Classifiers As A Reflection Of Mind:

A Cognitive Model Approach To Prototype Theory

By

George Lakoff

Cognitive Science Program

Institute of Cognitive Studies

University of California at Berkeley

January 1984

This research was supported in part by Grant No. BNS-8310445 from the National Science Foundation and by a grant from the Sloan Foundation to the University of California at Berkeley. Part of this report was delivered at a Conference on Categorization and Noun Classification, organized by Collette Craig and Talmy Givon at the University of Oregon in October, 1983. I would like to thank the organizers and participants for their comments and suggestions, especially R.M.W. Dixon and Pamela Downing who have been extremely kind in providing me with examples and allowing me to cite and comment on their research.

The production of the Berkeley Cognitive Science Report Series is supported by a grant in cognitive science to the University of California at Berkeley from the Alfred P. Sloan Foundation.

Classifiers As A Reflection Of Mind:

A Cognitive Model Approach To Prototype Theory

George Lakoff

Cognitive Science Program

Institute of Human Learning

University of California at Berkeley

January 1984

This research was supported in part by Grant No. BNS-8310445 from the National Science Foundation and by a grant from the Sloan Foundation to the University of California at Berkeley. Part of this report was delivered at a Conference on Categorization and Noun Classification, organized by Collette Craig and Talmy Givon at the University of Oregon in October, 1983. I would like to thank the organizers and participants for their comments and suggestions, especially R.M.W. Dixon and Pamela Downing who have been extremely kind in providing me with examples and allowing me to cite and comment on their research.

Preface

This is an exciting time to be studying the mind, especially those aspects of the mind that can be revealed through the study of language. For more than two thousand years, it has been generally assumed that people categorize things simply on the basis of shared properties. A category has been taken to be a kind of abstract container characterized by necessary and sufficient conditions for membership. But beginning with observations by Wittgenstein and empirical research by Berlin, Kay, Rosch, and others, that view is changing in a significant way.

> One consequence of the classical theory of categorization is that all members of categories should be equal -- that none should have any special status, none should be a 'better' category member than any other. A great many studies have shown that this is not the case. Some members of categories act as if they had privileged status. Such cases are called *prototype effects*, and the members with privileged status are referred to as *prototypes*. Prototype effects have many sources, and part of this study will be dedicated to sorting them out. In the course of doing so, we will suggest that prototype effects are by-products of what we will refer to as *cognitive models*.
>
> We will approach the role that cognitive models play in categorization through the study of systems of human categorization built overtly into natural languages -- classifier systems. Such systems are common throughout the languages of the world, and we will be discussing the basic mechanisms by which at least two such systems work. Cognitive models are included in those mechanisms.

Women, Fire, and Dangerous Things

Borges attributes the following taxonomy of the animal kingdom to an ancient Chinese encyclopedia entitled the *Celestial Emporium of Benevolent Knowledge*.

> On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance. (Borges, 1966, p. 108)

Borges, of course, deals with the fantastic. These not only are not natural human categories -- they could not be natural human categories. But part of what makes this passage art, rather than mere fantasy, is that it comes close to the impression a Western reader gets when reading descriptions of nonwestern languages and cultures. The fact is that people around the world categorize things in ways that both boggle the Western mind and stump Western linguists and anthropologists. More often than not, the linguist or anthropologist just throws up his hands and resorts to giving a list -- a list that one would not be surprised to find in the writings of Borges.

An excellent example is the classification of things in the world that occurs in traditional Dyirbal, an aboriginal language of Australia. The classification is built into the language -- as is common in the world's languages. Whenever a Dyirbal speaker uses a noun in a sentence, the noun must be preceded by a variant of one of four words: *bayi, balan, balam, bala*. These words classify all objects in the Dyirbal universe, and to speak Dyirbal correctly one must use the the right classifier before each noun. Here is a brief version of the Dyirbal classification of objects in the universe:

(I) *bayi*

> men, kangaroos, possums, most snakes, most fishes, some birds, most insects, the moon, storms, rainbows, boomerangs, some spears, etc.

(II) *balan*

> women, bandicoots, dogs, some snakes, some fishes, most birds, fireflies, scorpions, crickets, the hairy mary grub, anything connected with water or fire, sun and stars, shields, some spears, some trees, etc.

(III) *balam*

> edible plants, all trees with edible fruit, wine, cake, cigarettes

(IV) *bala*

> parts of the body, meat, bees and honey, wind, yamsticks, some spears, most trees, grass, mud, stones, noises and language, etc.

It is a list that any Borges fan would take delight in.

