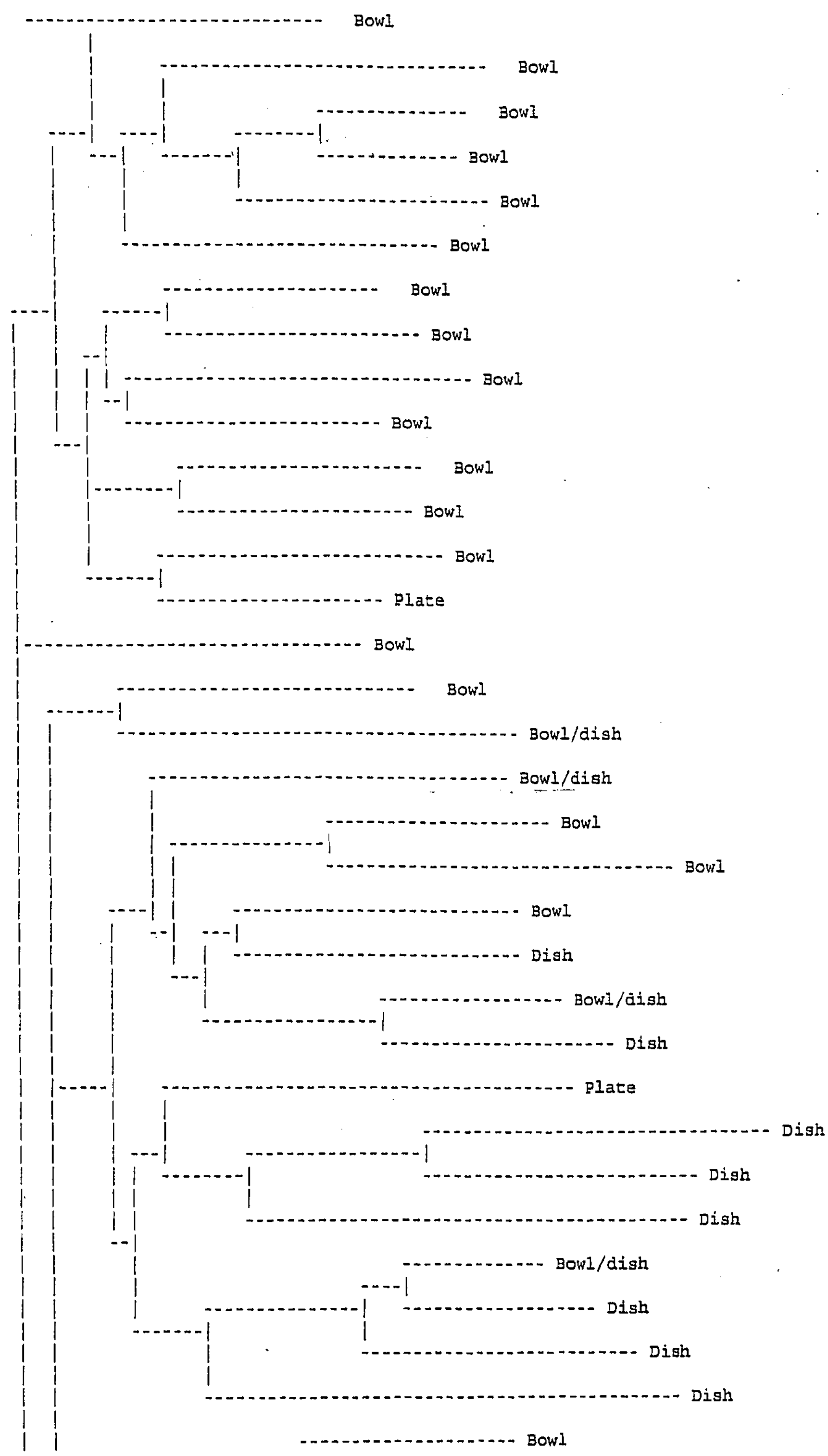


Plate
Plate
Dish
Dish
Bowl
Dish/plate
Bowl
Dish
Dish
Dish
Dish
Dish
Dish
Dish
Bowl/dish
Bowl/dish
Dish
Bowl/dish
Dish
Dish
Bowl
Dish
Plate
Dish
Dish
Dish
Dish

