Title
Categories

Author
Lakoff, George

Publication Date
1982

Peer reviewed

CATEGORIES

An Essay in Cognitive Linguistics

George Lakoff

Linguistics Department
and
Cognitive Science Program

University of California at Berkeley

to appear in

LINGUISTICS IN THE MORNING CALM

edited by In-Seok Yang

PREFACE

This paper does not cover the same ground as my SICOL '81 lectures on Cognitive Semantics. In those lectures, I was introducing to the Korean Linguistic Community work on Cognitive Linguistics that I and others had already completed, but was either unpublished or unwritten. Now, almost a year later, the unpublished and unwritten works have appeared, and I feel no need to rehash them. Most of the content of those lectures appears in Lakoff and Johnson (1980), Lakoff (1982), Lindner (1981,1982), Brugman (1981), and Lawler and Rhodes (1981).

I have decided instead to write a new paper for this volume, one that addresses issues that arose in my lengthy discussions with my Korean colleagues. The issues have to do not only with the foundations of Linguistics, but with the foundations of Cognitive Science in general. They concern the nature of categorization, which is as fundamental an issue as there is.

On the train to Taegu, two Korean colleagues who I admire immensely sat down next to me and asked me whether I had given up on formal linguistics. I said no. I had given up on a certain specialized conception of formal linguistics, not formal linguistics itself. Formal linguistics involves the following:

-The study of linguistic form.

-A notion of systematicity.

-A commitment to appropriate precision, which may include the development of notations which lead one to further insights.

My current work involves all of these. What I have given up on is a concept of formal syntax which is based on what I consider a bizarre metaphor: A GRAMMAR IS A RESTRICTED POST SYSTEM. This is the basic metaphor behind generative

linguistics, and it has consequences that I have found inconsistent with the linguistic data I have been concerned with for the past two decades. In this metaphor, a "language" is taken to be a set of "sentences", where a "sentence" is a sequence (that is, an ordered set) of uninterpreted, arbitrary symbols. All of the work on restricting the theory of formal syntax presupposes this without question. Among the consequences of the metaphor are:

- Linguistic structure is independent of cognition in general.
- Set theory is correct as an account of linguistic categorization.

The first follows since no interpretation of the symbols in cognitive terms can be made use of in a grammar, if results about generative power are to be maintained. And rules cannot make use of any aspect of general cognition for the same reason. The second conclusion follows, since set-theoretical categorization is all that the metaphor permits.

Moreover, formal semantics in the Fregean tradition makes basically the same assumptions, including the assumption that categories have sets as extensions and that concepts are to be characterized in set-theoretical terms. The edifices of generative syntax and formal semantics are built on classical set theory as a means of categorization.

But empirical studies both in psychology and linguistics have thrown into the question the assumption that set theory can serve as an adequate theory of categorization, both for language and for linguistics in general. Moreover, a cognitively real theory of language ought, it seems to me, to assume that human beings make use of their general cognitive capacities in language, rather than that they ignore them, as the RESTRICTED POST SYSTEM metaphor requires. Chomsky's claims about modularity in cognition seem to be a last-ditch attempt to salvage the RESTRICTED POST SYSTEM metaphor as a theory of linguistic structure, in the face of mounting evidence indicating that aspects of general

cognition are involved in language acquisition, as well as in the structure of language itself.

Among the empirical results that throw the foundations of generative linguistics and Fregean semantics in question are those concerning natural categorization. Having spent fifteen years investigating the frameworks of generative linguistics and Fregean semantics, and trying to make them fit the facts of language, I have become convinced that trying to make them work is pointless. The result has been to narrow the confines of linguistic investigation to those phenomena that have some hope of fitting, and ignoring not only most of language, but most of the phenomena that were being studied ten to fifteen years ago.

I am convinced that the results concerning natural categorization are basically correct, and am committed to finding out what changes they require in our views on human conceptual structure, human reason, and human language. I maintain my commitment to formal linguistics in the sense outlined above, but reject any concept of formal linguistics that is incompatible with the empirical results on natural categorization. Natural categorization has been established as an important aspect of general cognition, both within and outside of linguistics. Cognitive Linguistics is dedicated to discovering the ways in which language uses general cognitive mechanisms. At the same time, it proposes that the study of language can contribute to an understanding of such mechanisms. This is in sharp contrast with generative linguistics, which assumes that the linguistic system is independent of general cognitive mechanisms. From this assumption, it follows that generative linguistics has no contribution to make to the study of general cognition. As a working hypothesis, cognitive linguistics is more interesting, since, to the extent that it is correct, it makes linguistic phenomena important for the study of cognition in general.

Like many linguists, in America as well as Korea, my hosts were not familiar with all of the basic ideas of the theory of natural categorization, nor with the evidence supporting it. Moreover, they quite reasonably could not imagine a formal linguistics not based on a set-theoretical view of categorization.

It seems both urgent and to the point that I address such foundational issues as well as I can at the moment. The need for such a paper was made especially clear to me by a remark made to my Berkeley colleague Charles Fillmore at the conference by an American linguist who had heard Fillmore's paper. Fillmore had discussed a number of examples where natural categorization, which is often called prototype theory, was required for certain problems in lexical semantics. The American linguist asked Fillmore why he was still bothering with prototype theory when it had been thoroughly discredited in a paper by two psychologists, Osherson and Smith (1981).

Neither Fillmore nor I had seen the Osherson and Smith paper. Upon my return to America, I read the Osherson-Smith paper and decided that it confirmed, rather than discredited, the theory of natural categorization. However, it was an interesting paper that needed to be addressed seriously, especially since it indicated that psychologists were not aware of the linguistic studies in this area.

The present paper is an attempt to clarify the issues in the theory of categorization. It is also an attempt on my part to bring together as coherently as possible some ideas by my colleagues and myself on what an adequate theory of natural categorization would be like. The approach is called the theory of idealized cognitive models (or ICMs). It bears family resemblances to other ideas that have been used in cognitive linguistics in recent years: schemas, frames, linguistic gestalts, functional assemblies.

My debt to my colleagues -- faculty, postdoctoral fellows, and students -- in

the Berkeley Cognitive Science Program will be obvious throughout this paper. I would especially like to thank Ronald Langacker and the students from his Cognitive Grammar group at the University of California at San Diego for sharing their results with the members of our group. It is especially encouraging to me that the San Diego group, working independently, has arrived at such similar conclusions. And still more encouraging to find somewhat similar results emerging in Korea, in the work of Professor Kee Dong Lee.

THE CLASSICAL THEORY OF CATEGORIZATION

People use concepts to categorize things, and they act on those categorizations. Without the ability to conceptualize and categorize, we could not function at all, either in the physical world or in our social or intellectual lives. The theory of categorization is therefore central to any understanding of our conceptual system, and therefore necessary to any understanding of how we human beings function and what makes us human.

Most categorization is automatic and unconscious, and if we become aware of it at all, it is only in problematic cases. In moving about the world, we automatically categorize people, animals and physical objects, both natural and man-made. This sometimes leads to the impression that we just categorize things as they are, that things come in natural kinds, and that our categories of mind fit the kinds of things in the world. But a large part of our categorization, and maybe most of it, is not of this kind at all. It is abstract. We categorize events, actions, perceptions, emotions, spatial relationships, social relationships, and abstract entities of an enormous range: governments, illnesses, social practices and entities in both scientific and folk theories -- like electrons and colds. Any adequate account of the human conceptual system must provide an accurate theory for all our categorization, both concrete and abstract.

A good example of abstract categorization that is automatic and almost entirely unconscious occurs in language. Each human language is structured in terms of an enormously complex system of categories of various kinds: phonetic, phonological, morphological, lexical, syntactic, semantic, and pragmatic. Linguistic categories are among the kinds of abstract categories that any adequate theory of the human conceptual system must be able to account for. Linguistics is then an important source of evidence for the nature of cognitive categories. Conversely, general results concerning the nature of cognitive

categorization should apply to categories in linguistics. Linguistic theory is therefore very much bound up with general issues in cognition -- and none is more central than categorization.

For two millenia, categories were assumed within the Western tradtion to be well-understood. According to the Aristotelian tradition, concepts were defined by necessary and sufficient conditions, and people categorized things in terms of such concepts. The modern incarnation of this tradition should probably be credited to Gottlob Frege. Within Fregean semantics, set theory is the central tool for categorizing concepts and categories. Within set theory, a set can either be characterized by a list of its members or by a set of necessary and sufficient conditions for membership. Fregean models of reality consist of entities and sets. The extensions of concepts are sets of entities (or sets of sequences of entities.) The "meanings" or intensions of concepts are functions from possible situations to extensions. Sequences and functions are also defined in terms of sets, and situations are defined in terms of entities and sets. Fregean semantics is all entities and sets. The classical operations on sets are intersection, union and complementation. In recent work more exotic operations have been proposed, but they all involve functions, which are defined in terms of sets. Sets are at the heart of all modern versions of the classical theory of categorization.

This is also true of every aspect of generative linguistics, whether phonology, syntax or semantics. In generative phonology, distinctive features correspond to sets. Segments marked +F are in the set, and those marked -F are in the complement of the set. In generative phonological notation, square brackets indicate set intersection, and curly brackets indicate set union. The same is true of syntactic features, and since they define syntactic categories, syntactic categories are defined within generative linguistics in terms of classi-

cal sets. A language, within generative linguistics, is defined as a set of sentences, and a grammar as a set of rules that characterizes the set of sentences. The setences are sequences (ordered sets) of phonological feature matrices. The semantics is Fregean. In virtually every respect, generative linguistics rests on the classical theory of categorization as it has been interpreted in the Fregean tradition -- the assumption that the humanly relevant notion of a category can be adequately represented by the concept of a set.

Within recent years the classical theory has been challenged seriously within every branch of Cognitive Science: within Philosophy (by Wittgenstein and Putnam), within Psychology (by Rosch and Mervis), within Anthropology (By Berlin,Kay, McDaniel and Kempton) within Artificial Intelligence (by Zadeh and Winograd), and within Linguistics (by Lakoff, Ross, Fillmore, Labov, Langacker, Lindner, Brugman, Sweetser and Jaeger). Because the concept of a category is so central to all the cognitive sciences, these challenges, taken together, pose a serious threat to business as usual in these fields. This is especially true in Linguistics, which is currently dominated by generative approaches, which rely in virtually every detail on the classical theory of categorization. The collapse of the classical theory of categorization would pose nothing less than a major foundational crisis for generative linguistics. Yet, remarkably enough, the evidence against the classical theory has come at least as much from linguistics as from any other discipline.

Wittgenstein is usually cited as the source of the challenge. In the Philosophical Investigations (Wittgenstein 1953) he argued that the concept *game* is so heterogeneous that there could be no precise necessary and sufficient conditions for something to be a game. Instead he suggested that games bear "family resemblances" to one another, and it is on the basis of such resemblances that we consider activities as diverse as ring-around-the-rosie, solitaire, Monopoly,

and football as all being instances of the category *game*.

The more recent challenges differ somewhat from one another and from Wittgenstein's in their details. The two most widely discussed are those by Rosch and Zadeh, and it will be most convenient to start with their work and to move to other results later.

THE THEORY OF NATURAL CATEGORIZATION

Let us begin with Rosch's results. Their significance will be most apparent if we give as background some of the consequences of the classical view, as interpreted within the Fregean tradition. The classical tradition has two important classes of consequences:

I. CLEAR BOUNDARIES, SHARED PROPERTIES,

UNIFORMITY, AND INFLEXIBLILITY

II. OBJECTIVISM AND REDUCTIONISM

CLEAR BOUNDARIES: Everything is, or is not, a member of a category. There are no degrees of membership, borderline cases, or fuzziness of any kind.

SHARED PROPERTIES: There are necessary and sufficient conditions for category membership. All members of the category have something in common in that they meet these conditions.This is sometimes called a "checklist", since it amounts to a listing of the properties shared by all members of the category.

UNIFORMITY: No special status is conferred on any of the members of the category. All category members are equal. Similarly, no special status is accorded to any of the necessary and sufficient conditions for category membership. All conditions for category membership are equal: No condition for membership is more important than any other, and there are no different types of conditions.

INFLEXIBILITY: Category boundaries do not vary. Human purposes, features of context, etc. do not change category boundaries.

OBJECTIVISM: The Fregean view is part of an "objectivist" theory of meaning. All psychological factors – perception, mental images, human purposes, etc. –are ruled out. The world is assumed to be made up of objects with inherent properties and fixed relationships among them at any instant. Category membership is determined by *objective* necessary and sufficient conditions, which are made up of inherent properties and fixed relationships. No "psychological" properties or relationships play a role; e.g., how things are perceived, how people interact with objects, what people's intentions are, what their mental images are, etc. In any situation, the extension of a category is a set, consisting of the objects that meet the objective necessary and sufficient conditions for membership in the set.

REDUCTIONISM: The meanings of complex expressions are reducible to the meanings of simple expressions plus principles of combination. Hence, there are primitive irreducible predicates. Corresponding to such predicates are primitive categories. Complex categories are the result of operations on primitive categories. These operations are classically taken to be simple operations on sets – intersection, union, and complementation. These correspond to conjunction, disjunction, and negation in propositional logic. Some recent nonclassical extensions of the Fregean paradigm have proposed other, more complex kinds of operations.

Rosch's experimental results appear, at least superficially, to be at variance with **all** of these aspects of the classical theory. Her experimental results may be grouped into two types:

I. PROTOTYPE RESULTS

II. BASIC-LEVEL RESULTS

PROTOTYPE RESULTS:

-Some members of a category are judged by subjects to be more representative of the category than other members. For example, robins are judged to be more representative of the category BIRD than are chickens, penguins, and ostriches; and desk chairs are judged to be more representative of the category CHAIR than are rocking chairs, barber chairs, beanbag chairs or electric chairs. The most representative members of a category are often called "prototypical" members.

-Subjects give consistent goodness-of-example ratings across experimental paradigms. Some of the paradigms are:

Direct Rating: Subjects are asked to rate, say on a scale from one to seven, how good an example of a category (e.g.,BIRD) various members are (e.g., a robin, a chicken, etc.).

Reaction Time: Subjects are asked to press a button to indicate true or false in response to a statement of the form "An [example] is a [category name]" (e.g., "A chicken is a bird"). Response times are shorter for representative examples.

Production of examples: When asked to list or draw examples of category members, subjects were more likely to list or draw more representative examples.

Asymmetry in generalization: New information about a representative category member is more likely to be generalized to nonrepresentative members than the reverse. For example, when told that robins on an island had a disease, they were more likely to decide that ducks would catch it than that robins would catch the disease that the ducks had.

-These ratings correlated with "family resemblances", that is, perceived similarities between representative and nonrepresentative members.