But Dixon did not stop with a list. He was determined to learn what made these categories real categories of the human mind, categories that made sense to Dyirbal speakers -- that they could learn uniformly and use unconsciously and automatically. In the course of his fieldwork, Dixon observed that speakers do not learn category members one-by-one, but operate in terms of some general principles. Here is how they work. There is a basic, productive and simple categorization schema that operates unless some specialized knowledge, belief, or myth takes precedence. The basic schema is:

(I) *bayi*

human males

(II) *balan*

human females

(III) *balam*

edible plants

(IV) *bala*

everything else

The Special Case Principle:

Specialized knowledge takes precedence over general rules.

The Animal Principle:

Animals are categorized in I, with human males.

The Motivation for the Animal Principle:

There is no male-female subcategorization for animals. (For example, there are no separate words for male and female kangaroos.)

The male category is "unmarked" relative to the female category. Cases that are neutral between male and female fall into the male category. This is common in the languages of the world.

The categorization of humans is central; the categorization of animals is secondary to the categorization of humans. This is a general principle of human cognitive organization.

Thus, it makes sense that animals would be categorized with humans in general, and with human males in particular.

The Domain of Experience Principle:

If there is a basic domain of experience associated with A, then it is natural for entities in that domain to be in the same category as A. (For example, implements used in fishing are in the same category as fish.)

The Mythic Model Principle:

> If there is a basic cultural myth in which B is A, then it is natural for B to be in the same category as A.

These are the basic principles. Let us see how they apply in major cases. We will start with one of the most basic myths of all among the Dyirbal:

> The sun is the wife of the moon.

Thus, the sun is in category II with women, while the moon is in category I with men. Now consider some of the most basic domains of the experience:

> The main producers of light and heat are the sun and fire.

Thus, it makes sense for fire to be in the same category as the sun -- category II.

> Fire is dangerous.

By the domain of experience principle, it makes sense for fire and dangerous things to be in the same category.

> Fighting is dangerous.

Again, by the domain of experience principle, it is reasonable to find fighting in the same category as dangerous things in general. Similarly, it would make sense for fighting implements to be in the same category as fighting.

> Water extinguishes fire.

For aborigines living in the Australian bush, this is one of the most prominent phenomena of every-day experience. By the domain of experience principle, it would make sense for water and fire to be categorized together. But wait. Fire and water can also be viewed as opposites. Wouldn't that make them candidates for being in different categories? Yes and no. In order for things to be opposites, they first must be in the same domain of experience. Conceptually, water is closer to its opposite, fire, than it is to, say, numbers or words or marriages. But what about the sun and the moon? They too are in the same domain of experience. Why shouldn't they be categorized the same way. The reason seems to be that a basic myth imposes on the moon and the the sun a male-female distinction. The basic distinctions within the system seem to be male vs. female, and animal vs. edible plant. There seems to be nothing in the system that imposes such distinctions on water as opposed to fire. In order for opposite to catergorized as opposites, there must be an appropriate opposition within the linguistic sustem. In short, the *system as a whole* must be taken into account in order to make sense of individual cases of categorization.

At this point, we have the somewhat more complex categorization:

(I) *Human males; animals, the moon.*

(II) Human females; fire; dangers; fighting; water.

(III) Edible plants.

(IV) Everything else.

By taking into account other basic facts of Dyirbal experience and other basic Dyirbal myths, Dixon was able to make sense of the bulk of the Dyirbal classification system, and to show how the system was learned by speakers of the language. For example:

-Birds are believed to be the spirits of dead human females, so they are in class II. But, three willy wagtails are believed to be mythical men, so they are in class I. The satin bird is in class II, but for a different reason: it is the mythical bringer of fire.

-Fishes, being animate, are mostly class I, but the stone fish and the gar fish are dangerous, and are therefore in class II.

-Trees, bushes, vines and grasses are not edible, and therefore are class IV, not III. However, stinging trees and vines are harmful -- the pain can sometimes last for months -- and so they are in class II with other dangerous things.

-Things in the same domain as fire and the sun are in class II, for instance, light and the stars.

-Anything connected with fighting (implements, fighting ground) is in class II. This includes fighting spears. But multi-pronged spears used primarily for fishing are, like fishing line, in the fish class, class I.

-Crickets are old ladies in myths, and so are class II.

-The firefly, being associated with fire, is class II, not I.

-The hairy mary grub, whose sting is said to feel like sunburn, is in the class with the sun, class II, rather than in class I with the other insects.

The classification of loan words is predictable from this analysis:

-Fruit, flour, cake (made from flour), and wine (made from fruit) are in class III with the edible plants.

-Pipes and matches are in class II with fire, but cigarettes (leaves that are ‘consumed’) are in class III (edible plants).