Functional similarity



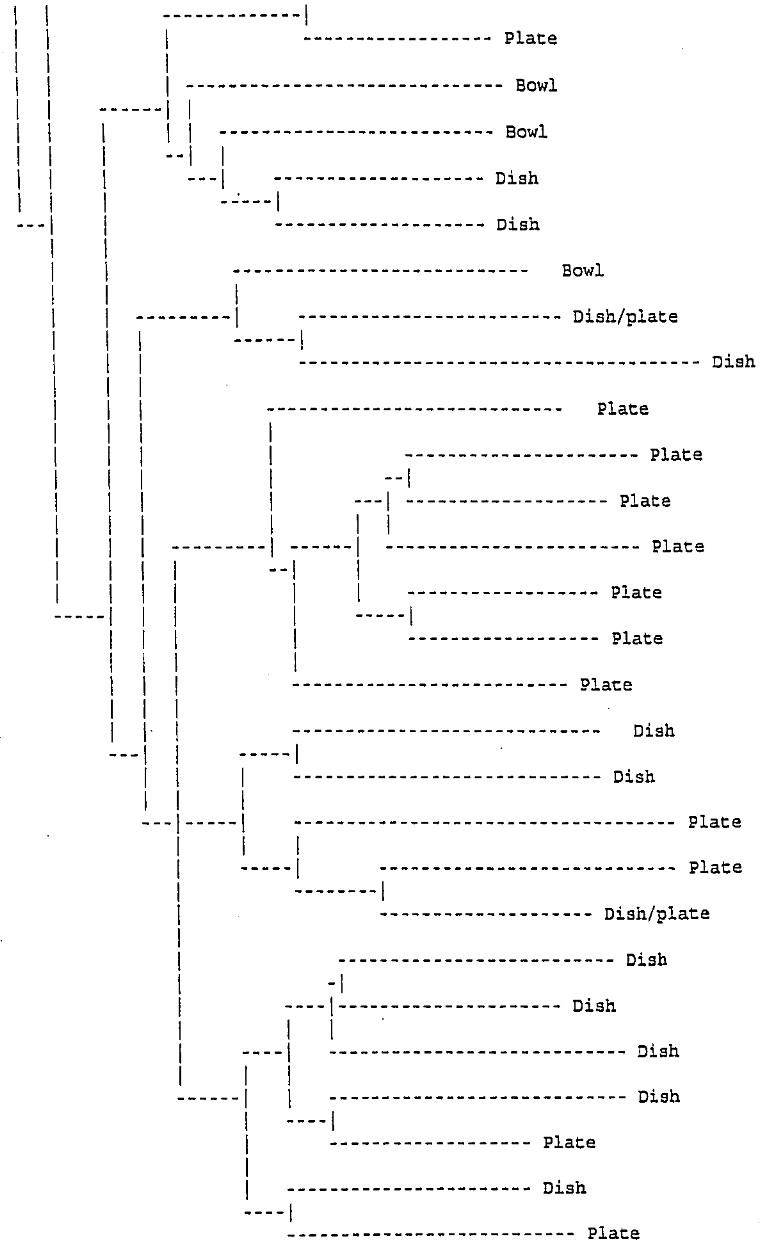


Plate
Bowl
Bowl
Dish
Dish
Bowl
Dish/plate
Dish
Plate
Plate
Plate
Plate
Plate
Plate
Plate
Dish
Dish
Plate
Plate
Dish/plate
Dish
Dish
Dish
Dish
Plate
Dish
Plate

explicitly descriptive account of the placement of the objects for each type of similarity. I will then move on to provide a statistically descriptive account of the differences between the function, physical, and overall trees. Finally, for each similarity type, I look at the clusters formed in the trees and give a qualitative account of what sorts of features they share.

Looking at the trees, one can see that overall similarity appears the best at separating the linguistic categories. Here the bowls are more or less clustered together at the top of the tree, followed by plates then dishes. There are only a few instances of plates being more similar (or placed within the bowls "section") to bowls than to other plates, and so on. In contrast, with the physical and functional similarities there were more heterogeneous groupings. For example, at the top of both the function and physical trees many bowls cluster together, but for both kinds of similarity there were many instances where there were plates or dishes interspersed within the bowls. This means that these plates and dishes were seen as sharing more features with bowls than with other plates or dishes for the particular similarity type in question. As with bowls, plates and dishes are also clustered best by overall similarity.

Visual inspection of the trees suggests that overall similarity judgments provide the best accounting for names as evidenced by the superior clustering in the overall similarity tree. However, a statistical test was devised to present a more rigorous analysis of the trees. Looking back at the trees in figures 5, 6, and 7 one can examine the apparent clusters of bowls, for example, and see that, at the top portion of the tree there are mostly bowls. Not every single object in the "bowl cluster" is a bowl, however. There are some dishes and some plates which appear within this larger cluster. This suggests that the homogeneity of labels within a cluster can be used as a measure of clustering.

To get a numeric description of the clustering of the various types of similarity for statistical analysis, I defined a bounded cluster of, bowls, for instance, and counted the number of intrusions by objects which were not called "bowl" into that cluster. I compared this to the total number of objects in the stimulus set not called "bowl". For reasons of clarity those objects which had two dominant names, like the "dish/bowls" in the trees were excluded from the analysis. This is because it is not clear what their category truly is. This left me with 54 objects to use in the analysis.

The next question is how to determine what a cluster of objects should be. One way to proceed would seem to be to find the first instance of bowl and count down through the tree until some specified number of bowls is included, then to count the number of intrusions of non-bowls in the cluster that method formed. This simple algorithm is problematic because often the first instance of a category occurred as an outlier very early in the tree. An alternative algorithm used which solved the problem was to find the median bowl of the total number of bowls in the tree and count symmetrically outward from the median until a specified percentage of bowls were included in the cluster.

I have mentioned that I counted for the analysis a specified number of bowls. There were several different numbers of objects considered. This is because the number of included objects is asking the question of how perfect or inclusive do I expect the clusters to be. Do I expect 100% of the objects to be clustered with few or no intrusions? This, in light of Malt et al.'s (1999) work suggesting names will never be perfectly specified by similarity, does not seem reasonable. Figures like 50% seem too low to make sense. Three percentages corresponding to three levels of inclusiveness were selected. I present analyses for 75%, 80%, and 90% inclusiveness, though I focus mainly on the 75% measure as it seems to me to be the best balance between inclusiveness and allowing for those instances where similarity does not fully predict the name given to the object.

I will be using Chi-square analyses to explore the data. I employed the Chi-square to compare the homogeneity of the clusters across the three similarity types at each of the three levels of inclusiveness already discussed. Specifically, if a type of similarity were good at clustering together like-named objects, then we should find few intrusions of objects with different names within the cluster. In contrast, if a given similarity type does not cluster objects well, we would expect many intrusions. The Chi-square test will reveal if the types differ from each other in the degree of clustering, So, for each level of inclusiveness I counted the number of intrusions of objects without the target label as the measure employed in the Chi-square. The expected values for the Chi-square are generated from the table of actual values and the number of possible intrusions (the number of objects without the category label in question). Specifically, the expected values are generated by multiplying the total number of intrusions found for all three similarity types divided by the total of the row total (the grand total), and the number of possible intrusions also divided by the grand total, and multiplying this number by the grand total. Chi-squares are done separately for each category, "bowl", "dish", and "plate" since the categories themselves are non-independent. Non-independence is a violation of the assumptions behind the Chi-square analysis if we were to compare the categories directly. Since it is the case that the categories, themselves, are non-independent the analyses are done for completeness, but treated separately.

Bowl. I discuss the largest category in the study, "bowl" first. Table 2 gives the number of intrusions in the bowl cluster and the number of non-bowls outside the bowl cluster for each type of similarity at the 90% inclusiveness level, Table 3 is for 80% inclusiveness, and Table 4 is for 75% inclusiveness. The clusters were obtained by counting symmetrically out from the median bowl. The values in parentheses are the expected values from the Chi-square analysis.