- Category membership seems to characterized not by necessary and sufficient conditions, but by clusters of attributes that characterize the most representative members. None of these attributes need be either necessary or sufficient for category membership. And it may be that no single member of the category has all the attributes in the cluster. Some attributes are typically more important for category membership than others. That is, attributes in general do not have equal status. It may even be the case that two nonrepresentative members have no relevant common attributes and are members of the category only by virtue of bearing family resemblances of very different kinds to representative members.

-Representative members serve as "cognitive reference points" for certain kinds of reasoning and other tasks.

-Category boundaries are indeterminate. There is a good deal of variation at the edge of categories.

-The properties of representative members do not determine category membership as a whole. An object is not necessarily a member of a category to some degree just because it bears some degree of similarity to prototypical members. Pigs aren't nonrepresentative examples of dogs just because they bear some similarity to prototypical dogs. Neighboring categories have an effect on category membership. Moreover, near the edge of a category, there may be arbitrary cultural conventions that determine category boundaries.

BASIC-LEVEL RESULTS:

-Certain categories are psychologically more "basic" than others: they are recognized more rapidly, learned earlier, used more frequently, have shorter names, etc. In American Sign Language, basic level categories are generally denoted by single signs, while superordinate and subordinate categories are almost always denoted by multiple sign sequences.

-These categories are set-theortically "in the middle", in the sense that they are not primitive, and have both subordinate and superordinate categories. Some examples:

SUPERORDINATE	MAMMAL	FURNITURE
BASIC-LEVEL	DOG	CHAIR
SUBORDINATE	RETRIEVER	ROCKING CHAIR

-The basic level is the most general level at which

(a) a person uses similar motor actions for interacting with category members

(b) people perceive category members as having similar overall shapes

(c) a mental image can reflect the entire category.

In general, basicness depends upon *perceived attribute structures*. In what follows I will refer to general motor actions, perceived shapes, mental images, intentions and functions as *interactional properties* .

-Basic levels are not absolute, but can vary as a function of both cultural significance of the domain and the level of expertise of the individual.

Rosch's hypothesis: At the basic level, the information value of attribute clusters is maximized. This is the level at which categories maximize within-category similarity (of relevant interactional properties) relative to between-category similarity.

Note that basic levels are characterized relative to entire systems of categories, rather than on a category-by-category basis.

One way of thinking about basic level categories is that they are "human sized". They depend not on objects themselves, independent of people, but on the way people interact with objects: the way the perceive them, image them, and use motor actions on them. The relevant properties clustering together to define such categories are not inherent to the objects, but are interactional properties, having to do with the way people interact with objects.

Basic level categories have somewhat different properties than superordinate categories. For example, superordinate categories seem not to be characterized by images or motor actions. For example, we have mental images of chairs -- abstract images that don't fit any particular chair -- and we have general motor actions for sitting in chairs. But if we go from the basic level category CHAIR to the superordinate category FURNITURE, a difference emerges. We have no abstract mental images of furniture that are not images of basic level objects like chairs, tables, beds, etc. Try to imagine a piece of furniture that doesn't look like a chair, or table, or bed, etc., but is more abstract. People seem not to be able to do so. Moreover, we do not have motor actions for interacting with furniture in general that are not motor actions for interacting with some basic level object -- chairs, tables, beds, etc. But superordinate categories do have other humanly-based attributes -- like purposes and functions.

In addition, the complements of basic level categories are not basic-level. They do not have the kinds of properties that basic-level categories have. For example, consider nonchairs. What do they look like? Do you have a mental image of a general, or an abstract nonchair? People seem not to. How do you interact with a nonchair? Is there some general motor action one performs with nonchairs? Apparently not. What is a nonchair used for? Do nonchairs have

general functions? Apparently not.

In the classical theory, the complement of a set that is defined by necessary and sufficient conditions is another set that is defined by necessary and sufficient conditions. But the complement of a basic-level category is not itself a basic level category.

We can now see just how Rosch's experimental results appear to clash with the classical theory:

I. THE PROTOTYPE RESULTS ARE INCONSISTENT WITH CLEAR BOUNDARIES, SHARED PROPERTIES, UNIFORMITY, AND INFLEXIBILITY.

II. THE BASIC-LEVEL RESULTS ARE INCONSISTENT WITH OBJECTIVISM AND REDUCTIONISM.

The range of experimental results presented above limits the range of possible theories of categorization. We will call any theory consistent with such results a *theory of natural categorization*. Rosch's views are also referred to as "prototype theory", as though it were a single, completely specified theory. This is somewhat misleading, since these results are open to many specific theoretical interpretations, and Rosch has taken pains not to be any more specific than is warranted by experimental results.

CLASSICAL FUZZY SET THEORY

Let us now turn to a particular theory of category membership, Zadeh's fuzzy set theory, as presented in Zadeh (1965). Fuzzy set theory is an extension of classical set theory to allow for degrees of category membership between 0 and 1. Every entity in the universe of discourse has a degree of membership between 0 and 1 in every fuzzy set. This is its membership value for that set. In the original version of fuzzy set theory, operations on fuzzy sets are defined in a

very straightforward way. Intersection is defined by taking the minimum of the membership values for the two sets. Set union takes the maximum of the values. And set complementation takes 1 - the membership value. It is a natural and ingenious extension of the classical theory, to get around the CLEAR BOUNDARIES condition of the classical theory.

It should be noted that the original version of fuzzy set theory preceded Rosch's results by several years. It was not intended to do anything more ambitious than to provide a natural way to get around the CLEAR BOUNDARIES condition. It does not, and was not intended to, provide an adequate account of most of Rosch's results. More recent versions have attempted to model some of the prototype results. But the spirit of fuzzy set theory seems not at all to be in accord with Rosch's BASIC-LEVEL results. Fuzzy set theory is both objectivist and reductionistic in all of the versions I know about. It is Fregean in spirit, and is likely to remain that way. I view it as an extension of the classical theory rather than a basic challenge to it.

THE OSHERSON-AND-SMITH DEFENSE OF THE CLASSICAL THEORY

Rosch's experimental results have been replicated repeatedly, and extended. To my knowledge, the experimental results themselves have not been challenged. However, her interpretation of the results has been challenged vigorously by defenders of the classical theory. As we have noted, the fall of the classical theory would have serious consequences throughout the cognitive sciences, especially in Linguistics. It is therefore no surprise to find attempts to reconcile the classical theory with Rosch's results. The most sustained defense to date is that of Osherson and Smith (1981). The O&S paper is of particular interest because it doesn't merely make an arbitrary counterproposal to keep the classical theory while accounting for Rosch's results. What it provides is a counterproposal of a very general type. As I will endeavor to show, it is a

counterproposal that fails, and that the way it fails suggests that all counterproposals of that type will fail as well.

The O&S paper has two parts:

1. A putative argument against "prototype theory".
2. The presentation of their alternative.

The paper raises an extremely important issue: What are complex categories like and how are they related to less complex categories? The paper proceeds from a number of assumptions, some of which are unstated. The basic assumptions are:

ASSUMPTION 1: OBJECTIVISM

The world is made up of objects, with inherent properties and fixed relationships among them at any time.

A statement is true if it objectively fits the world.

Meaning is based on objective truth.

ASSUMPTION 2: REDUCTIONISM

The meaning of a complex expression is obtained from the meaning of simple expressions by completely general compositional functions.

ASSUMPTION 3: EXCLUSIVENESS

Assumption 2 is the only way to account for the fact that people can comprehend an infinity of new expressions.

ASSUMPTION 4: PROTOTYPE THEORY EQUALS CLASSICAL FUZZY SET THEORY

Zadeh's original version of fuzzy set theory accurately represents Rosch's

claims about categorization. The Fregean paradigm (including assumptions 1 and 2) extended to fuzzy set theory is the *only* way to characterize a theory of meaning based on Rosch's results.

ASSUMPTION 5: OBJECTIVE SIMILARITY DETERMINES CATEGORY MEMBERSHIP

Prototype theory claims that the degree of category membership for a given object is determined solely by the degree of objective similarity between that object and a prototypical member of the category.

Each of these assumptions is either gratuitous, or under serious challenge, or just plain false. Assumptions 4 and 5 are just plain false. Let us begin with Assumption 4.

As we saw above, Zadeh's original version of fuzzy set theory attempted to account only for the lack of clear boundaries in a category.It says nothing about family resemblances, which are at the heart of Rosch's prototype results. It does not attempt to account for any of Rosch's other results, especially the BASIC-LEVEL results, which are inconsistent with both objectivism and reductionism. Since fuzzy set theory is both objectivist and reductionistic, it cannot possibly provide an accurate rendering of Rosch's basic level results. And the original version of fuzzy set theory could not, and did not try to, account for FAMILY RESEMBLANCES, NONUNIFORMITY and FLEXIBILITY. Assumption 4 can be made only by ignoring most of the most interesting aspects of prototype theory.

Assumption 5 is also false. Although Rosch has claimed that nonrepresentative members bear family resemblances to representative members, she nowhere claims that mere similarity to a representative member of a category is enough to guarantee an object membership in that category. In fact, she goes to some lengths to point out that the existence of neighboring categories, as well

as contextual factors and arbitrary conventions override similarity in the case of elements which are not representative members.

Assumptions 1 and 2 are made without justification. They contradict Rosch's BASIC-LEVEL results. By making Assumptions 1 and 2, O&S are assuming a significant part of what they set out to prove. Moreover, Assumptions 1 and 2 are under serious challenge not just from Rosch's BASIC-LEVEL results, but from other quarters as well. Objectivism has been challenged by the results of Lakoff and Johnson (1980) and Quinn (1981). Reductionism has been challenged by Fillmore (1978), Lakoff(1977) and Langacker (1982). Though it is not at all uncommon for many philosphers and linguists to make Assumptions 1 and 2, it is particularly inappropriate for O&S to assume them without question in this setting, where they are among the things at issue.

Assumption 3 is gratuitous. Other suggestions for accounting for complex categorization have been suggested, e.g., by Fillmore (1975), Lakoff (1977) and Langacker (1982).

However, if we grant O&S all their assumptions, it turns out that they do make a correct and interesting observation:

-The Fregean paradigm, extended to fuzzy set theory, makes incorrect predictions about the understanding of complex categories.

The examples they give are well worth considering. Like classical set theory, classical fuzzy set theory has only three ways of forming complex categories: intersection, union, and complementation. O&S take each of these and show that they lead to incorrect results. Their first counterexample involves three drawings:

a. A line drawing of a normally-shaped apple with stripes superimposed on

the apple.

b. A line drawing of a normally-shaped apple.

c. A line drawing of an abnormally-shaped apple with only some stripes.

They now consider three concepts: *apple*, *striped*, and *striped apple*. They correctly observe that within classical fuzzy set theory there is only one way to derive the complex category *striped apple* from the categories *apple* and *striped*, namely by intersection of fuzzy sets – which is defined by taking the minimum of the membership values in the two component fuzzy sets.

They assume the following:

- (a) is a good example of a striped apple

- (a) is not a good example of an apple, since apples generally aren't striped.

- (a) is not a good example of a striped thing, since apples are not among the things that are typically striped.

It follows that:

(a) will have a high value in the category *striped apple*.

(a) will have a low value in the category *apple*.

(a) a will have a low value in the category *striped*.

But since the minimum of two low values is a low value, it should follow from fuzzy set theory that (a) has a low value in the category *striped apple*. Thus fuzzy set theory makes an incorrect prediction. It predicts that an excellent example of a striped apple will have a low value in that category, since it has low values in the component categories *apple* and *striped*.

There is a general moral here: GOOD EXAMPLES OF COMPLEX CATEGORIES ARE OFTEN BAD EXAMPLES OF COMPONENT CATEGORIES.

O&S cite a similar example: *pet fish*. A guppie might be a good example of a pet fish, but a bad example of a pet and a bad example of a fish. Set intersection in classical fuzzy set theory will give incorrect results in such cases.

O&S also use some of what might be called "logician's examples":

A AND NOT A: an apple that is not an apple

A OR NOT A: a fruit that either is, or is not, an apple

They assume the correctness of the usual logican's intuitions about such cases: There is no apple that isn't an apple and so the first category should have no members to any degree; and all fruits either are or are not apples, so the second category should contain all fruits as full-fledged members. Such intuitions have been disputed: a carved wooden apple might be considered an apple that is not an apple. And a cross between a pear and an apple might be considered a bad example of a fruit that clearly either is, or is not, an apple. O&S do not consider such possibilities. They correctly argue that classical fuzzy set theory cannot account for the usual logician's intuitions in such cases.

The argument goes like this. Take an apple that is not a representative example of an apple, say a crabapple. According to classical fuzzy set theory, this would have a value in the category *apple* somewhere in between zero and 1. Call the value 'c'. Its value in the category *not an apple* would then be 1 - c, according to the definition of set complementation in fuzzy set theory. If c is in between zero and 1, 1 - c will also be between zero and 1. And both the maximum and the minimum of c and 1 - c will be in between zero and 1. Thus, according to fuzzy set theory, a nonrepresentive apple, like a crabapple, would have a value

greater than zero in the category *an apple that is not an apple*, and it would have a value less than one in the category *a fruit that either is, or is not, an apple*. This is inconsistent with the intuitions assumed to be correct by O&S.

O&S's last major argument depends crucially on their fallacious Assumption 5: In prototype theory, degree of membership is determined by degree of similarity to a prototypical member. They correctly produce a counterexample to this nonexistent principle of prototype theory, taking classical fuzzy set theory to represent prototype theory. It is based on the following use of Assumption 5. Consider grizzly bears and squirrels. Since one can find some (possible small) similarities between grizzly bears and squirrels, it follows by Assumption 5 that squirrels are members of the category *grizzly bear* to some degree greater than zero. Now consider the statement:

-All grizzly bear are inhabitants of North America.

Suppose someone were to find a squirrel on Mars. Since that squirrel is a member of the category *grizzly bear* to some extent, and since Mars is far from North America, the discovery of a squirrel on Mars would serve as disconfirmation of the claim that all grizzly bears are inhabitants of North America. But this is ridiculous. The existence of squirrels on Mars should have nothing to do with the truth or falsity of that statement. O&S take this to be a counterexample to prototype theory. Of course, it has nothing to do with prototype theory, since prototype theory does not make Assumption 5.

O&S consider themselves as having refuted prototype theory as a viable theory of concepts. As we have seen, they have said nothing whatever about prototype theory, but they have shown that Fregean semantics extended to classical fuzzy set theory won't work very well.

O&S then present their own proposal for saving the classical theory while

accounting for the experimental results of prototype theory. What they propose is a hybrid theory: each concept has a *core* and an *identification procedure*. The core is works according to the traditional theory; the identification procedure accounts for the empirically validated prototype affects. As they put it:

> The core is concerned with those aspects of a concept that explicate its relation to other concepts, and to thoughts, while the identification procedure specifies the kind of information used to make rapid decisions about membership... We can illustrate this with the concept *woman*. Its core might contain information about the presence of a reproductive system, while its identification procedures might contain information about body shape, hair length, and voice pitch.

The core, in other words, would be where the real work of the mind -- thought -- is done. The identification procedure would link the mind to the senses, but not do any real conceptual work. As they say,

> Given this distinction it is possible that some traditional theory of concepts correctly characterizes the core, whereas prototype theory characterizes an important identification procedure. This would explain why prototype theory does well in explicating the real-time process of determining category membership (a job for identification procedures), but fares badly in explicating conceptual combination and the truth conditions of thoughts (a job for concept cores).

The hybrid theory assumes that traditional theories actually work for complex concepts. The fact is that this is one of the most notorious weaknesses of traditional theories. The only traditional theories there are are based on classical set theory. Such theories permit set-theoretical intersection, union, and complement operations, and occasionally a small number of additional operations. But on the whole they do very badly at accounting for complex categorization. We can see the

problems best by looking first at the classical theory, unextended by additional operations. The traditional set-theoretical treatment of adjective-noun phrases is via set intersection. That is the only option the traditional theory makes available. So, in the classical theory, the complex concept *striped apple* would denote the intersection of the set of striped things and the set of apples.

The literature on linguistic semantics is replete with examples where simple set intersection will not work. Perhaps we should start with some that O&S themselves mention (fn.8,p.43 and fn.12, p.50).

small galaxy -- not the intersection of the set of small things and the set of galaxies

good thief – not the intersection of the set of good things and the set of thieves

imitation brass – not the intersection of the set of imitations and the set of brass things

Other classic examples abound:

electrical engineer -- not the intersection of the set of electrical things and the set of engineers

mere child – not the intersection of the set of mere things and the set of children

red hair – since the color is not focal red, it is not merely the intersection of the set of red things and the set of hairs.

happy coincidence -- not the intersection of the set of happy things and the set of coincidences

topless bar – not the intersection of the set of topless things and the set of bars

heavy price – not the intersection of the set of heavy things and the set of prices

past president – not the intersection of the set of past things and the set of presidents

set theory – not the intersection of the set of sets and the set of theories

Such examples can be multiplied indefinitely. There is nothing new about them, and no serious student of linguistic semantics would claim that such cases could be handled by intersection in traditional set theory. At present there is no adequate account of most kinds of complex concepts within a traditional framework, though a small number of isolated analyses using nonstandard set-theoretical apparatus has been attempted. For example, Kamp has attempted a treatment of the "small galaxy" cases using Montague semantics, and there have been occasional attempts to account for the "good thief" cases, and a couple of the others. But the vast number have not even been seriously studied within traditional approaches, and there is no reason whatever to think that they could be ultimately accounted for by traditional set theory , or any simple extension of it.

Let us turn now from the adequacy of the traditional set-theoretical core of the O&S hybrid theory to the identification procedures. O&S do not give an indication as to what such identification procedures might be like. But what is more important is that O&S do not consider the question of what the identification procedures for complex concepts would be like and how they would be related to the identification procedures for component concepts. Take, for example, O&S's case of *pet fish*. As O&S correctly observe, "A guppie is more prototypical of *pet*

fish than it is of either *pet* or *fish*." In the hybrid theory, the identification procedure for *pet* would not pick out a guppie as prototypical, nor would the identification procedure for *fish*. How does the hybrid theory come up with an identification procedure for the complex concept *pet fish* that will pick out a guppie as prototypical? In short, the hybrid theory has not solved the problem of how to account for the prototypes of complex concepts. It has just given the problem a new name.

Even the O&S hybrid theory would have to have as a subpart a theory for showing how prototype phenomena for complex categories are related to prototype phenomena for their component categories. But such a theory would be, in itself, a theory of complex categorization. Not only is a set-theoretical "core" inadequate as a theory of complex categorization; it is also extraneous, since a prototype-based theory of complex cateogorization is required independently.

THEORIES OF THE FORM: CORE + EVERYTHING ELSE

As an attempt to retain the classical theory of categorization, the hybrid theory is a dismal failure. But the fact that it has failed is important. The reason is that the hybrid theory is no ordinary run-of-the-mill theory of the sort that pops up and disappears regularly. I believe it represents a last gasp at saving a whole range of traditional theories in a number of areas of cognitive science in general, and in linguistics in particular. The issue of categorization pervades all of cognitive science, whether one is concerned with perception, imagery, reasoning, memory, syntax, or phonology. All of cognition uses categories – the question is what kind of theory of categorization is empirically adequate. If traditional theories cannot be defended, then a whole new theoretical apparatus has to be developed for all areas of cognition.

The O&S hybrid theory is not an isolated attempt to preverve tradtional approaches. The strategy behind the theory has been used many times before.

The general strategy is this: Bifurcate the subject matter of the theory into two parts A & B, with the following properties:

1. A is independent of B.

2. B depends on A.

3. There should be simple, general principles governing A. B is messy.

4. A is considered "really important" and of immediate concern. B is considered "less important" and its concerns can wait.

5. Since A is independent of B, no results about B can affect A. Therefore, if you are intersted in A, you can safely ignore B. But if you are interested in B, you must be concerned with A.

6. Traditional methods will work for A, while new methods must be worked out for B. These traditional methods are most typically taken from developments in the foundations of logic and mathematics, with the implicit assumption that tools from this area can be applied wholesale to the study of cognition.

7. Business as usual can go on in the study of A.

8. Students can be trained more easily in the study of A.

9. More research effort will be directed to the study of A than to the study of B.

10. This is appropriate since the methods for studying A are well-known, since A is "really important", and since B depends on A, but not vice versa.

This seems to be the hidden agenda behind the hybrid theory. The term "core" connotes centrality -- what is really important. The study of concepts and thoughts are central.

Here are some of the areas where this strategy has been suggested in the

recent past:

A	B
Propositional reasoning	Perception, imagery, metaphor, etc.
Formal semantics	Pragmatics
Linguistic competence	Linguistic performance
Autonomous Syntax	Empirical semantics
Core grammar	Most of grammar
Abstract phonology	Cognitively real phonology

What is particularly interesting is that those who proposed such bifurcations were attempting to preserve some aspects of the traditional theory of categorization. Thus, the bifurcation suggested by the hybrid theory: A (core) vs. B (identification procedure) can be thought of as the general case, the one on which all these suggested bifurcations is based. I would like to suggest that the failure of the hybrid theory indicates that all such bifurcations will ultimately fail as well. Just as prototye phenomena are the "real stuff" of categorization, so, I believe, those things in the B column above will turn out to be the real stuff of cognition.

CORE GRAMMAR + MOST OF GRAMMAR

As we saw above, the O&S hybrid theory consisting of a core + identification procedure for each concept failed for two major reasons. First the traditional theory would not work as a theory of core meaning, even if one accepted the separation. Secondly, there still needed to be an account of prototypes for complex categories, even if they were to be relegated to identification procedures. But given such a theory, there would be no need to separate out core and identification procedures; such a theory would supercede a compositional theory of core meaning.

A similar argument can be made in the area of syntax. There Chomsky, in recent work,(e.g. Chomsky 1981) has separated out "core grammar" from the rest of grammar. Core grammar is supposed to be that portion of grammar that happens to work by general principles of compositionality. Noncore grammar does not work by such general compositional principles and is left unstudied in generative grammar, just as prototype phenomena would be left unstudied by a return to a traditional theory of concepts by limiting it to "the core". But Chomsky's division of grammar into core + noncore, does not eliminate the need for a theory of noncore grammar. I would like to suggest that such a theory of noncore grammar would eliminate the need for a theory of core grammar.

Those who are familiar with the enormous amount of syntactic literature in the late 60's and early 70's will be acutely aware that the range of constructions included in core grammar is tiny by comparison with the range of constructions relegated to noncore grammar. Core grammar includes a very small range of phenomena: passives, subject-to-subject raising, WH-movement, each-movement, and perhaps a few more, depending on whose version one accepts. Core grammar thus excludes most of English syntax. This can be seen by a perusal of the collected works of Paul Postal, John Robert Ross, Dwight Bolinger, and myself, together with all the proceedings of the Chicago Linguistic Society, and such classic compendiums as those by Jespersen and by Quirk et al. As Dwight Bolinger and Charles Fillmore have repeatedly observed in recent years, most constructions in English are not fully compositionally productive, but are productive to a certain degree. There appears to be a continuum between productive constructions and completely frozen constructions. Presumably a theory that could account for compositionality in the range from almost completely frozen to almost completely productive should be able to account for fully productive constructions as well. Since this partially productive range of

constructions includes, at my best estimate, 95% to 98% of the constructions in English, the attempt to isolate a handful of fully productive constructions and view them as the really significant constructions seems as pointless as the O&S hybrid theory.

ODD NUMBERS

It should be pointed out that Rosch's results apply to natural categories, that is, those that arise spontaneously and are used by ordinary people, as opposed to artificial categories that are constructed for some technical purpose. It is common for man-made systems of thought to to be constructed so that they fit the classical theory of categorization. This is typically the case in expert theories of natural phenomena. Such man-made systems are often set up so that they by definition have yes-no answers, clear boundaries, primitive concepts, necessary and sufficient conditions, etc. It should be no surprise when artificial systems that were constructed to fit the traditional theory of categories do in fact do so. Classical systems of mathematical logic are examples of such artificial systems which, by and large, have the properties that they were constructed to have.

Confusions sometimes arise when expert theories are developed to fit natural concepts. Arithmetic is, for example, an expert theory developed to fit the natural concept "number". Arithmetic began as the study of the so-called "natural numbers": 1,2,3,4,5,6,... In the the history of mathematics, there have been heated controversies over whether various mathematically defined entities should be admitted into the category "number": zero, the negative numbers, the fractions, the reals, the imaginary numbers, infinitesimals, transfinite cardinals, inaccessible ordinals, etc. Some mathematicians accept all of these as numbers; others do not. It depends on the assumptions they make. Numbers do not have an objective existence outside of conceptual systems. There is no single,

objectively correct set of necessary and sufficient conditions defining "number". What your concept of number is depends partly on your training in mathematics, and perhaps on your philosophical beliefs. For people who are not highly trained in mathematics, some kinds of numbers will be more representative of the concept number than others. I have not done any experiments, but my guess is that relatively low positive integers are the most representative examples of the category "number". I would be very surprised if experiments did not show that six is more representative than pi, 2i+1, and aleph-null, or even very large positive integers like a googolplex.

There is a very good reason for this. Our everyday concept of number -- that is, our understanding of what a number is -- depends on our experience with numbers, and that depends on training, degree of expertise, etc. The activity of counting gives one experience with low positive integers. Experiences like dividing up a batch of cookies gives one experience with division (and remainders). Experiences like dividing up a batch of cookies between two people gives rise to concepts like "even number" and "odd number". Whether one is engaged in dealing cards or choosing up teams, from early childhood on, evenness is associated with balance, proportionality, and fairness, and oddness with imbalance, lack of proportionality, and unfairness. This invests concepts like 'even number and "odd number" with considerable experiential importance.

The fact that our orthography and our naming system for numbers is in base 10 is important for comprehending such concepts, since it allows us to tell at a glance when an integer is odd or even -- just look at the last digit. If we counted in base 9 instead of base 10, that would not be true. In base 9, 12 would be an odd number, 15 would be even, 22 would be even, 25 would be odd, 100 would be odd, and offhand, without calculation, I don't know whether 1745638 would be odd or even. Similarly, our cognitive reference points depend on

orthography. We have as cognitive reference points numbers like 10, 100, 1000, a million, a billion, etc. The numbers corresponding to these orthographic representations would probably also serve as cognitive reference points if we used a base 9 or base 7 orthography, but they would designate very different numbers! -- at least in the sense of "number" as it is defined in formal arithmetic. One of the very confusing things for young schoolchildren learning the "new math" is distinguishing between a number and its orthographic representation. The reason is that we make important use of orthographic representations in comprehending what numbers are, and we have to unlearn much of this in order to learn formal arithmetic as it exists free of the constraints of an orthographic system. Although the concept of "number" is independent of orthographic systems in formal mathematics, the properties of orthographic systems are very much part of formal mathematics. Our orthographic system for numbers is based on a representation of integers in terms of decompositions into sums of multiples of powers of ten. This particular decomposition is part of what we understand integers to be. Much of our experience with numbers has to do with computation, and computation is based on particular decompositional representations. Other systems of computation are based on other decompositions. Binary arithmetic is an important subject matter, and the advent of certain types of computers has made octal arithmetic important.

The point is this: Our base 10 orthographic system gives a special cognitive status to the single-digit numbers, since we determine oddness and evenness for large numbers in terms of oddness and evenness for single- digit numbers. Thus it should not be surprising to find that single-digit odd numbers are more representative of odd numbers than multiple-digit odd numbers. This result has in fact been reported by Gleitman (1981).

Gleitman did not interpret these results in the same way I do. She inter-

preted them as showing that prototype phenomena ~~has~~ have nothing whatever to do with what she calls "the real concepts." Here is where I disagree with her interpretation:

The theory of natural categorization has to do with the way human beings conceptualize things, and the way they understand them in terms of their experience. Prototype theory claims that representative members have a special status in terms of conceptualization and understanding. If people understand multiple-digit odd numbers in terms of single-digit odd numbers, then we would expect single-digit odd numbers to have special status of the sort predicted by the theory of natural categorization.

Gleitman, however, did not distinguish natural categories, which have to do with human conceptualization and understanding, from artificial categories, which are constructed to fit traditional theories of categorization. She assumed that the only 'true' category of "odd number" was the one defined within the man-made system of formal arithmetic. This had, of course, been defined to fit the traditional theory of categorization, rather than the theory of natural human categorization. In formal arithmetic, the way human beings happen to understand a concept is irrelevant. In formal arithmetic, the fact that people comprehend some odd numbers in terms of others is irrelevant. By definition, given the traditional theory, all that matters in characterizing a concept are its necessary and sufficient conditions. In this case, there is only one simple condition: that it be of the form $2n+1$, where n is an integer. Within formal arithmetic, anything else is extraneous to the category of odd number, where "category" is defined in the traditional theory.

Gleitman took the artificial concept of odd number as defined in formal arithmetic as the only true concept of odd number. She then performed Rosch-style experiments, which showed that people understood "odd number" in terms

of prototypical odd numbers. She concluded correctly that such prototype effects were irrelevant to the concept of odd number as it is defined in formal arithmetic. Her mistake was in assuming that this artificial concept was the only true concept, and therefore that it was the humanly relevant one, the concept in terms of which people comprehend what an odd number is.

I feel that Gleitman missed the point of the theory of natural categorization, which is to show the way that human beings conceptualize and comprehend things. Her experiments showed that in the case of "odd number", as in the case of other natural categories, they use prototpes.