The chi-square for bowls was not significant for the 90% inclusiveness level, $\underline{X}^2(2) = 4.36$, $p > 0.05$, indicating no difference in how well the three similarity types clustered the objects. At the 90% inclusiveness level the test was requiring that 20 out of the 22 bowls in the set be included in a cluster for maximum possible clustering. As one can see in Table 2 there were a large number of intrusions into the bowl cluster across types of similarity. At an inclusiveness level of 80%, the chi-square is also not significant, $\underline{X}^2(2) = 4.3$, $p > 0.05$. Eighteen out of 22 bowls must be included at this level. Looking back at the trees, these levels of inclusiveness do not seem to reflect what to the eye seems obvious, that overall similarity definitely performs better than the other two types.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	16 (20.32)	17 (20.32)	28 (20.32)	61
non-bowl outside	16 (11.68)	15 (11.68)	4 (11.68)	35
Total	32	32	32	96

Note. The number of intrusions of non-bowls in the tree analysis clusters and the number of non-bowls outside the cluster are listed for each type of similarity.

Table 2

Chi-square table for bowls for 90% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	16 (17.34)	12 (17.34)	24 (17.34)	52
non-bowl outside	16 (14.66)	20 (14.66)	8 (17.34)	44
Total	32	32	32	96

Note. The number of intrusions of non-bowls in the tree analysis clusters and the number of non-bowls outside the cluster are listed for each type of similarity.

Table 3

Chi-square table for bowls for 80% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	1 (12.33)	12 (12.33)	24 (12.33)	37
non-bowl outside	31 (19.67)	20 (19.67)	8 (19.67)	59
Total	32	32	32	96

Note. The number of intrusions of non-bowls in the tree analysis clusters and the number of non-bowls outside the cluster are listed for each type of similarity.

Table 4

Chi-square table for bowls for 75% inclusiveness.

Despite the apparent lack of difference in clustering by the three similarity types at the 90% and 80% inclusiveness, at 75% inclusiveness, there is a dramatic difference. This is because the other two levels were always including an outlier which pushed the number of intrusions higher. For instance, with bowl cluster at the 80% inclusiveness level,. 17 out of the 18 bowls I was attempting to include in a cluster were clustered fairly well for overall similarity. However, to capture the final bowl an large number of additional nodes had to be included. At 75% inclusiveness, there is but a single intrusion in the cluster of 17 bowls for overall, 12 intrusions for function, and 24 for physical. The chi-square on these differences is significant, $\underline{X}^2(2) = 21.47$, $p < 0.001$. Looking at Table 2 it appears that overall similarity is indeed the one with the best clustering power. However, contrary to prediction function does not fare the worst in its ability to separate categories. The greatest number of intrusions occurs in physical similarity. I compared physical and functional similarity for bowls in another chi-square to see if this difference is significant, and it was, $\underline{X}^2(1) = 11.06$, $p < 0.001$. Thus, function truly does outperform physical similarity at category separation, at least for bowls.

Though the superior performance of functional similarity over physical similarity is contrary to prediction, it still does not salvage the function as core position. The function-as-core position dictates that physical features can play no role in categorization. This position is untenable given that overall similarity outperformed functional as well as physical similarity, and we have already noted, some purely physical features were present in the overall similarity tree. Thus physical features do indeed play a role in category decisions. These data suggest that multiple dimensions underlie categorization, which is consistent with Malt et al.'s (1999) general view.

Plates. Chi-square tests were also performed on the plates. The number of intrusions for plates at 90% inclusiveness is listed in Table 5, Table 6 is for 80% inclusiveness, and Table 7 is for 75%inclusiveness. There was near perfect clustering of plates for overall similarity (perfect for 90% and below). Looking at the table it is obvious that neither function or physical similarity do nearly as well. The chi-square for plates at the 90% inclusiveness level was significant, $\underline{X}^2(2) = 31.6$, $p < 0.001$. Again, I ran a chi-square on physical and functional similarity alone. This chi-square, too, was significant, $\underline{X}^2(1) = 11.6$, $p < 0.05$. And again, function does outperform physical for the plates, this holds true as well for 80% and 75% inclusiveness levels.

The 80% and 75% inclusiveness levels for plates are identical and so are reported as one. The chi-square is significant $\underline{X}^2(2) = 40.37$, $p < 0.001$. The chi-square between physical and function was also significant $\underline{X}^2(1) = 24..7$, $p < 0.001$.

Dishes. Finally, the same analyses were run on the dishes category. The intrusion data for dishes at 90% inclusiveness are presented in Table 8, Table 9 is for 80% inclusiveness, and Table 10 is for 75%. At 90% inclusiveness, the chi-square is significant, $\underline{X}^2(2) = 7.7$, $p < 0.05$. A glance at Table 4 will reveal superiority for overall and near identical performance for physical and functional. The tables and data for 80% and 75% inclusiveness are also the same. The chi-square is significant, $\underline{X}^2(2) = 9.47$, $p <0.01$. Here it seems that physical might

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	0 (20)	26 (20)	34 (20)	60
non-plates outside	40 (20)	14 (20)	6 (20)	60
Total	40	40	40	120

Note. Number of intrusions of non-plates in the tree analysis clusters and the number of non-plates outside the plate cluster for each type of similarity. The expected values are in parentheses.

Table 5

Chi-square table for plates for 90% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	0 (15.67)	12 (15.67)	35 (15.67)	47
non-plates outside	40 (24.33)	28 (24.33)	5 (24.33)	73
Total	40	40	40	120

Note. Number of intrusions of non-plates in the tree analysis clusters and the number of non-plates outside the plate cluster for each type of similarity. The expected values are in parentheses.

Table 6

Chi-square table for plates for 80% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	0 (15.67)	12 (15.67)	35 (15.67)	47
non-plates outside	40 (24.33)	28 (24.33)	5 (24.33)	73
Total	40	40	40	120

Note. Number of intrusions of non-plates in the tree analysis clusters and the number of non-plates outside the plate cluster for each type of similarity. The expected values are in parentheses.

Table 7

Chi-square table for plates for 75% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	4 (11.52)	15 (11.52)	16 (11.52)	35
non-dishes outside	32 (24.48)	21 (24.48)	20 (24.48)	73
Total	36	36	36	108

Note. Number of intrusions of non-dishes in the tree analysis clusters and the number of non-dishes outside the dish cluster for each type of similarity. The expected values are in parentheses.

Table 8

Chi-square table for dishes for 90% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	7 (7.67)	14 (7.67)	2 (7.67)	23
non-dishes outside	29 (28.33)	22 (28.33)	34 (28.33)	85
Total	36	36	36	108

Note. Number of intrusions of non-dishes in the tree analysis clusters and the number of non-dishes outside the dish cluster for each type of similarity. The expected values are in parentheses.

Table 9

Chi-square table for dishes for 80% inclusiveness.

	Type of Similarity			
Intrusions	Overall	Function	Physical	Total
intrusions in cluster	7 (7.67)	14 (7.67)	2 (7.67)	23
non-dishes outside	29 (28.33)	22 (28.33)	34 (28.33)	85
Total	36	36	36	108

Note. Number of intrusions of non-dishes in the tree analysis clusters and the number of non-dishes outside the dish cluster for each type of similarity. The expected values are in parentheses.

Table 10

Chi-square table for dishes for 75% inclusiveness.

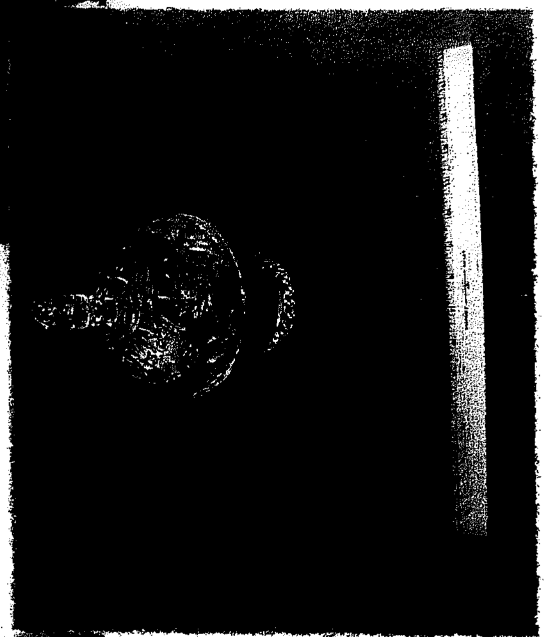
actually come close to performing as well as overall, the chi-square however, was significant, $\underline{X}^2(1) = 4.25$, $p < 0.05$. Therefore, for dishes as well, overall does the best. However, with dishes physical actually outperforms functional, $\underline{X}^2(1) = 9.41$, $p < 0.01$. Some potential reasons for this reversal will be presented in the general discussion.

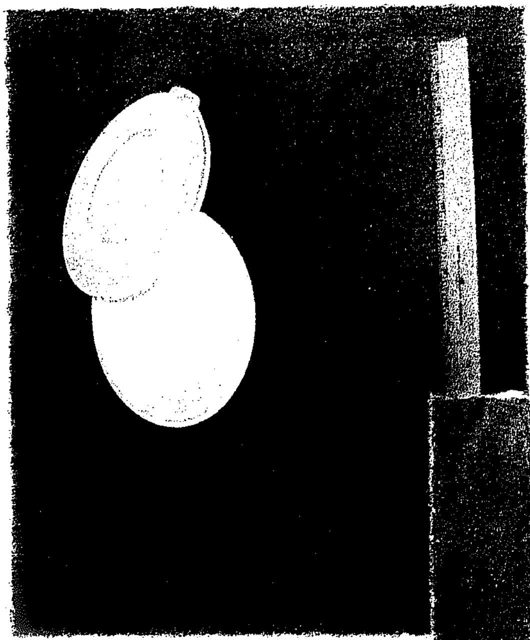
In summary, the chi-square tests, for each category, support the hypothesis that overall similarity will perform best. The hypothesis that physical similarity will outperform functional similarity was not supported for bowls and plates but was for dishes. In any event, the fact that overall similarity did perform best sheds further doubt on core theories.

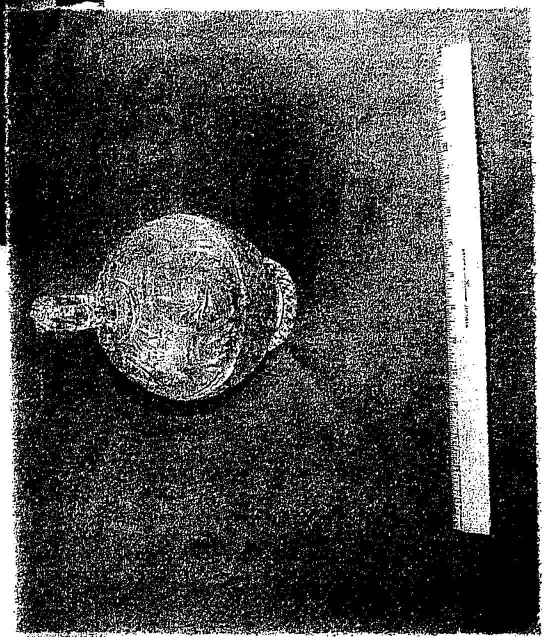
It would be instructive to see what sorts of features objects closer to each other in the tree analysis share. This is done for several reasons. First, it can act as a sort of check on my similarity type manipulation. If, upon examination, the tree for physical similarity for instance, contained many functional features, I would have reason to suspect my participants were either not following directions or that physical features simply are not important for categorization. Also, especially with overall similarity, it is important to see if participants were indeed considering combinations of features and not just focusing on a single type of feature. Second, since a major part of this research is to examine the contention that functional features form a core for artifact concepts, I need to look at the features in overall similarity. If they turn out to be entirely or mostly functional ones, my hypotheses about core theories will be demonstrated inadequate. Finally, as was discussed in the introduction, there may be a relationship between shape and function. Examining shape information in the function similarity tree may help to elucidate the nature of that relationship.