THE LINGUISTIC STUDY OF NATURAL CATEGORIZATION

Within cognitive psychology, most of the study of categorization has involved physical objects. Within cognitive linguistics, on the other hand, there has been relatively little focus on physical objects. Most work has concerned categorization of "psychological" entities -- colors, events, actions, perceived spatial relations, causation, social institutions, syntactic entities (nouns, verbs, subjects, grammatical constructions), phonological entities, mental images, etc. This research has produced overwhelming support for prototype theory, or more properly for the need to develop further a theory of natural categorization along the lines of Rosch's results.

In what follows I will be discussing some of the research in cognitive linguistics that bears on issues in the theory of categorization. In the process I will be discussing some of the ideas that have developed in cognitive linguistics and how they are related to the theory of natural categorization.

The need to distinguish representative from nonrepresentative members of categories arises as soon as one considers issues of the appropriateness of descriptions.For example, if there is a sparrow on your front lawn, it would be

appropriate to describe the situation by "There is a bird on my front lawn". On the other hand, if you saw that there was a chicken or a kiwi or an ostrich or a penguin on your front lawn, such a description would be inappropriate and misleading.

English has expressions that function semantically to form categories consisting of representative members. Consider for example the expression "par excellence". It is true that a robin is a bird par excellence, and false that chickens, ostriches and penguins are birds par excellence. On the other hand, expressions like "sort of" and "kind of" function semantically to define a derived category consisting of nonrepresentative members. Thus, it is false that a robin is sort of a bird, but one might describe an ostrich or a penguin as "sort of a bird".

The point is this: The expressions "par excellence" and "sort of" are defined in terms of prototype structure. One cannot describe their meanings at all without reference to the notion of representative members. Within the O&S hybrid theory, for example, one could not give an account of the meanings of such expressions in principle, since "meaning" would be defined only in terms of concept cores and not in terms of identification procedures. Since the concept cores would be classical, they would have no prototype structure at all for "par excellence" and "sort of" to operate on.

Linguistic hedges offer other kinds of evidence against traditional theories of categorization and in favor of theory of natural categorization. Take the issues of objectivism and uniformity. In the classical theory, properties are objectively given, inherent properties of the objects themselves, rather than humanly relevant interactional properties like perceptual properties, general motor actions, functions and human intentions. And in the classical theory, all properties are equal – different kinds of properties do not have different sta-

tuses. Linguistic hedges like "fake" cannot be accounted for adequately in traditional theories that maintain objectivism and uniformity.

Attempts have been made to account for expressions like "fake" within Montague semantics, which is an attempt to extend and build on classical theories. Various Montague semanticians have correctly observed the folowing distinction between adjectives like "fake" and adjectives like "black":

That is a black gun

entails

That is a gun.

On the other hand,

That is a fake gun.

entails

That is not a gun.

They correctly recognize that this is a problem for logic, and for a theory of concepts. The approach that has been taken is the following: Bifurcate the class of adjectives into two kinds: those like "black" and those like "fake". Assume that those like "black" work by set intersection: in any context, the category "black gun" will denote the intersection of the set of black things and the set of guns.This will guarantee that black guns will always be guns. Adjectives like "fake", on the other hand, will be functions that work by mapping the category modified into a subset of its complement: thus "fake gun" will be mapped into a subset of the set of non-guns. This will guarantee that a fake gun is not a gun.

We have already seen that colors in general do not work by set intersection, but we will let that pass for the moment. The problem with such an analysis is that it doesn't tell us much. Consider, for example, the entailments:

That is a fake gun.

Therefore, that is not an elephant.

That is a fake gun.

Therefore, that is not a picture of Fats Waller.

That is a fake gun.

Therefore, that is not a theory of categorization.

There are an indefinitely large number of such entailments, and these too are a problem for logic and for the theory of concepts. The problem with the Montague analysis is that it does not specify what a fake gun is. It only provides a minimal constraint on what it is not.

Let us think for a moment about the meaning of "fake". Fake things are manufactured; they do not occur in nature. Fake things are made to deceive certain people into thinking they are real. To do this, they generally have to be perceived as real; that is, they must share enough of the perceptual properties as is necessary in context. They also have to not function in the way the real thing does: if it shoots, it's a real gun, not a fake. Moreover, it has to not have been manufactured to function like the real article: if a real gun breaks, or malfunctions, or rusts, that doesn't make it fake. And fake items are made to deceive for a purpose, generally a purpose that the real item would serve. For example, a fake gun might be used in a robbery or a prison break, serving the purpose of a real gun.

Relative to a given context, a fake item must preserve enough of the perceptual-motor attributes to deceive, it must not function, or have been made to function, like the real thing, and it must have some purpose beyond its basic function that the real article has. These are not inherent properties of objects;

rather they are humanly-relevant interactional properties, and an adequate account of "fake" must distinguish among types of interactional properties. Thus, the way that "fake" works semantically is inconsistent with theories of categorization that require objectivism and uniformity. It thus provides confirmation of the theory of natural categorization and disconfirmation of classical theories.

Research on color categorization provides further evidence confirming the theory of natural categorization. Kay and McDaniel (1978) report that focal colors function as representative members of color categories, with nonfocal colors functioning as less representative members. They also report on biophysical research on color perception that shows that focal colors are not "in the world" objectively, but depend on rates of neural firings within the eye. The categorization of nonfocal colors depends partly on closeness to the focal colors, but not entirely; it may vary according to cultural convention. Color categorization research thus confirms the the theory of natural categorization in the following respects:

-There are representative and nonrepresentative members.

-Color is an interactional property.

-The boundaries of color categories are not clearcut, vary with context, and are subject to cultural conventions. They are thus not predictable from the prototypes. The study of causative constructions also suggests that prototype theory makes better sense of the concept of causation than does the classical theory. Lakoff (1977) argues that prototypical causation is direct manipulation, which is characterized most typically by the following cluster of interactional properties:

1. There is an agent that does something.

2. There is a patient that undergoes a change to a new state.

3. (1) and (2) constitute a single event; there is an overlap in time and space; the agent comes in contact with the patient.

4. Part of what the agent does (either the motion or the exercise of will) precedes the change in the patient.

5. The agent is the energy source; the patient is the energy goal; there is a transfer of energy from agent to patient.

6. There is a single definite agent, and a single definite patient.

7. The agent is human.

8. a.The agent wills his action. b. The agent is in control of his action. c. The agent bears primary responsibility for both his action and the change.

9 The agent uses his hands, body, or some instrument.

10. The agent is looking at the patient, the change in the patient is perceptible, and the agent perceives the change.

The most representative examples of humanly-relevant causation have all ten of these properties. This is the case in the most typical kinds of examples in the linguistics literature: Max broke the window, Brutus killed Caesar, etc. Billiard-ball causation, of the kind most discussed in the natural sciences, has properties (1) - (6). Indirect causation is not prototypical, since it fails in (3), and possibly other conditions. Many languages of the world meet the following generalization: The more direct the caustion, the closer the morphemes expressing the cause and the result. This accounts for the distinction between *kill* and *cause-to-die*. *Kill* expresses direct causation, with cause and result expressed in a single morpheme – the closest possible connection. When would anyone ever say "cause to die"? In general, when there is no direct causation, when there is causation at a distance. For a discussion of corresponding facts in Japanese and

Mixtec, see Shibatani (1976) and Hinton (1982).

This view of the concept of causation also provides an explanation for various kinds of debates over whether or not something is a cause, or how good an example of a cause it is. For example, according to this account, indirect causes are less representative examples than direct causes. Multiple causes are less representative than single causes. Involuntary causation is less representative than voluntary causation. Causation with a separation of agent and patient (say throwing a ball) is more representative than causation where the patient is a body-part of the agent (say, raising your arm).

Conditons (1) - (10) are obviously not all equally important. The relative importance of these conditions is a matter for further empirical study. As is the question of how accurate these conditions are, and what other properties there might be.

IDEALIZED COGNITIVE MODELS

Much of the evidence of a theory of natural categorization comes from the area of lexical semantics. Most of the empirical researach in this area has been done by Fillmore. Rather than try to summarize it here, I refer my readers to his contribution to the present volume. Fillmore's research is set within his theory of frame semantics. Out of that research has come the concept of what I will call an **IDEALIZED COGNITIVE MODEL**, or, for short, an ICM.

As their name suggests, ICMs provide idealized models of reality. The idealizations involve oversimplifications, and often, metaphorical understandings and theories of reality – both expert theories and what anthropologists have referred to as folk theories. Fillmore has hypothesized that the meanings of lexical items are defined relative to ICMs, and that ICMs provide the motivation for the existence of the lexical item.

Fillmore has suggested that the theory of ICMs can account for at least some prototype phenomena. I would like to discuss how such a theory would work, what constraints there would have to be on it, what phenomena are outside the theory as it now stands, and how the theory might be extended to account for those phenomena.

Let us start with Fillmore's account of the word *bachelor* (Fillmore, 1982):

> The noun *bachelor* can be defined as an unmarried adult man, but the noun clearly exists as a motivated device for categorizing people only in the context of a human society in which certain expectations about marriage and marriageable age obtain. Male participants in long-term unmarried couplings would not ordinarily be described as bachelors; a boy abandoned in the jungle and grown to maturity away from contact with human society would not be called a bachelor; John Paul II is not properly thought of as a bachelor.

In other words, *bachelor* is defined with respect to an ICM in which there is a human society with (typically monogamous) marriage, and a typical marriageable age. The idealized model says nothing about the existence of priests, "long-term unmarried couplings", homosexuality, etc. The ICM might include some folklore about people who are unmarried, e.g., unmarried women want to get married and are on the lookout for eligible unmarried men; unmarried men can't cook or keep house; etc. With respect to this idealized congnitive model, a *bachelor* is simply an unmarried adult man. This idealized model, however, does not fit the world very precisely. It is oversimplified and contains folk theories about unmarried people that are not by any means completely accurate. There are some segments of society where the idealized model fits reasonably well, and when an unmarried adult man might well be called a bachelor (e.g., Joe Namath). But the ICM does not fit the case of priests or people abandoned in the jungle. In such cases, unmarried adult males are certainly not

representative members of the category of bachelors.

The theory of ICMs would account for the prototypicality effects of the category "bachelor" in the following way: An idealized cognitive model may fit one's understanding of the world either perfectly, very well, pretty well, somewhat well, pretty badly, badly, or not at all. If the ICM in which *bachelor* is defined fits perfectly and the person referred to by the term is unequivocally an unmarried adult male, then he qualifies as a prototypical member of the category *bachelor*. The person referred to deviates from prototypical bachelorhood if either the ICM fails to fit the world perfectly or the person referred to deviates from being an unmarried adult male.

The theory of ICMs depends upon keeping one's idealized cognitive models separate from one's knowledge about and understanding of the world. Then one needs the concept of "fitting" one's ICMs to one's understanding of a given situation, and keeping track of the respects in which the fit is imperfect.

It should be borne in mind that this is an oversimplified version of the theory of ICMs. Further constraints needed on the theory will be discussed below. But for the moment, it might appear to be a reasonable compromise between the classical theory and prototype theory: Definitions are precise relative to the ICMs, but not relative to either the world or one's knowledge of it. Prototype effects arise from imperfect fits. It looks like the classical theorist can maintain his position relative to the ICMs. But as we shall see, things are not so simple.

Before going on to more complex cases, let us consider two examples that appear to be handled reasonably well by the oversimplified theory of ICMs presented so far. The first is taken from Coleman (1975), and concerns protypicality effects in presuppositions. It is well known that the verb *manage* is typically restricted to contexts in which it is taken for granted that the subject tried to accomplish the specified action. For example, if one says

Harry managed to break down the door.

it is most typically assumed that

Harry tried to break down the door.

As Coleman observes, typically one only has to try to do something if there is expected to be some difficulty. And if there is difficulty then there is typically some reason to believe that the event might not happen, that is, that the event is unlikely. Thus, Coleman observes that the following hierarchical relationship holds:

MANAGING takes for granted TRYING
TRYING involves the expectation of some DIFFICULTY
DIFFICULTY suggests UNLIKELIHOOD

Corresponding to this hierarchy, Coleman observes that there is a hierarchy of representativeness in the use of the verb *manage* . The most representative uses of *manage* have all three conditions:

Harry managed to break down the door.
The A's managed to get five runs off Gossage.
Max managed to get a date with Sylvia.

Here there is TRYING, DIFFICULTY and UNLIKELIHOOD. However, there are uses of *manage* where there is only DIFFICULTY and UNLIKELIHOOD:

Harry spent all evening trying not to insult Ursula, but he managed to insult her all the same.
It's pretty difficult to flunk this course without trying, but Fred managed to do it.

INDENT

like

Such cases seem less representative uses of *manage* than the ones given above. Still less respresentative are uses where only the UNLIKELIHOOD condition is met. In such cases it is possible to have an inanimate subject of *manage* .

It always manages to rain on my day off.

That low-budget movie managed to make thirty million dollars.

That old house has managed to remain unoccupied for years.

These are possible, but unrepresentative uses of *manage* . If *manage* is defined relative to an ICM in which someone is trying to do something, then the ICM will have the three conditions given above. How representative the use of *manage* is will depend on how many of those conditions hold in the situation given.

The most complex example of the theory of ICMs discussed to date has been given by Sweetser (ms). Sweetser's analysis is based on experimental results by Coleman and Kay (1981) on the use of the verb *lie*. Coleman and Kay found that their informants did not appear to have necessary and sufficient conditions characterizing the meaning of *lie* . Instead they found a cluster of three conditions, no one of which was necessary, and which varied in relative importance:

> A consistent pattern was found: falsity of belief is the most important element of the prototype of *lie* , intended deception the next most important element, and factual falsity is the least important. Informants fairly easily and reliably assign the word *lie* to reported speech acts in a more-or-less, rather than all-or-none, fashion,... [and]... informants agree fairly generally on the relative weights of the elements in the semantic prototype of *lie* .

Thus, there is agreement that if you, say, steal something and then say you didn't, that's a good example of a lie. A less representative example of a lie is

when you tell the hostess "That was a great party!" when you were bored stiff. Or if you say something true but irrelevant, like "I'm going to the candystore, Ma" when you're really going to the poolhall, but will be stopping by the candystore on the way.

An important anomaly did, however, turn up in the Coleman-Kay study. When informants were asked to define a *lie*, they consistently said it was a false statement, even though actual falsity turned out consistently to be the least important element by far in the cluster of conditions. Sweetser (ms) has observed that the theory of ICMs provides an elegant way out of this anomaly. Sweetser points out that, in most everyday language use, we take for granted an idealized cognitive model of social and linguistic interaction. Here is a somewhat oversimplified version of the ICM Sweetser proposes:

- People intend to help rather than harm one another.
- Giving correct information is helpful.
- People say what they believe.
- People have adequate reasons for their beliefs.
- What people have adequate reason to believe is true.

Most of the time, in ordinary everyday discourse, we function in terms of such an idealized model unless we have reason not to.