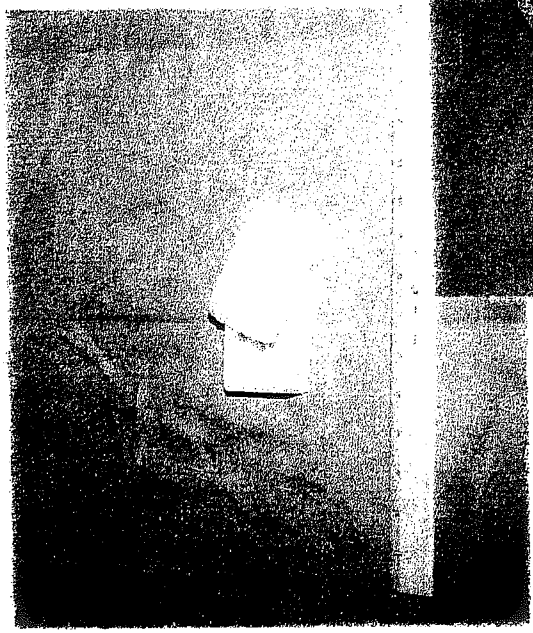
While a lengthy, technical analysis and treatment of the features issue within the trees is not the focus of this research it is possible to obtain a qualitative account of the features shared by clusters of objects in order to answer the first question posed above. To achieve this, I looked more closely at what objects fell into the clusters used in the Chi-square analysis, and searched for features common to these clusters. I used the clusters which contained the central 75% of the objects of a given linguistic category.

For functional similarity, the features which appeared to be pivotal were size, or diameter and the presence of dividers or lids. Objects with lids or dividers tended to cluster together irrespective of size even though in most cases objects tended to cluster on the basis of size. So, for example, the covered plate in Figure 7 was nearest to a small plastic soap box and a large covered bowl. These objects were not placed near each other for any of the other types of similarity. These features seem at first glance to be purely physical ones; however, I argue that this is an example of a case of some physical features, like size, being important indicators of function. This can easily be demonstrated. For example, the size of a "plate" indicates whether it can be used to serve food to a large group of people or used to serve a single person a small dessert. Other physical features, like color and translucence, are not necessarily indicators of function, or how the object is used. The other features mentioned as important for separating the categories in the tree analysis for function were lids and dividers, which I argue are more





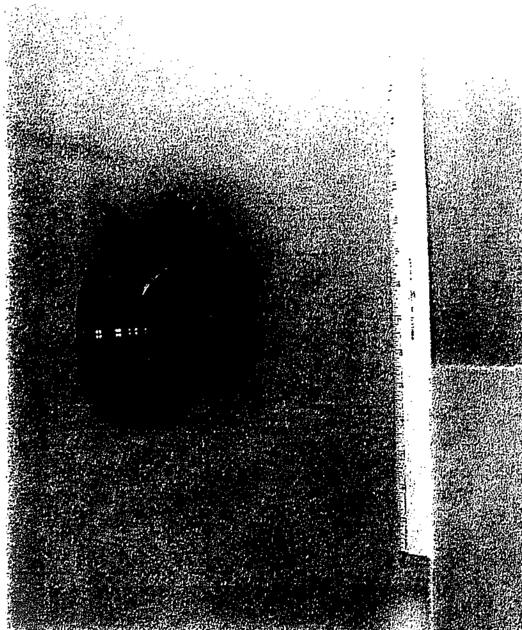


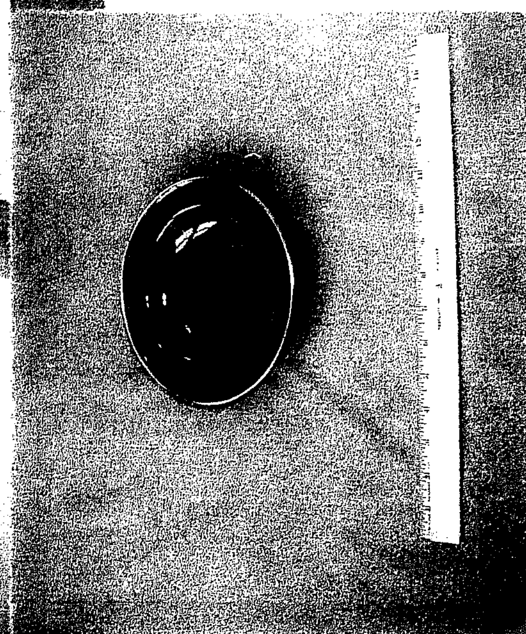


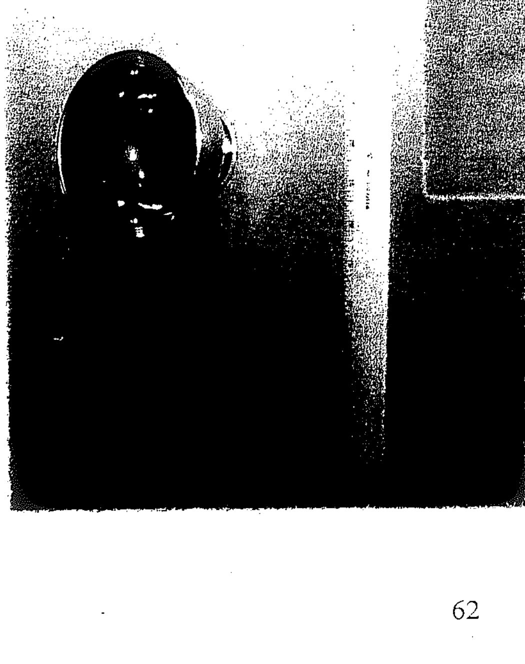
certainly functional features. This is because lids and dividers, unlike color or material, have important consequences for how an object is used. For instance, the presence of a lid indicates the potential for food storage. In contrast, color or material have little or no effect on whether or not one can use a given object to store food.

Within the tree for physical similarity the objects tended to cluster on the basis of size, color, and, opacity. Size, as I argued above, could be looked at as a physical feature that is an indicator of function, although participants were told to focus on physical similarity and to disregard function. Given that size could have functional significance, it is worth stressing that size was not the primary feature shared by clusters of objects for physical similarity. Color, in particular, had a strong effect on the clustering, and color is clearly a non-functional feature. Objects mostly clustered on the basis of size within larger clusters of color. For example, Figure 8 shows a typical small cluster of objects from the tree analysis for physical similarity. The objects depicted were all blue and were ordered within that cluster of blue objects by size.

Finally, in the overall sort, objects clustered mainly by size and also by depth of the vessel, with the physical material of construction also being important. Arguably depth of the vessel is a functional feature, but size I have argued is more difficult. It could be either a purely physical feature or a physical indicator of functional attributes of the object. The importance of the depth of the vessel feature could explain why the bowls were more clearly bounded to the upper part of the analysis, as depth played a less significant role in the other





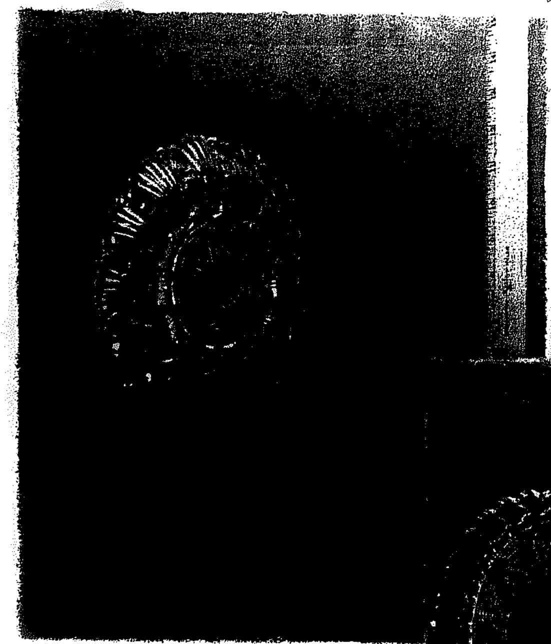


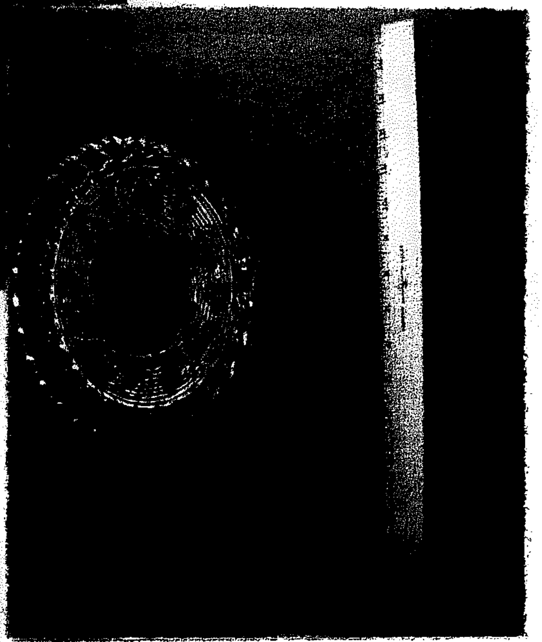
types of similarity. Indeed, few bowls are included outside that cluster of bowls at the beginning. There are, however a couple. Figure 9 is an illustration of one such example. While the bowl is made of glass, it is noticeably deeper than the objects around it. This distinction between clear glass objects and more opaque objects was also an important feature for overall similarity. Note that translucence is a purely physical feature. Again, this supports the contention that physical features are important in making category decisions, and goes against core theories.

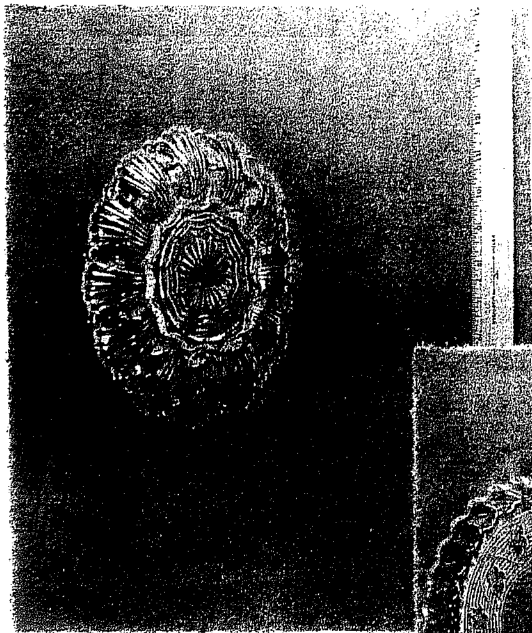
Hence overall similarity does appear to be tapping into a set of features, both physical and functional, which is superior to physical or functional similarity alone at separating category boundaries. This is as predicted and lends support to the Malt et al. framework.

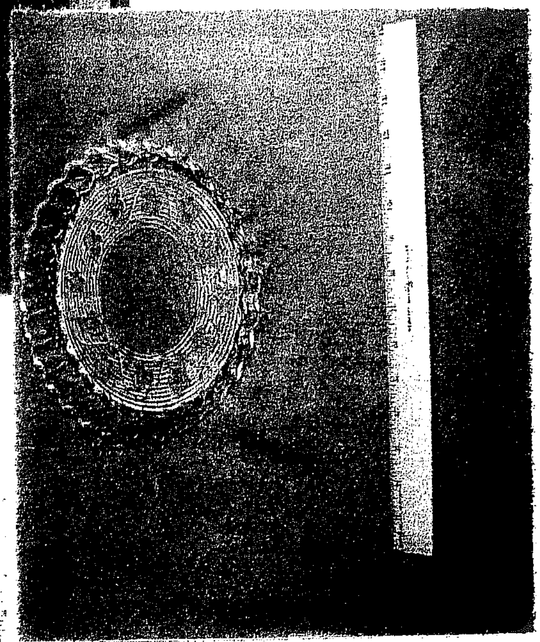
Importantly, while overall similarity produced clearly the best clustering, it was by no means perfect. For instance, interspersed within the cluster of bowls is a dish and another object labeled both a dish and a plate. Likewise, there are a few bowls within the main cluster of dishes. This too was predicted, because I maintain that no single similarity type or combination of similarity types will be sufficient for full category separation.

It does appear, from this qualitative analysis, that the manipulation of different similarity types was successful, as well as instructive. Participants did indeed focus on different features as instructed. Finally, these different features led to a differential ability to separate the categories, with overall similarity performing the best. Glancing back at the discussion for overall similarity









suggests it is not the case that all or most of the features considered by participants were functional ones. Therefore, my hypothesis that core theories are incorrect is not weakened by the types of features which seems important overall similarity judgments.

As to whether shape could be determined by function or vice versa, these results provided no clear-cut answer. In some cases, it appeared as if shape was being determined by function, but in other cases that relationship did not hold. Further research is needed to settle this issue, as there are many subtleties to be teased apart.