Sweetser then shows that within this ICM Coleman and Kay's three conditions for *lie* are equivalent. Saying what you believe not to be true is saying what is false. And giving incorrect information is harmful not helpful. Thus, if one defines a *lie* simply as a false statement with respect to this ICM (not the real world), the additional Coleman-Kay conditions will follow as a consequence. Sweetser also shows that their relative importance will also follow as a consequence of the logical relations among the elements of the idealized cognitive

model. And, moreover, the nonrepresentative cases will be accounted for by imperfect fits of this ICM to the situation at hand. For example, white lies and social lies occur in situations where the condition "Giving correct information is helpful" does not hold.

Sweetser then goes on to show how other expressions, such as, *social lie*, *white lie*, *exaggeration*, *joke*, *kidding*, *oversimplification*, *tall tale*, *fiction*, *fib*, *mistake*, etc. can be accounted for in terms of systematic deviations from the above ICM. Although neither Sweetser nor anyone else has attempted to give a theory of complex concepts in terms of the theory of ICMs, it is worth considering what would be involved in doing so.

As should be obvious, adjective-noun expressions like *social lie* do not work according to traditional theories. The category of social lies is not the intersection of the set of social things and the set of lies. The term *social* brings with it an idealized cognitive model in which being polite is more important than telling the truth. This conflicts with the condition "Giving correct information is helpful", and it overrides this condition. Saying "That was a great party!" when you were bored stiff is a prototypical social lie, though it is not a prototypical lie. Thus, any general account of complex concepts like *social lie* in terms of ICMs will have to indicate how the ICM evoked by *social* can cancel one condition of the ICM evoked by *lie*, while retaining the other conditions. An obvious suggestion would be that in conflicts between modifiers and heads, the modifiers win out. I don't know whether this will hold generally.

ICMS AND SCENE SEMANTICS

Each of the cases considered so far is a case where the theory of ICMs permits a simple definition while still accounting for prototype effects. On the surface, it would thus appear from such cases that one could salvage some version of the traditional theory by relativizing it to idealized cognitive models, rather

than having it hold directly with respect to the world. One way of seeing some of the differences between the traditional theory and the theory of ICMs is by comparing the theory of ICMs with scene semantics as developed by Barwise and Perry (1980). A scene for Barwise and Perry is a partial model, which contains some entities and some specification of their properties and the relationships between them. The specifications are left incomplete, just as our knowledge of the world is necessarily incomplete. A B&P scene can thus be understood as a subpart of the world as observed from a particular viewpoint. B&P's scenes thus contrast with possible worlds in intensional semantics in that possible worlds are complete specifications of all entities in a world and all their properties and interrelationships.

This incompleteness is one thing that B&P's scenes share with ICMs. But there is a very important difference between the two views. Within scene semantics, truth is defined with respect to scenes and so are classical entailments. Scene semantics is an objectivist semantics. What's there in the scene is really there in the world; it's just not all that's there. An idealized cognitive model is very different. First of all, it's idealized. It provides a conventionalized way of comprehending experience in an oversimplified manner. It may fit real experience well or it may not. ICMs are not part of an objectivist semantics, and this will become clearer as we go through more examples.

Let us begin with Barwise's attempt to provide a logic of perception (Barwise, 1980). Barwise discusses naked infinitive (NI) constructions like *Harry saw Max eat a bagel*, where *eat* is in naked infinitive form. He proposes several principles governing the logic of such sentences, including the following:

(A) **The Principle of Veridicality** : *If a sees P, then P.*

For example, if Harry saw Max eat a bagel, then Max ate a bagel.

(B) **The Principle of Substitution:** *If a sees* $F(t_1)$, *and* $t_1 = t_2$, *then a sees* $F(t_2)$.

Barwise's example is:

Russell saw G.E. Moore get shaved in Cambridge.

G.E. Moore was (already) the author of *Principia Ethica* .

Therefore, Russell saw the author of *Principia Ethica* get shaved in Cambridge.

I find both of these principles problematic if taken as objectively true and absolute – which is how they are intended. The basic problem is this: SEEING TYPICALLY INVOLVES CATEGORIZING. For example, seeing a tree involves categorizing an aspect of your visual experience as a *tree* . Such categorization in the visual realm generally depends on conventional mental images: You categorize some aspect of your visual experience as a tree because you know what a tree looks like. In the cases where such categorizations are unproblematical, we would say that you really saw a tree.

An important body of 20th century art rests on the fact that ordinarily seeing is seeing-as, that is, categorizing what is perceived. A good example is discussed in Lawrence Weschler's *SEEING IS FORGETTING THE NAME OF THE THING ONE SEES* , a biography of artist Robert Irwin (Weschler, 1982). A substantial part of Irwin's career was devoted to creating art pieces that could not be seen as something else, that were exercises in pure seeing without categorization. Irwin's discs provide the best examples I have come across of such experiences. The point is that seeing experiences of this sort – seeing without seeing-as, seeing without categorization – are rare. They require extraorinary works of art, or very special training, often in a meditative tradition. The existence of such extraordinary seeing experiences highlights what is typical of seeing: SEEING TYPICALLY INVOLVES CATEGORIZING.

Because people do not all categorize the things they see in the same way, the semantics of seeing cannot be purely objective. Suppose Harry sees a woman walk by him. He thinks she's beautiful; I think she's ugly. He says, speaking truly of his experience:

(1) I saw a beautiful woman walk by me.

And he would take a third-person report of his experience to be true:

(2) Harry saw a beautiful woman walk by him.

According to Barwise's principle of veridicality, (3) would follow:

(3) A beautiful woman walked by Harry.

Harry would agree but I wouldn't. I would take (3) to be false. My true account of the situation would be:

(4) I saw an ugly woman walk by Harry.

And by Barwise's principle of veridicality, (5) would follow:

(5) An ugly woman walked by Harry.

Given that only one woman walked by Harry, it would seem that (3) and (5) are incompatible. Yet they follow ~~from~~ the principle of veridicality from (2) and (4), which are true.

For an objectivist, seeing-as is a matter of human psychology and should never enter into questions of meaning, which are objective. But seeing-as, in the form of visual categorization, typically enters into seeing. This fact creates all sorts of problems for any logic of perception set up on objectivist principles. Among the places where the problems show up are in the principles of veridicality and substitution. Take for example the following case where Barwise's principle of substitution applies:

(a) Oedipus saw Jocasta get into bed with him.

(b) Jocasta was Oedipus' mother.

(c) Therefore, Oedipus saw his mother get into bed with him. Is (c) a valid conclusion to draw from (a) and (b)? Well, yes and no. The case is problematic. Oedipus did not see Jocasta as his mother, so you can't unproblematically say he saw his mother get into bed with him. On the other hand, he did see someone who happened to be his mother get into bed with him. The problem is very much the same as the classic problem with *want* .

(a) Oedipus wanted to marry Jocasta.

(b) Jocasta was Oedipus' mother.

(c) Therefore, Oedipus wanted to marry his mother.

Logicians generally agree that the principle of substitutability does not apply with *want* , since (c) does not unproblematically follow from (a) and (b). Yet the cases with *want* and *see* are parallel.

Though Barwise's principles of veridicality and substitution do not hold unproblematically in the way they were intended, they are not altogether wrong. They seem to follow from our common- sense folk theory of seeing, which might be represented as an idealized cognitive model of seeing.

THE ICM OF SEEING

1. You see things as they are.
2. You are aware of what you see.
3. You see what's in front of your eyes.

These aspects of our idealized cognitive model of seeing have various consequences, among them the folk-theoretical forms of Barwise's principles.

Consequences of 1:

Veridicality: If you see an event, then it really happened.

Substitution: You see what you see, regardless of how it's described.

Exportation: If you see something, then there is something real that you've seen.

Consequence of 1 and 2:

To see something is to notice it and know it.

Consequences of 3:

The Causal Theory of Perception: If something is in front of your eyes, you see it.

Anyone looking at a certain scene from the same viewpoint at a given place and time will see the same things.

You can't see what's not in front of your eyes.

You can't see everything.

This idealized cognitive model of seeing does not always accurately fit our experience of seeing. Categorization does enter into our experience of seeing, and not all of us categorize the same things in the same way. Different people, looking upon a scene, will notice different things. Our experience of seeing may depend very much on what we know about what we are looking at.

If Barwise's principles of veridicality and substitution are taken as special cases of the ICM OF SEEING, rather than as logical principles, then the problematic cases we noted above can be accounted for straightforwardly as cases where the situation is not normal and where the idealized cognitive model does not fit the situation in certain respects. For instance, the Oedipus-Jocasta example, however real, is not a representative example. This is exactly what is predicted by the ICM OF SEEING. Whenever what one sees as happening depends on knowledge that is not generally shared, cases like the Oedipus-Jocasta

example can arise. Such cases are not representative, and will be judged so. Here the theory of ICMs works where the Barwise principles as interpreted within scene semantics fails. Within scene semantics, the Barwise principles are absolute logical principles, and the problematic cases appear to be counterexamples.

The theory of ICMs cannot be properly understood as a version of traditional Fregean semantics. The use of ICMs to account for prototype phenomena would not constitute a return to the traditional theory of categorization. The reason is that ICMs do not fit the world as it is, but provide a way of understanding experience. Within the theory of ICMs, bizarre examples of the sort constructed by linguists and philosophers are just that -- nonrepresentative examples. And the fact of their bizarreness can be accounted for by the theory of ICMs. Within traditional theories, all examples are on a par: it is not the job of such theories to distinguish representative examples from nonrepresentative ones. All examples either fit or don't, and those that don't fit when they are supposed to are counterexamples.

HEDGES REVISITED

Through a reanalysis of the data on linguistic hedges, first studied in detail in Lakoff (1973), Paul Kay (1979) has made a number of remarkable observations that have important implications for any theory of meaning, especially the theory of ICMs. The observations that intrigue me the most are:

- Certain linguistic hedges are defined relative to idealized cognitive models of language itself.

-Within English, there are hedges whose ICMs are not consistent with each other.

As we shall see, both of these observations place the theory of ICMs even further outside the bounds of traditional theories. But before proceeding with examples, let us consider why Kay's observations are important.

Traditional theories of meaning, especially those within the Fregean tradition, do not tolerate contradictions in their models. Those models are, presumably, models of the world, and the world as objectivists conceive of it, cannot, by definition, contain contradictions. But the theory of ICMs is not an objectivist theory, for two reasons.

-First, the models themselves are both idealized and cognitive. They do not characterize the world as it is.

-Second, the ICMs are not matched to the world itself, but rather to *understandings of experience* .

Thus, it would not be at all surprising to find ICMs that are inconsistent with each other. ICMs often characterize what anthropologists call "folk theories" of experience, and our folk theories often contradict each other. Kay has documented a number of such cases where we have inconsistent folk theories, which are representable as ICMs. What makes this interesting is that in each case, the meaning of one or more English expressions is defined relative to one of the ICMs. In other words, this subpart of English does not have a single overall consistent semantics. This means that is it is not amenable to analysis in terms of traditional semantic theories.

As if that weren't enough, the relevant semantic domain covered by these ICMs is language itself. We have cognitive models of language that are not consistent with each other, and we have words that are defined with respect to each such cognitive model. In other words, the semantics of our language about language is not consistent. We have a variety of folk theories about language,

each with its own vocabulary, and they don't fit together to make a coherent whole.

This, in itself, should not be surprising. There are many areas of human experience in which we have conflicting modes of understanding. We have both folk and expert theories of medicine, politics, economics, etc. Each theory, whether folk or expert, involves some idealized cognitive model, with a corresponding vocabulary. A given person may hold one or more folk theories and one or more expert theories in areas like medicine or economics or physics. It is commonplace for such idealized cognitive models to be inconsistent with each other. In fact, recent studies of adults' understanding of physics shows that most of us do not have a single coherent understanding of how the physical world works. Instead, we have a number of cognitive models that are inconsistent (cf. Gentner, 1981). For an objectivist theory of meaning, this would be an irrelevant matter of psychology. Traditional theories of concepts, which are objectivist, would take people's idealized cognitive models as irrelevant to meaning. In such theories, meaning is a matter of reference and truth. The only question that matters is whether the words actually fit the world, which is consistent by definition. Any cognitive models of the world that someone might happen to have is irrelevant in traditional objectivist theories. What makes Kay's work of interest is firstly the claim that meanings of words has to do with cognitive models, not with objectively given reality. Kay's idea makes sense of the fact that there is no overall consistent semantics of a traditional sort, at least for English and probably for other natural languages as well.

But what makes Kay's observations especially interesting is that the domain of these cognitive models is language itself. Kay's work seems to imply that, at least in some respects, the way language actually works depends on what we understand it to be. I read him as implying that language actually works – in

part – in terms of conflicting cognitive models of language, and that, at least in some areas, our cognitive models of language create linguistic reality.

Kay goes about arguing for this position through a careful reanalysis of the hedges *loosely speaking*, *strictly speaking*, *regular*, and *technically* , and the ICMs with respect to which they are defined. Kay's strategy is to show the following:

> *Loosely speaking* and *strictly speaking* are defined with respect to an ICM which has a strict semantics/pragmatics distinction of a traditional sort.
>
> *Technically* is defined with respect to ICMs that do not have a strict semantics/pragmatics distinction of a traditional sort.

Kay takes a principle like (I) to be a folk-theoretical counterpart of traditional semantic theories:

(I) WORDS CAN FIT THE WORLD BY VIRTUE OF THEIR INHERENT MEANING.

He argues convincingly that *loosely speaking* and *strictly speaking* are defined with respect to an ICM embodying (I). Kay puts it this way:

> One of the implicit cognitive schemata by which we structure, remamber, and image acts of speaking assumes that there is a world independent of our talk and that our linguistic expressions can be more or less faithful to the non-linguistic facts they represent. Thus we can lie, innocently misrepresent, speak loosely, speak strictly, and so on.

In short, if words can fit the world, they can fit it either strictly or loosely, and the hedges *strictly speaking* and *loosely speaking* indicate how narrowly or broadly one should construe the fit. For instance, take Kay's example:

> Loosely speaking, the first human beings lived in Kenya.

In a strict sense, there were no such things as "the first human beings" – at least assuming continuous evolution. But loosely speaking, this expression can

be taken to refer to primates with important human characteristics. And Kenya, if you want to be picky, didn't exist then. But loosely speaking, we can take "in Kenya" to be in the general part of Africa where Kenya now exists.

Kay thus identifies "loosely speaking" and "strictly speaking" as pragmatic hedges, which take for granted the ICM in (I) above. That is, they assume words can fit the world by virtue of their inherent meanings, either strictly or loosely. With respect to this sentence "the first human beings" and "in Kenya" properly fit things they wouldn't fit under a strict construal, given the inherent meaning of the word.