As another convergent test of what kinds of information were being utilized in the overall similarity condition, each individual similarity matrix was correlated with each other similarity matrix. If more functional features than physical were going into the overall similarity judgment, then a higher correlation between the functional and overall matrices than between the physical and overall matrices would be expected. In contrast, if there are more physical than functional features going into the overall similarity judgment the opposite pattern of results would be expected.

Table 5 gives the results of this correlation. The correlation between physical and overall similarity was the highest ($r = 0.87$). The correlation between functional and overall similarity was lower ($r = 0.78$). This suggests that more physical than functional features were going into the overall similarity judgment. This is consistent with the hypothesis, that physical features are important for categorization.

It is of interest that the correlation between physical and functional similarity, while lower than the other two correlations, was actually quite high ($r = 0.68$). This could be taken as more indirect evidence that some, but not all, physical features indicate, or diagnose, functional features or vice versa.

As a final point of discussion, the naming versus recognition view maintains that most, but importantly not all, of naming will be accounted for by similarity. The trees in my analysis also support this claim, because, especially with overall similarity, there are a minority of objects that appear far outside their cluster. These are the objects where similarity does not fully account for naming.

So far I have discussed the features that objects in the same cluster share, but I now wish to look at those objects which ended up outside of their linguistic category in the tree analysis. What, for instance, made a lone bowl appear in the midst of a cluster of dishes? In examining those instances I find that there is generally a feature of the object which is highly obvious and is shared mainly by objects outside of that particular objects' linguistic category. The salience of the feature is strong enough so that it is displaced to object clusters dissimilar to it in most other respects. Figure 10 illustrates one such example. The three objects depicted are adjacent in the tree analysis and occur within a cluster of 'dishes' The object at the top left and the one at the bottom are labeled 'dish' and dish/plate' respectively, while the one at the top right is called 'bowl' by the majority of participants. In most respects the bowl is different from the other two objects, but because it is cut or molded clear glass, it is placed with other cut or

	Type of Similarity		
Similarity Type	Functional	Physical	Overall
Functional	–	0.68	0.78
Physical	–	–	0.87
Overall	–	–	–

Table 11

Correlations between the three types of similarity matrices.

molded clear glass objects which are not deep and also not labeled 'bowl'. Some possible explanations of the processes which affect these objects is offered in the general discussion.

General Discussion

This study set out to examine a set of questions about what sorts of features separate linguistic categories. Core theories predict that function should be the core feature or set of features for artifact categories, and as such would delineate category boundaries best.

I had hypothesized that overall similarity would separate categories better than functional or physical features. I also expected that no single type of similarity would ever completely separate linguistic categories. This was indeed the case. Looking again at the performance of overall similarity, we see a combination of features, depth, size, and material, which does a good job of separating the categories. Notice that these features include physical ones.

Thus I can conclude that multiple features, including physical features, will likely be necessary to separate linguistic categories. This is contrary to core theories that argue that a single feature, in this case function, will separate categories.

One of the central tenets of the naming versus recognition view is that processes which affect recognition categories may be different than processes that affect naming. Two of these processes were convention and chaining. We see in the tree analysis several instances of objects appearing outside their clusters. So a bowl whose nearest neighbor is a dish warrants further examination. While the present pair of studies is not sufficient for determining why a given object appears within a cluster that does not share its linguistic category, I propose that some of these objects are there because of the two processes mentioned above. These are objects where similarity is not predicting their names. With objects like these we can begin to examine those processes which operate on naming alone. Now that this study has identified several objects whose linguistic categorization may have been influenced by these processes, later studies can be aimed at elucidating more clearly how these processes create linguistic categories which are not perfect reflections of the similarity relations between the categories.

A final point warrants discussion, in the Chi-square, for bowls and plates the pattern of results was similar; overall similarity performed best at separating the categories, while function was second best and physical was the worst at separating the categories. For dishes the pattern was somewhat different. Overall similarity was still the best at separating the dishes, but here physical actually was second best while function came in last. Why there was this reversal is not obvious, but some examination of the categories reveals a possible answer.

First, looking back at the three trees, notice that the dish category is generally the most spread out. There are numerous dishes spread throughout the trees. Even in the overall similarity tree the dish category shows the least coherent clustering. I believe the reason for this is that there are two different senses of dish. Therefore the name "dish" can apply to objects in two different ways; as a label for specific objects such as a "casserole dish" and for the superordinate category of the objects in the stimulus set. When we ask someone to "put the dishes on the table" we are referring to objects called "bowl", "plate", and probably some only called "dish".

I argue that this property could account for the pattern observed in the trees for the dish category. In the naming study, when participants were faced with problematic items such as an object which looked like a plate but seemed too big to receive the plate label, they reverted to labeling them with the superordinate "dish". Clearing this duality up would require attempting to separate objects called dish in the superordinate sense from the other objects called dish, perhaps asking participants some questions about their labeling strategies would more stringently identify these two sets. Then we could look back at the trees and see if there was actually sub-clusters of dishes for the two senses of dish.

Conclusion. In conclusion, I have demonstrated that function is not the only feature used to make category decisions. In fact, functional features were not even necessarily the primary features used in making category decisions. Core theories maintain that function will be the core for artifact categories, hence my results suggest that core theories are likely untenable as theories of artifact concepts. Instead, it appears as though the naming versus recognition view may better account for how people categorize, this theory predicted that multiple features or dimensions will be necessary to better separate the categories because naming is a process with added complexity. This process will be influenced by many factors, not just similarity along a single dimension. Overall similarity with its multiple features and dimensions was clearly the best at separating the categories, suggesting that the naming versus recognition view has merit. To the extent that physical features are the focus of prototype theories, my results weaken prototype theories as physical features were generally poorest at separating the categories.

References

Ashby, G.F. (1992). Multidimensional models of categorization. In Multidimensional models of perception and cognition, F. Gregory Ashby (Ed.). Lawrence Erlbaum Associates:Hillsdale, NJ.

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

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

Hampton, J.A. (1981). An investigation of the nature of abstract concepts. Memory and Cognition 9, 149-156.

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

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

Malt, B.C. (1990). Features and beliefs in the mental representation of categories. Journal of Memory and Language, 29, 289-315.

Malt, B.C. (1994). Water is not H20. Cognitive Psychology, 27, 41-70.