But "technically" is defined relative to a different folk theory of how words refer. "Technically" assumes the following folk conception of the relation between words and the world:

(II) THERE IS SOME BODY OF PEOPLE IN SOCIETY WHO HAVE THE RIGHT TO STIPULATE WHAT WORDS SHOULD DESIGNATE, RELATIVE TO SOME DOMAIN OF EXPERTISE.

Sometimes these people are taken to be experts who know better than the common man what the world is like, as in the example:

Technically, a dolphin is a mammal.

In other cases, technically may refer in context to some immediately relevant body of experts:

Technically, a tv set is a piece of furniture.

This may be true with respect to the moving industry and false with respect to the insurance industry. Kay suggests that *technically* should be glossed as something like "as stipulated by those in whom society has designated the right to so stipulate". Let us call such people "experts". Now when relevant area of expertise of these experts happens to be the nature of the world, then their stipulation as to how a term should be used dovetails with our assumptions about how

the world really is. In this case, technically has a 'semantic' effect, and it produces truth conditions that converge with those of *strictly speaking*.

Technically, a dolphin is a mammal.

Strictly speaking, a dolphin is a mammal.

Both sentences have the same truth conditions, but for different reasons. Since "mammal" is a term from scientific biology, the relevant body of expertise for *technically* is biology, which is about how the world is. *Strictly speaking* assumes that words, via inherent meanings, fit the world as it is. Of course, the sentences have very different linguistic meanings and conditions of use, since the two hedges evoke different ICMs.

Now when the relevant body of experts are people like Quaker church officials and Internal Revenue agents, the truth conditions for *technically* and *strictly speaking* diverge, as we would expect.

Technically, Richard Nixon is a Quaker.

Strictly speaking, Richard Nixon is a Quaker.

Technically, Ronald Reagan is a cattle rancher.

Strictly speaking, Ronald Reagan is a cattle rancher.

In both cases, the first sentence would normally be taken as true and the second as false.

The folk-theoretical principles given in ICMs (I) and (II) will happen to produce the same truth conditions for technically and strictly speaking if the domain of expertise in (II) happens to be the nature of the world.But in general they produce different truth conditions. Since (I) and (II) make very different assumptions about the relationship between words and how they come to designate things, they are not consistent with each other. Yet there are English expressions that make use of each of them. Both are needed if one is to charac-

terize the meaning and actual use of the hedges given above. In communicating, we make use of both principles. Despite their inconsistency, (I) and (II) each play a role in characterizing the reality of language use.

Language, as a domain of human activity, is affected in certain ways by the ways in which people understand the nature of the domain. Those understandings can create a reality, even when they are in part incompatible with each other. This is very much in keeping with the results reported on in Lakoff and Johnson (1980) in the area of metaphor. Among those results were:

-Metaphors are conceptual, rather than linguistic, in nature.

-Such conceptual metaphors are ways of understanding one domain of experience in terms of another.

-Each conceptual metaphor sets up an idealized cognitive model of a domain.

-A domain of experience may have many metaphorical cognitive models, each with a different ontology.

-These metaphorical models of a domain are typically not consistent with each other.

-Concepts can be defined by clusters of metaphorical cognitive models which are mutually inconsistent.

-Each metaphorical cognitive model characterizes a different aspect of the concept.

-Human concepts are, in general, multi-faceted; that is, they have more than one aspect.

-Multi-faceted concepts allow one to understand a domain of experience in more than one way and thus permit understanding of many aspects a domain of experience.

-In domains of experience that are social, interpersonal and personal, where the way people act depends in a significant way on what they understand the domain to be like, cognitive models have a role in [illegible] determining actions and therefore in creating reality.

Thus, the view of concepts and categories that comes out of metaphor research fits very well with the view of categorization that comes out of the theory of ICMs. And even though theory of ICMs permits definitions to be precise, it is by no means a version of the traditional theory of categorization.

In fact the theory of ICMs diverges from traditional theories in still another way, namely, in its characterization of pragmatics. In the theory of ICMs, pragmatics is simply the semantics of language itself. Pragmatic principles are embodied in idealized cognitive models of language that we make use of in actual communication. Many linguistic expressions, such as hedges and performative verbs, evoke such ICMs, as do syntactic constructions like imperative, question, and declarative constructions. This permits the old performative hypothesis of generative semantics to be recast in cognitive terms, with idealized cognitive models of speech acts replacing the old abstract performative predicates. Thus, the same ICM would be evoked by the imperative construction and verbs like *order* and *request*. Syntactic generalizations that depend upon the nature of speech acts can thus be stated in terms of ICMs and their relationship to surface syntactic structure.

Kay's ideas can also make sense of some of Gleitman's findings on odd numbers that she takes to be bizzare:

> We found that people say 100% of the time that the concept *odd number* is all or none and it is absurd to think otherwise. Even more surprisingly, they said over 60% of the time that such concepts as vegetable were all or none. ... Not five minutes after saying that it was absurd to suppose the

concept odd number was graded, they rated odd numbers and vergetables and gave graded responses.

Such results are not at all bizzarre in the light of Kay's findings. People seem to use more than one idealized cognitive model to comprehend a given domain, and sometimes these models are inconsistent. Moreover, such models may have very different statuses. Some may be part of consciously learned expert theories like formal arithmetic, while others may have to do with everyday comprehension. If you ask people if *odd number* is an all-or-none concept, it appears that they will consciously employ the model of arithmetic they learned in school and reply that it's an all-or-none concept. But if you perform Rosch-style experiments, ~~they employ~~ they employ the largely unconscious model they use to comprehend odd numbers, in which there are prototypes.

Kay's research also throws some light on the notion of an inherent property. In my 1973 paper on hedges, I observed that the hedge *regular* seemed to distinguish inherent from noninherent properties. In examples like

Esther Williams is a regular fish.

and

Sam is a regular Henry Kissinger.

regular seems to cancel out inherent properties and focus on salient incidental properties. In this way, *regular* is not at all like *loosely speaking*, as the following examples show:

Loosely speaking, Esther Williams is a fish.

Loosely speaking, Sam is Henry Kissinger.

Kay suggests that the idea that things have inherent and noninherent properties is part of our idealized cognitive model of what things are like — or at least one of our natural ICMs. We may have other ICMs that don't make such distinctions.

Anyway, Kay proposes that *regular* evokes and makes uses of the INHERENT PROPERTY ICM. The idea is that INHERENT PROPERTY is a humanly-relevant concept which is part of one common idealized cognitive model of reality, and one that we make regular use of.

Let us conclude this section with a further observation of Kay's: Expert theories of language are often attempts to take folk-theoretical principles about how language works and extend them consistently as if they were true in general. Since the folk-theoretical principles about how language works are not in general consistent with each other, it is of the utmost importance for linguistic theorists to be aware of what they are and where the inconsistencies lie.

FURTHER LINGUISTIC DATA

In the remainder of this paper we will be surveying additional linguistic studies that have relevance for a theory of categorization. Rather than merely describe the studies, we will ask in each case what consequences they would have for using the theory of ICMs to account for categorization generally.

BASIC COLOR TERMS

So far, the theory of ICMs has used the following strategy:

-Let the idealized cognitive model characterize the properties of the prototype in propositional terms.

-Category members will be those that fit the ICM more or less well.

The phenomena of color categorization force us to make some changes. In the case of colors, the ICMs would not be given in propositional terms, but in terms of images of focal colors. Languages that have blue and red basic color terms, but no basic term for purple, seem to work in one of three ways:

(a) Purples are assigned to blue.

(b) Purples are assigned to red.

(c) Purples are split up between blue and red.

Therefore, conventions will be needed that will specify, for things that are not "close" to the prototype, where the general areas of the boundaries are. This is in accord with Rosch's observation that category membership cannot be predicted from the prototype in cases of nonrepresentative members, though degree of representativeness will still correspond to closeness to the prototype, given conventions about boundary areas. The theory of ICMs will have to include such conventions about boundary areas for categories.

Certain hedges, like *very* and *sort of* , will take both prototypes and conventional boundary areas into account. "Very red" has the same prototype as "red" and narrower boundary areas. "Sort of red" widens the boundary area of "red", excludes the prototype of "red" and takes as prototypical areas of intermediate representativeness in the category "red". Within the theory of ICMs, the specification of conventional boundary areas is important in characterizing new categories generated by hedges of this sort.

Since languages with no blue-green distinction among basic color terms have focal blue and focal green both as most representative members of the 'grue' category, the theory of ICMs will have to admit disjunctive prototypes. This is similar to the situation in superordinate cetegories (e.g.,furniture) which require disjunctive specification of their most representative members (e.g., chair, table, bed).

LEXICAL ITEMS AS CATEGORIES OF SENSES

Fillmore (1982) observes that the adjective *long* has two senses, one spatial and one temporal. The spatial sense is generally taken to be more representa-

tive, or prototypical, and the temporal sense related to it via metaphor, in the sense of Lakoff and Johnson (1980). Another example would be the word *up*, which can mean happy in "I'm feeling up today", or can have a spatial sense in "The rocket went up". The spatial sense is generally taken as the more representative sense.

These and other observations about prototypical uses of lexical items can be united with other data on natural categorization by viewing lexical items as constituting natural categories of senses. Thus some senses of a word may be more representative than other senses. The senses of a word are related to one another more or less closely. There are various ways in which words can be related to one another. One is by conceptual metaphor. As Lakoff and Johnson (1980) observe, a metaphor can be viewed as an experientially-based mapping from an ICM in one domain to an ICM in another domain. This mapping defines a relationship between the idealized cognitive models of the two domains. It is very common for a word that designates an element of the source domain's ICM to also designate the corresponding element in the ICM of the target domain. The metaphorical mapping that relates the ICMs defines the relationship between the senses of the word. It is most common for the sense of the word in the source domain to be viewed as more representative. Thus, in the case of *up*, the source domain is spatial and the target domain is emotional, and the spatial sense is viewed as being more representative.

In other cases, a single idealized cognitive model can be the basis on which a collection of senses forms a single natural category expressed by a single lexical item. Quinn (1982) has shown that the word *commitment* is one of the most important words in the lexicography of marriage. In a detailed study of hundreds of examples from interviews, she demonstrates that *commitment* has three senses: PROMISE (as in "make a commitment"), DEDICATION (as in "feel a

commitment") and ATTACHMENT (as in "our commitment has grown stronger"). Quinn poses the question of why the same word should have these three senses in discussions of marriage. She hypothesizes the following answer:

There is an idealized cognitive model of marriage in America that follows the scenario:

MARRIAGE STARTS WITH A PROMISE

FROM THE PROMISE, A DEDICATION DEVELOPS

THROUGH DEDICATION, AN ATTACHMENT DEVELOPS

The senses PROMISE, DEDICATION, AND ATTACHMENT are thus linked though an idealized cognitive model of marriage. The ICM thus provides a basis for the relationship among the senses of *commitment*, and makes it possible for them to form a natural category of senses.

The idea that lexical items are natural categories of senses has been studied the most in the domain of English prepositions, and we will turn to those results next.

PREPOSITIONS

Most of the research on categorization with cognitive psychology has been in the domain of physical objects and physical perception. But perhaps the strongest evidence against traditional views of categorization and for a prototype approach comes from the prepositions, which specify relations, both spatial and abstract. The most detailed studies of prepositions by far are those done by Lindner (1981) and Brugman (1981). Lindner's study looked at 1800 verb-particle constructions based on the two prepositions *up* and *out*, and surveyed the contributions to meaning made by the particles. Brugman's study is an extended survey of the single most complicated preposition in English – *over* – and covers nearly one hundred kinds of uses. The two studies reach

substantially the same conclusions, though Brugman's has a more explicit discussion of the consequences for the theory of categorization, and Brugman is the first to explicitly propose the idea that lexical items are natural categories of senses.

Let us begin with a survey of the Lindner-Brugman results:

-The senses of *up*, *out*, and *over* are either spatial, or metaphorically-based on spatial senses.

-Each spatial sense can be represented by an image-schema.

-For each preposition, there are representative senses and nonrepresentative senses.

-The senses are related to one another by either minimal spatial transformations (in the case of two spatial senses represented by image-schemas), or by conceptual metaphors, each linking a spatially-grounded image-schema to an abstract domain.

-Each preposition forms a natural category of senses.

-In addition, each spatial image-schema constitutes a prototype relative to situations in the world it can fit.

Thus there are two levels of prototype structure for each preposition:

At the lexical level there are representative and nonrepresentative senses. The lexical item is the category; the senses are its members.

At the sense level, there are also representative and nonrepresentative members. Each sense, or image-schema, is a category; the situations it fits are its members. The situations the image-schema fits well are representative; the situations it fits less well are nonrepresentative.

In addition:

-The spatial senses are not always discrete, but often flow into one another, in the course of the continuous spatial transformations that relate them.

-As expected, the range of senses of each preposition cannot be predicted entirely from its most representative sense or senses; as with other categories, conventions specifying the range are necessary. Closeness among image-schemas is defined by simple spatial relationships, which *motivate* extensions of the category to nonrepresentative members, but do not *predict* them.

-Nonrepresentative members may differ from prototypical members in such different ways that they may bear little or no resemblance to each other, and are members of the same category only by virtue of their relationship to the prototypical members.

-For spatial senses, similarity is characterized not by shared properties, but via a network of minimal image transformations.

To get some sense of the data and analyses that the conclusions are based on, let us take some examples of sort discussed by Brugman.

The plane flew over the hill.

John walked over the hill.

The helicopter is over the hill.

The town is over the hill.

He spread the tablecloth over the table.

The flies are all over the wall.

He knocked the lamp over.

He turned the book over.

They talked over the plan.

They talked over lunch.

The play is over.

Do it over.

Don't overdo it.

When the image-schemas for these examples and many more were plotted relative to each other, the schemas for the first two examples turned out to be in the center of a network. An oversimplified version of the network is given in figure 1. The variation among the schemas was resolved into variation among seven factors, such as the size and shape of the trajector, the horizontal-vertical orientation of the landmark, the boundaries of the landmark, etc. When the factors were considered one-at-a-time, each schema in the network was relatively close to the central members in the network *with respect to that factor*. Given seven spatial factors that vary, plus various metaphorical mappings into abstract domains, the result is a rich and diverse category.

[INSERT FIGURE 1 AS CLOSE TO HERE AS POSSIBLE]

So far, I have not been able to find any preposition, postposition, or case in a non-Indo-European language that comes close to matching the same range of senses as *over* in English. As natural as it may seem to speakers of English to use the word *over* for all these senses, it is quite unnatural for speakers of non-Indo-European languages, and they seem have great difficulty mastering it. If Brugman's analysis is correct, it would seem that in the word *over*, English has a conceptual category that does not exist in most of the world's languages. Learning the English word *over* like a native speaker would involve learning a new category of senses, not just a list of unrelated and separate meanings. Thus, learning a language is learning a new way of categorizing, not just learning a new list of labels for old categories.