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

Malt, B.C. and Sloman S.A. (1995, November). Evaluating theories of artifact naming. Paper presented at the 36th annual meeting of the Psychonomic Society.

Malt, B.C., Sloman, S., Gennari, S., Shi, M, and Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. Journal of Memory and Language, 40, 230-262.

Medin, D.L. and Wattenmaker, W.D. (1987). Category cohesiveness, theories, and cognitive archaeology. In U. Neisser (Ed.) Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.

Medin, D.L. and Schaffer, M.M. (1978). Context theory of classification learning. Psychological Review, 85 (3), 207-238.

Miller, K.F. (1987). Geometric methods in developmental research. In J. Bisanz, C.J. Brainerd, and R. Kail (Eds.), Formal methods in developmental psychology. New York: Springer Verlag.

Murphy, G.L. and Medin, D.L. (1985). The role of theories in conceptual coherence. Psychological Review, 3, 289-316.

Nosofsky, R.M., Gluck, M.A., Palmeri, T.J., McKinley, S.C., and Glauthier, P. (1994).Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). Memory and Cognition, 22 (3), 352-369.

Rosch, E. and Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 573-603.

Sattath, S. and Tversky, A. (1977). Additive similarity trees. Psychometrika, 42(3), 319-343.

Smith, L.B. and Samuelson, L.K. (1997). Perceiving and remembering: Category stability, variability, and development. In K. Lamberts and D. Shanks (Eds.) Knowledge, concepts, and categories. Cambridge, MA: MIT Press.

Appendix A: Naming Instructions

Naming Instructions.

We will go through the pile of pictures, and I would like you to tell me, for each object, what you think you would call that object; just whatever name you think you would use for it in ordinary conversation. I will record your answer on the computer.

You can give a single word name, or the name can be more than one word. Some objects may be hard to name, but try to give an answer for each one.

Appendix B: Similarity Instructions

Function Similarity Instructions.

When you make your judgments, I'd like you to focus on the FUNCTION or USE of each object; that is, how it is used to prepare and serve or hold food or other items (containing a liquid, with solid items resting on it, etc.) I'd like you to judge as similar those objects that you think are very similar to each other in how they FUNCTION, and judge as less similar the ones that are less similar in FUNCTION.

Note that we are interested in how similar the OBJECTS themselves are, not what is prepared or served in the objects. Only judge two pictures as very similar if the objects are like each other in how they work. Do NOT judge them as similar just because the objects are used to serve or prepare things which are found together. For instance, if one object is for preparing salad and another is for serving or eating salad from, or if one is for cooking casserole and one is used to eat casserole from, DON'T call them similar unless you really think the objects themselves perform their function in a similar way.

Physical Similarity Instructions.

When you make your judgments, I'd like you to focus on the PHYSICAL QUALITIES of each object. To do so, use just those properties that you can see from looking at the picture. Specifically, consider the object's SHAPE, SIZE, COLOR, TEXTURE, and RELATIVE SHININESS. Please don't use any other properties you might know about from having seen the objects before; focus only on the information provided by the picture. I'd like you to judge as similar those objects that you think are very similar to each other PHYSICALLY and judge as less similar the ones that are less similar PHYSICALLY.

Note that we are interested in how physically similar the OBJECTS themselves are, not what is prepared or served in the objects. Only judge two pictures as very similar if the objects are like each other in the physical properties listed above. Do NOT judge them as similar just because the objects are used to serve or prepare things which are found together. For instance, if one object is for preparing salad and another is for serving or eating salad from, or if one is for cooking casseroles an one is used to eat casseroles from, DON'T call them highly similar unless you really think the objects themselves are alike in their shape. Size, color, texture, and relative shininess.

<u>Overall Similarity Instructions.</u>

When you make your judgments, I'd like you to consider ALL the qualities of each object; that is, all its features including what it looks like, how it holds any substances (with solid objects placed on it or inside it, holding a liquid, etc.) And any other aspect of the object that seems important or natural to you. I'd like you to judge as similar those objects that you think are very similar to each other OVERALL, and judge as less similar the ones that are less similar OVERALL.

Note that we are interested in how similar the OBJECTS themselves are, not what is prepared or served in the objects. Only judge two pictures as very similar if the objects are similar to each other in all respects. Do NOT judge them as similar just because the objects are used to prepare or serve things which are found together. For instance, if one object is for preparing salad and another is for serving salad or eating salad from, or if one is for cooking casserole and another is used to eat casserole from, DON'T call them highly similar unless you really think the objects themselves are similar in all respects.

Vita
Molly J. Stanton

Biographical

Birthdate:	24 May 1973
Birthplace:	Ontario, Oregon
University Address:	Department of Psychology 17 Lehigh University Bethlehem, PA 18015
Parent's Names:	Michael and Rebecca Stanton

Education

1991-1995	Smith College, Northampton, MA B.A., Psychology 1995.
1995-date	Lehigh University, Bethlehem, PA Academic Advisor: Barbara Malt, Ph.D.

Academic Awards

1992-1995	Smith College Alumni Association Scholarship.
1993-1995	Smith College Dean's List.
1994-1995	Smith College First Group Scholar.
1994-1995	Inducted an associate member of Sigma Xi, The Scientific Research Society.
1994-1995	Inducted as a member of Psi Chi, The Psychology Honor Society.

Membership in Professional Organizations

American Psychological Association

Professional Experience

Teaching Experience:

Student Instructor, Smith College, Northampton, MA 01063. Spring 1994 and Spring 1995. Introductory Psychology.

Academic Tutor, Smith College Center for Academic Development, Northampton, MA 01063, 1994-1995, Statistical Methods in Psychology, Introductory Physics.

Teaching Assistant, Lehigh University, Bethlehem, PA 18015, 1995-1996; 1998-1999, Introductory Psychology recitations, Personality, Child Development, Personality recitations, Research Methods.

Research Experience:

Research Assistant to Dr. Donald Reutener, Smith College, Department of Psychology, Northampton, MA 01063, 1994-1995.

Research Assistant to Dr. Jill deVilliers, Smith College, Department of Psychology, Northampton, MA 01063, Spring 1995.

Research Assistant to Dr. Barbara Malt, Lehigh University, Department of Psychology, Bethlehem, PA 01063, 1996-1998.

END

OF

TITLE