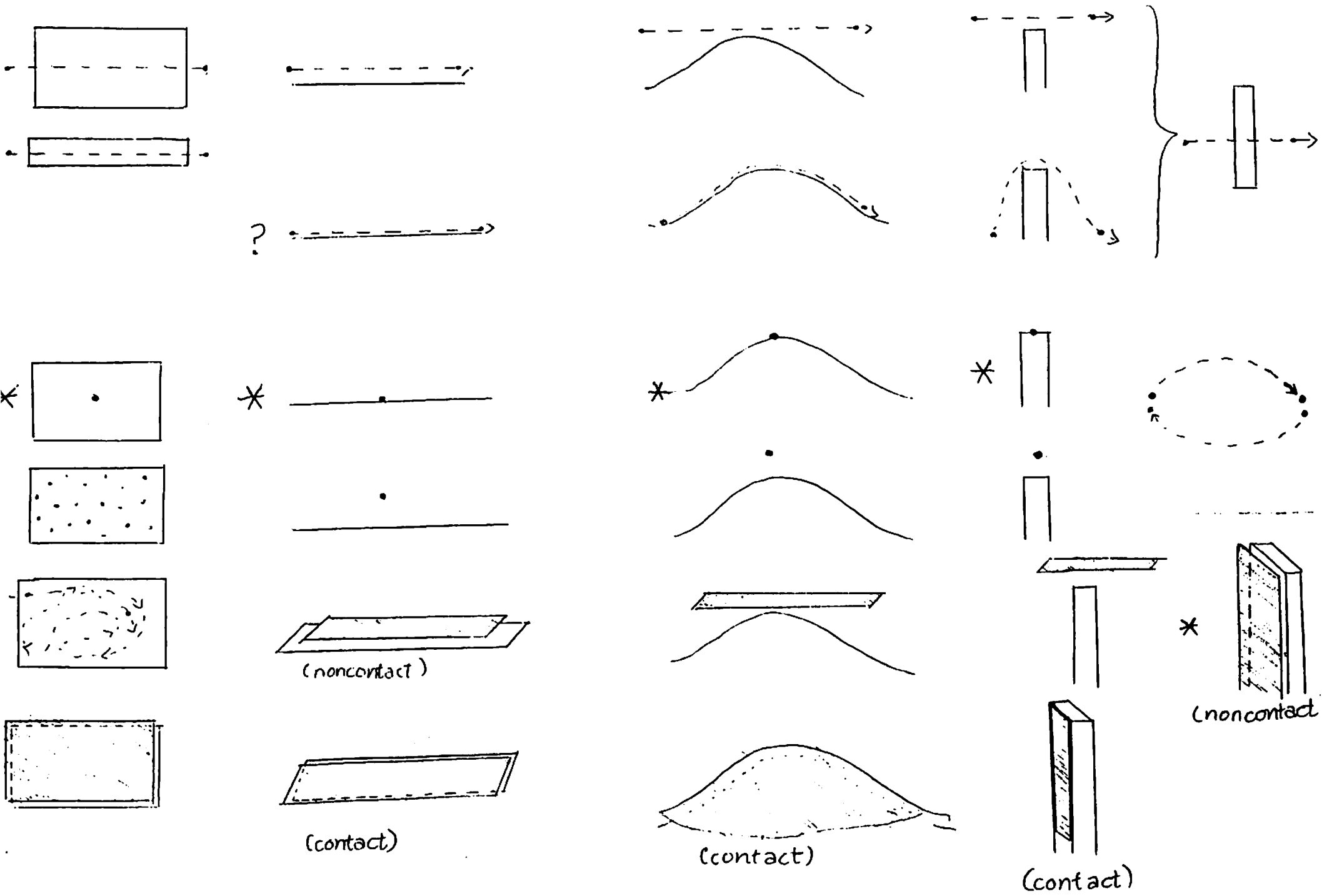

FIGURE I

CLASSIFIER SYSTEMS

A large range of the world's languages have rich classifier systems, especially Asian, African, Native American, and Oceanic languages. Among the world's languages, such systems are probably more the norm than the exception. As linguistic elements, classifiers are rather diverse. They can occur as affixes to numerals (Burmese), demonstratives (Chinese), incorporated elements in the middle of verbs (Algonquian), noun prefixes (Bantu), and as semantic components in classificatory verb stems (Athapaskan) and classificatory locative roots (Eskimo). The number of classifiers can vary from less than ten to the 557 discovered by Berlin in Tzeltal (Berlin, 1968). In Southeast Asian languages, it is common for there to be more than one hundred.

Noun classifiers vary from being completely frozen to being relatively free. For example, the Swahili noun *chura* (frog) consists of a classifier in the form of a noun prefix and a noun root, *ch-ura* (artifact-frog). The classifier *ch-* indicates lowliness of status -- a mere thing, as opposed to, e.g., *ng-ombe* (animal-cow); cows have higher status as animals. These are frozen forms, with the classifier attached directly to the noun and the classifier indicating what is, relative to a folk theory of the culture, an inherent property of the object. In Burmese, on the other hand, classifiers can vary depending on what one is talking about, as Becker (1975) exemplifies:

myi? tə tan river one line (e.g., on a map)
myih tə 'sin river one arc (e.g., a path to the sea)
myi? tə thwe river one connection (e.g., linking two villages)
myi? tə khu' river one thing (e.g., in a discussion of rivers in general)

As Denny (1976) observes, "...the semantic function of noun classifiers is to place objects within a set of classes different from and additional to those given by the

nouns. These classes are concerned with objects as they enter into human interactions..." Denny notes that, cross-linguistically, classifiers fall into three basic semantic types, all having to do with human interaction: "*physical interaction* such as handling, *functional interaction* such as using an object as a vehicle, and *social interaction* such as interacting appropriately with a human compared to an animal, or a high status person compared to a low status one." Denny argues persuasively that the range of physical interaction classifiers correlates with the kinds of significant physical activities performed in the given culture. Translated into the terminology adopted above, classifiers are defined relative to the idealized cognitive models prevalent in the culture, and they specify interactional properties, rather than purely objective inherent properties, even though such properties may be considered inherent by the folk theories of the culture. This is in accord with both the theory of ICMs and Rosch's BASIC LEVEL results.

Classifier systems comprise an extraordinarily rich source of data for the study of human categorization. The question of whether they have the kinds of rich internal structures, like those of English prepositions, has not been seriously investigated. However, there is some indication not only that such cases do occur, but that they may be the norm. The literature on classifiers is full of cases where classifiers do not neatly fit a classical theory based on necessary and sufficient conditions. Pamela Downing (personal communication) offers the example of the Japanese classifier *hon*, whose most representative use seems to be for long, thin, rigid objects. Thus, it can classify sticks, canes, pencils, trees, candles, etc. Not surprisingly, it can be used to classify dead snakes and dried fish, both of which are long and rigid. But *hon* can be extended to, presumably, less representative cases:

-Pitches in baseball (straight trajectories, formed by a solid object)

-Rolls of tape (which can be unrolled into something long and thin)

-Telephone calls (which come over wires and which are instances of the CONDUIT metaphor,as described in (Reddy,1979) and Lakoff and Johnson, 1980))

-Radio and tv programs (like telephone calls, but without the wires)

-Letters (communication; and in traditional Japan, letters were scrolls, and hence stick-like -- and this is still a very strong conventional image)

-Movies (like radio and tv, plus they come in reels, like rolls of tape)

These uses of *hon*, though conventional, do not appear to be arbitrary. They seem to be motivated conventional uses, like the senses of English prepositions. The motivations are:

-Relations between image schemas, like the relation between a long, thin, rigid object and a baseball trajectory

-Metaphors, like the CONDUIT metaphor for communication

-Idealized cognitive models, say of what dried fish look like and what letters used to look like.

Viewed in this way, *hon* can also be considered as a category of senses structured around a prototypical sense, with less representative senses linked to the prototype via image relationships, metaphors, and culturally-based idealized cognitive models. As is usual, the less representative senses are motivated, but they are not predictable from the prototype and have to be learned. The alternative to such an analysis is that *hon* designates a disjunctive category -- a list of things that are not understood as having anything to do with each other. That seems unlikely in the case of *hon*, but it may be true. It may also be true that some senses are understood as motivated, and therefore as "natural extensions"

of the prototypical sense, while others are understood as being arbitrary. Moreover, this may vary from person to person. In short, we do not know how much of the above analysis of *hon* is real and how much is fanciful analysis. The question is an empirical one and has not been settled, for this classifier or for any other. My best guess is that the categories of senses picked out by classifiers will, on the whole, have a prototype structure like that suggested above for *hon*, and that most uses of classifiers will be motivated by image relationships, metaphors, and culturally-based ICMs; however, I would also be surprised if there were no arbitrary, unmotivated, and disjunctive members of such categories.

SOME CHARACTERISTICS OF NATURAL CATEGORIES

Within the theory of ICMs, natural categories will have at least the following characteristics:

- One or more ICMs, each characterizing representative members. The ICMs, if there are more than one, may be discrete or they may overlap with one another.

- Conventions specifying boundary areas. These may vary with context.

- Motivations for the inclusion of members in the category.

SUMMARY

The senses of polysemous lexical items seem to constitute natural categories. These categories have the following characteristics:

-They have prototypical members.

-They have conventional boundary conditions, such that members very close to the prototype can be predicted from properties of the prototype and a concept of perceived similarity, but members further from the

representative members must be specified by convention.

-Part of what constitutes 'perceived similarity' are links defined by image relationships, conceptual metaphors, and culturally-based idealized cognitive models. These provide motivation for the inclusion of nonrepresentative members in the category. -/// The properties that play a role in categorization are 'humanly relevant' interactional properties, rather than purely objective properties that "exist out there in the world independent of human beings". These include perceptual properties (images), interactional properties in the domains of motor actions, functions, and social roles.

Insert #

-Nonrepresentative members may bear little or no perceived similarity to one another.

-Each sense of such a lexical item is itself a natural category, whose members are in realm of human experience.

In short, the senses of polysemous lexical items seem to form natural categories in Rosch's sense.

COMPLEX CATEGORIZATION

The traditional view of the problem of complex categorization arises from the assumptions of the classical theory of concepts and categories. Suppose you assume that:

-There are primitive concepts.

-Meaning is truth conditional.

-The meaning of the whole is a truth conditional function of the meanings of the parts.

Then then classical problem of complex categorization arises:

-Exactly what are the primitive predicates and exactly how do you get the meanings of the wholes from the meanings of the parts?

But Rosch's BASIC LEVEL results contradict the assumptions of the classical theory. They suggest that:

-There are basic level concepts, but not necessarily primitive concepts.

-Meaning is based on human perception, interaction, and understanding, and is therefore not truth conditional.

Within the theory of natural categorization, the problem of complex categorization in its classical form does not arise at all. But the classical problem was based on a correct empirical observation:

-People create new sentences all the time, and are able to understand new sentences they've never heard before.

The question naturally arises: How is this possible? People do learn a finite stock of linguistic expressions and they do put them together to form new ones that they can understand. Exactly how?

Given the theory of natural categorization, this problem is very different from the classical problem of complex categorization. The problem is set within a cognitive theory that is neither reductionistic nor objectivist. The things available to such a theory are mental images (not just visual images, but sound images, force images, etc.), perceptual and other cognitive processes, patterns of motor activity, intentions, cognitive models, and an extremely rich background of knowledge and experience.

In the classical theory, you have two choices for characterizing set membership: you can predict the members (by precise necessary and sufficient conditions, or by rule) or you can arbitrarily list them, if there is a finite list. The only choices are predictability and arbitrariness. But in a theory of natural

categorization, the concept of *motivation* is available. Cases that are fully motivated are predictable and those that are totally unmotivated are arbitrary. But most cases fall in between -- they are partly motivated.

Differences like these make possible suggested solutions to O&S' examples of *striped apple* and *pet fish*. Consider for example Kay's Parsimony Principle (Kay, 1982), which was originally inroduced for entirely different reason -- to handle discourse-based inferences. Adapted to the theory of ICMs, it says (informally and somewhat oversimplified): **When a number of ICMs are evoked make them overlap as much as possible, consistent with your background knowledge.** In this case, the relevant aspects of the evoked ICMs in the *striped apple* example are our idealized image of stripes and our idealized image of an apple. The Parsimony Principle, yields a simple image overlap -- an apple with stripes -- for our new complex ICM. This is O&S' prototypical striped apple, and it works just as it should. The clause "consistent with your background knowledge" is a version of a general principle used both in AI research and in Linguistics: More specific knowledge takes precedence over more general knowledge. In other words, if you don't know about specific cases, use whatever general principles you have. But if you know something about a specific case, use what you know. This accounts for cases like *pet fish*. We happen to know about the kind of fish many people (at least in America) keep in their houses in fishbowls and fish tanks, and that guppies are typical of such fish. That knowledge overrides the general Parsimony Principle. An incidental consequence is that the expression *pet fish* as used to describe guppies is not completely motivated by the meanings of *pet* and *fish*, but it is partly motivated. This accounts for the feelings on the part of most of the people I've asked that the expression *pet fish* is not an ideal description of the guppy-like creature in the fishtank, but in the absence of anything better it will do.

These suggestions for accounting for the *striped apple* and *pet fish* examples are not available in the classical theory. They involve mental images, idealized cognitive models, background knowledge, the concept of partial motivation, and cognitive processing (use of the Parsimony Principle with its specific-information proviso). They seem like plausible suggestions, but they would hardly satisfy a classical theorist since they use ideas not permitted within the classical theory.

NONREDUCTIONISTIC CONCEPTS

One of the problems that has faced the theory of natural categorization is that most academics, at least in America, are trained within the classical theory and have trouble comprehending what a nonreductionistic theory of concepts could possibly be like. I have been asked again and again what it could possibly mean for there not to be primitive concepts. How are concepts grounded? What is there to hold onto?

Here is the way I think about concepts which cannot be decomposed into primitives: Concepts are grounded in human experience -- in perception, in action, in physcial and social interaction. Recall Rosch's basic level ~~objects~~ categories – in the middle of the set-theoretical hierarchy. These are 'human-sized' concepts. They are characterized by clusters of 'interactional properties', that is, perceptual properties (what a chair looks like to people), motor action properties (what people do when they sit in a chair), etc. Such concepts are basic relative to human experience, but they are not primitive building blocks. Concepts are grounded at the basic level. Basic level concepts, and the kinds of experiences that give rise to them, are what we have to hold onto.

Our 'starting point' is at the basic level. But what does it mean that the basic level is "in the middle' of the hierarchy of categories? I take it as meaning that, using our cognitive abilities, we can both generalize 'upward', forming

superordinate categories, and analyze 'downward' forming subordinate categories. Starting at the basic level, we can generalize up and analyze down as far as our cognitive capacities can take us, perhaps indefinitely.

Instead of primitive concepts, we have basic-level concepts, many of which seem to be fundamental to human experience. The theory of natural categorization thus provides a very different approach to the study of conceptual universals than does the classical theory of primitive concepts.

PROTOTYPES IN PHONOLOGY AND SYNTAX

Since phonology and syntax both involve categorization, it would be surprising if the theory of natural categorization did not apply in those areas. To date, prototype effects have been found in both phonology and syntax. Let us begin with phonology.

JAEGER'S RESULTS

Jaeger (1980) is an extensive experimental study which replicates Rosch's results in phonology. In short, Jaeger has demonstrated that, so far as psychological reality is concerned, prototype theory appears to be correct for phonology. Here is a brief summary of her results:

-Phonemes are natural categories of speech sounds and they are psychologically real. Phonemic categories have a prototype structure, that is, they may have representative and nonrepresentative members (allophones). Specific experimental results show:

In English, the [k] after word-initial [s] is part of the /k/ phoneme and not either the /g/ phoneme or some velar archiphoneme.

In English, the affricates [tʃ] and [ʤ] are considered to be unitary phonemes.

English speakers consider the following vowel pairs to belong together in a psychologically unified set: [ey-ae], [i-ɛ], [ow-a], [u-ʌ]. The source of the speaker's knowledge about this set of alternations is the orthographic system of English.

-Phonetic features in general have psychological reality, but not all the features proposed in various theories do. [Continuant],[sonorant], and [voice] are confirmed as real by the experiments, but [anterior] is brought into question.

-Phonetic features are not binary, but consist of a dimension along which segments can have varying values. A psychologically real theory must allow for the possibility of more than one correct feature assignment for a segment.

-Psychologically real phenomena in phonology can originate from a number of different sources. Most knowledge about phonology comes from pre-literate language acquisition, but orthography, education, and community myths play a role as well.

These experimental results call into question much of orthodox generative phonology.

ROSS' SQUISHES

In a number of studies ranging widely over English syntax, John Robert Ross (1973,1974, 1978,1981) has shown that just about every syntactic category in the language shows prototype effects. These include categories like noun, verb, adjective, clause, preposition, noun phrase, verb phrase, etc. Ross also demonstrated that general constructions in English show prototype effects, for

example, passive, relative WH preposing, question WH preposing, topicalization, conjunction,etc.

Ross' data is so rich that it is difficult to give any simple examples. Among the simplest I could find were these:

- *Near* has properties of both an adjective and a preposition. It takes the suffix *-ness*, as in *nearness to NP*, which otherwise goes exclusively on adjectives. It pied-pipes like a preposition, as in:

Near which tree did you see him digging?

- Tensed *that* clauses act like prototypical clauses, and constructions like *John's house* act like prototypical NPs with respect to constructions that take clauses and NPs. However, infinitival clauses, gerundive clauses, expressions of the form *the verb+ing of NP*, and NPs with lexicalized nominalizations (like *destruction*) all partake of some clausal properties and some NP properties, and form a continuum between clause and NP.

-Constructions that Chomsky has described in terms of a single movement transformation (move alpha), namely, WH-QUESTION PREPOSING, WH-RELATIVE PREPOSING, TOPICALIZATION, ADVERB PREPOSING, OBJECT-TO-SUBJECT RAISING, etc., all show prototype effects. In fact, they show different prototype effects, depending on the nature of what is "moved over" (in transformational terminology).

- Nodes which define islands and constrain movement rules show prototype effects, with respect to their ability to constrain various kinds of movements.

In general, Ross' results are not merely inconsistent with various current theories within generative syntax; they are inconsistent with the whole endeavor of generative syntax, which depends strongly on the classical set-theoretical account of categories in almost every respect.

One unfortunate aspect of Ross' investigations is that they came a bit too early in the history of natural categorization studies. They preceded Rosch's work on basic level categories, and on certain other aspects of natural categorization. At the time Ross did his work, fuzzy set theory seemed a viable approach to prototype phenomena. Ross therefore limited his investigation to linear phenomena, which he referred to as 'squishes'. Ross even tried, in the spirit of fuzzy set theory, to quantify syntactic phenomena explicitly, and the result was a dismal failure as a theory. Still, the data remain, and they still show prototype effects almost everywhere in syntax. Hopefully future syntactic investigations will take into account the subtleties of the theory of natural categorization that have emerged since Ross did his research in the area.

ICMS IN SYNTAX

Suggestions have been made to use devices like ICMs to characterize prototypical syntactic constructions. Lakoff's (1977) linguistic gestalts and Langacker's (1982) functional assemblies are moves in that direction.

GRAMMATICAL RELATIONS: THE BATES-MACWHINNEY HYPOTHESIS

Bates and MacWhinney (1980) have proposed that the theory of natural categorization can be used to characterize the grammatical relation SUBJECT in the following way:

Prototypical subjects are both agents and topics.

-AGENT and TOPIC are both natural categories centering around prototypes.

-Inclusion in the category SUBJECT cannot be completely predicted from the properties of agents and topics. As usual in natural categories, things that are very close to prototypical members will most likely be in the category and be relatively representative members. And as expected, the boundary areas will differ from language to language. Category membership will be motivated by (though not predicted from) family resemblances to prototypical members.

-This predicts that items that are neither prototypical agents nor prototypical topics can be highly representative subjects, providing that they have important agent and topic properties.

-This permits what we might call a 'prototype-based' universal: SUBJECT IS A NATURAL CONCEPTUAL CATEGORY WHOSE MOST PROTOTYPICAL MEMBERS ARE BOTH PROTOTYPICAL AGENTS AND PROTOTYPICAL TOPICS.

This definition of subject is semantically-based, but not in the usual sense; that is, it does not attempt to predict all subjects from semantic and pragmatic properties. But it does define the prototype of the category in semantic and pragmatic terms. This leaves room for language-particular conventions, not arbitrary ones, but conventions that are motivated by family resmblances to prototypical members.

Some preliminary work is now being done to check out this hypothesis:

- Research by Jack Hawkins and Jeanne van Oosten on German and Dutch respectively indicates that German and Dutch differ from English and from each other in the range of NPs that can be subjects. However, in each case, NPs close to the Bates-MacWhinney prototype can be subjects. The variation occurs at some distance from the prototype.

(based on ROHDENBURG, 1974)

In Standard Brasilian Portugese, verbs agree with their subjects. In Non-standard Portugese, Margarida Salomao reports (in-progress dissertation) that verbs can agree with locative adverbs, if they have sufficient topic properties. This is in accord with the hypothesis, since topic properties motivate subject-hood for a nonrepresentative subject. Bates-MacWhinney would have been contradicted had locatives been capable of agreement only if they were new information and not the discourse topic, for example.

Jeanne van Oosten, in ongoing dissertation research, reports that the range of uses of the passive in discourse is predicted by the Bates-MacWhinney hypothesis: the subject of the passive is closer to the agent-topic prototype than is the object of the by-phrase (overt or understood). This is accord with the results of Van Oosten (1977) and Lakoff (1977) on patient-subject constructions like *This car drives well*, in which the patient has more subject properties than the agent.

John DuBois (in a 1981 lecture at Berkeley, not yet published) reported that studies of ergativity in Mayan languages had led him to a variant of the Bates-MacWhinney hypothesis: the prototype for the ABSOLUTIVE category in ergative languages is both PATIENT and NEW INFORMATION. DuBois suggested that this is a conceptual category that is highly functional, and he provided evidence that, although there are only bare syntactic traces of it in English syntax, it plays an important role in English discourse structure.

It is too early to tell how much empirical suport these preliminary investigations will provide for the Bates-MacWhinney hypothesis and for prototype-based universals in general. The hypothesis is, however, extremely interesting, and provides an alternative to theories of grammatical relations that have no

semantic-pragmatic underpinning at all. It also adds to the study of universal grammar an important new type of universal -- the prototype-based universal. It is particularly interesting that this hypothesis was advanced not by linguists, but by developmental psycholinguists attempting to account for the data of language acquisition. And it accords with what is known about natural categorization in general.

CONCLUSION

The classical theory of concepts and categories has been studied in the West for two thousand years. It has become so much a part of Western culture and education that it is hard to think in other terms. When I first heard Rosch present her results on basic-level categorization, I was thrown almost into a state of shock. They contradicted the world-view that I was brought up to accept as if no other could possibly exist. My subsequent research on metaphor with Mark Johnson (1980) has reinforced Rosch's BASIC-LEVEL RESULTS in suggesting that neither reductionism nor objectivism can be maintained given a close look at the linguistic evidence.

REFERENCES

Barwise, Jon. 1980. Scenes and Other Situations. In Barwise and Sag, eds., *Stanford Working Papers in Semantics, Vol. 1.*

Barwise, Jon and John Perry. 1980. The Situation Underground. In Barwise and Sag, eds., *Stanford Working Papers in Semantics, Vol. 1.*

Bates, Elizabeth and Brian Mac Whinney. 1980. Functionalist Approaches to Grammar. In L. Gleitman and E. Wanner, eds., *Language Acquisition: The State of the Art*, Cambridge: Cambridge University Press.

Becker, Alton L. 1975. A Linguistic Image of Nature: The Burmese Numerative Classifier System. *Linguistics*, Whole No. 165, 109-121.

Berlin, Brent. 1968. *Tzeltal Numeral Classifiers*. The Hague: Mouton.

Brugman, Claudia. 1981. *Story of* Over. University of California, Berkeley, M.A. Thesis. unpublished.

Chomsky, Noam. 1981. *Lectures on Goverment and Binding*. Dordrecht: Foris.

Coleman, Linda. 1975. The Case of the Vanishing Presupposition. In *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society.*

Coleman, Linda and Paul Kay. 1981. Prototype Semantics: The English Verb *Lie*. {LANGUAGE}. 57:1.

Denny, J. Peter. 1976. What are Noun Classifiers Good For? In *Proceedings of the Twelfth Regional Me4eting of the Chicago Linguistic Society.*

Fillmore, Charles. 1975. An Alternative to Checklist Theories of Meaning. In *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society.*

Itzl

Fillmore, Charles. 1978. The Organization of Semantic Information in the Lexicon. In *Chicago Linguistic Society Parasession on the Lexicon.*

Fillmore, Charles. 1982. Towards a Descriptive Framework for Spatial Deixis. In Jarvella and Klein, eds., *Speech, Place, and Action.* London: John Wiley.

Gentner, Dedre. 1981. Gnerative Analogies as Mental Models. In *Proceedings of the Third Annual Conference of the Cognitive Science Society.*

Gleitman, Lila. 1981. What Some Concepts Might Not Be. Address to Annual Meeting of the Jean Piaget Society. Psychology Department, University of Pennsylvania, Philadelphia. ms.

Hinton, Leanne. 1982. How to Cause in Mixtec. In *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society.*

Jaeger, Jeri. 1980. *Categorization in Phonology: An Experimental Approach.* University of California, Berkeley, Ph.D. Dissertation. Unpublished.

Kay, Paul. 1979. The Role of Cognitive Schemata in Word Meaning: Hedges Revisited. Berkeley Cogntive Science Program. ms.

Kay, Paul. 1981. Three Properties of the Ideal Reader. Berkeley Cognitive Science Program. Unpublished ms.

Kay, Paul and Chad McDaniel. 1978. On the Linguistic Significance of the Meanings of Basic Color Terms. *LANGUAGE.* 54:3.

Lakoff, George. 1973. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic.* 2:458-508.

Lakoff, George. 1977. Linguistic Gestalts. In *Proceedings of the Thirteeth*

Regional Meeting of the Chicago Linguistic Society.

Lakoff, George. 1982. Experiential Factors in Linguistics. In Simon and Scholes, eds., *Language, Mind, and Brain.* Hillsdale, N.J.:Erlbaum.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By.* Chicago:University of Chicago Press.

Langacker, Ronald. 1982. *Foundations of Cognitive Grammar.* Chapter 1 of Draft. Linguistics Department, University of California, San Diego. Ms.

Lawler, John and Richard Rhodes. 1981. Athematic Metaphor. In *Proceedings of the Seventeenth Regional Meeting of the Chicago Linguistic Society.*

Lindner, Susan. 1981. *A Lexico-Semantic Analysis of Verb-Particle Constructions with* Up *and* Out. University of California, San Diego. Ph.D. Dissertation. Unpublished.

Mervis, Carolyn, and Eleanor Rosch. 1981. Categorization of Natural Objects. *Annual Review of Psychology.* 32:89-115.

Osherson, Daniel and Edward Smith. 1981. On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition.* 9:1,35-58.

Quinn, Naomi. 1981. Marriage is a Do-It-Yourself Project: The Organization of Marital Goals. In *Proceedings of the Third Annual Conference of the Cognitive Science Society.*

Quinn, Naomi. 1982.'Commitment' in American Marriage: A Cultural Analysis. Presented at the 79th Annual Meeting of the American Anthropological Association. Ms.

Reddy, Michael. 1979. The Conduit Metaphor. In A. Ortony, ed., *Metaphor and Thought*. Cambridge: Cambridge University Press.

Rosch, Eleanor. 1973a. Natural categories. *Cognitive Psychology*. 4:328-350.

Rosch, Eleanor. 1973b. On the Internal Structure of Perceptual and Semantic Categories. in T. E. Moore, ed., *Cognitive Development and the Acquisition of Language*. New York: Academic.

Rosch, Eleanor. 1975. Cognitive Reference Points. *Cognitive Psychology*. 7:532-547.

Rosch, Eleanor. 1977. Human Categorization. In N. Warren, ed., *Studies in Cross-Cultural Psychology*. London: Academic.

Rosch, Eleanor. 1978. Principles of Categorization. In E. Rosch and B. B. Lloyd,eds., *Cognition and Categorization*. Hillsdale, N. J.: Lawrence Erlbaum.

Ross, John Robert. 1972. The Category Squish: Endstation Hauptwort. in *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*.

Ross, John Robert. 1973a. A Fake NP Squish. In Bailey and Shuy, eds., *New Ways of Analyzing Variation in English*. Washington: Georgetown University Press.

Ross, John Robert. 1973b. Nouniness. In Osamu Fujimura, ed., *Three Dimensions of Linguistic Theory*. Tokyo: TEC Corporation.

Ross, John Robert. 1974. Clausematiness. In E. Keenan, ed., *Semantics For Natural Languages*. Cambridge: Cambridge University Press.

Ross, John Robert. 1981. Nominal Decay. Unpublished ms.

Shibatani, Masayoshi, ed. 1976. *The Grammar of Causative Constructions*. New

York: Academic.

Sweetser, Eve Eliot. 1981. The Definition of *Lie*: An Examination of the Folk Theories Underlying a Semantic Prototype. Presented at the Eightieth Annual Meeting of the American Anthropological Association. Unpublished ms.

Weschler, Lawrence. 1982. *Seeing is Forgetting the Name of the Thing One Sees.* Berkeley: University of California Press.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. New York: Macmillan.

Zadeh, Lotfi. 1965. Fuzzy Sets. *Information and Control*. 8:338-353.