Dixon was unable to find explanations for all the classifications, partly because he never became sufficiently familiar with the relevant myths and beliefs. Moreover, he suggests that since many myths and beliefs die out while the classification remains, some classifications will be without explanation, and hence (synchronically) arbitrary and unmotivated. For example, three animals --

bandicoots, dingos, and platypuses -- are exceptions. That is, speakers know that they are categorized in class II rather than in class I with the other animals, and they know that they are exceptional cases, but they can provide no reaons as to why they are exceptional. Dixon points out, however, that these cases are still not utterly random. It is no accident that these exceptions went in class II (defined by women as central members), rather than in class III with the edible fruits or class IV with others. These are, after all, exceptions to the rule that would place them in class I, with men as central members. Thus, as exceptions, they are in opposition to the normal cases, which are in the men-centered-class. Therefore, if they are to be marked as exceptions, it makes sense to place them in a class that is in opposition to the men-centered-class, namely, class II -- the women-centered-class. Even the exceptions are not randomly distributed.

Dixon's achievement is remarkable. He has shown that what might look superificially, to the Western eye, as a fantastic classification out of Borges, is from the perspective of the people doing the classifying a relatively regular and principled way to classify things.

In the process, Dixon has provided a remarkable example of how human cognition works. Though the details of categorization may be unique to Dyirbal, the general principles at work in the Dyribal system show up again and again in systems of human categorization. They are:

-*Centrality*: Human males and females are the central members of categories I and II. Willy wagtails and the moon are less central members of category I. Stinging vines, gar fish and the hairy mary grub are less central members of category II.

-*Chaining*: Complex categories are structured by chaining: central members are linked to other members, which are linked to other members, and so on. For example, women are linked to the sun, which is linked to sunburn, which is linked to the hairy mary grub.

-*Experiential Domains*: There are basic domains of experience, which may be culture-specific. These can characterize links in category chains. Fishing, for example, is a significant domain of experience in Dyirbal life. The implements used in this domain are categorized in the same class as the fish.

-*Idealized Models*: There are idealized models of the world -- myths among them -- that can characterize links in category chains. The myth in which birds are bringers of fire provides a link for categorizing birds with fire.

-*Specific Knowledge*: Specific knowledge overrides general knowledge.

-*The Other*: Borges was right about this. Conceptual systems can have an "everything else" category. It, of course, does not have central members, chaining, etc.

-*Overall Systematicity*: Categorization depends on the kinds of distinctions made in the system as a whole. For example, the male-female distinction is basic to the system and leads to the sun and moon being categorized differently, while there is nothing in the system that imposes a similar distinction on fire and water.

-*No Common Properties*: Categories on the whole need not be defined by common properties. There is no reason to believe that the Dyirbal find anything in common among women, fire, dangerous things, etc. Nor do they assume, so far as is known, that there is anything feminine about fire or danger, or anything fiery or dangerous about women. On the other hand, common properties seem to play a role in characterizing the basic schemas (edible plant, human male, human female).

-*Motivation*: The general principles given make sense of Dyribal classification, but they do not predict exactly what the categories will be.

Of these principles, motivation is perhaps the least obvious and is worth some discussion. There is a big difference between giving principles that *motivate,* or *make sense* of, a system, and giving principles that *generate*, or *predict*, the system. Dixon's analysis explains why the Dyirbal system is the kind of a system that human beings can function with. It does not predict exactly what the system will be. For example, one must learn which domains of experience are relevant to categorization and which are not. Thus, fish live in water, and fish are in class I, but that does not make water class I with fish, nor does it make fish class II with water. The domain of habitation is simply not important for Dyirbal classification. Dyirbal speakers simply must learn which domains of experience matter for classification, and which myths matter.

What *is* predicted is that systems of classification tend to be structured in this way, that is, that there tends to be centrality, chaining, etc. The theory of categorization makes predictions as to what human category systems can and cannot be like. It does not predict exactly what will be in a given category in a given culture or language.

Evidence from Language Death

Dixon's analysis on the whole makes sense of the Dyirbal system. But how do we know that it is right? Couldn't this just be some arbitrary analysis imposed by an outside analyst? Couldn't it simply be akin to an analysis of a literary work imposed by a critic? How do we know that there is anything psychologically real about Dixon's analysis?

In fact, there is evidence that Dixon's analysis was basically correct. It comes from an unfortunate source. Dixon did his intitial field research in 1963. At that time, English-speaking Australian culture had not yet impinged much on the Dyirbal community. But in the years since then, Dyirbal children have been sent to English-speaking schools and Australian culture has gone a long way toward wiping out Dyirbal culture. As of 1983, only twenty years later, Dyirbal culture and the Dyirbal language are dying. Young people in the Dyirbal community grow up speaking English primarily, and learn only an extremely oversimplified version of traditional Dyirbal. Their experiences are very different from those of their parents and they either don't learn the myths or find that the myths don't mean much in their lives.

As one might expect, this has resulted in a drastic change of the categorization system. This has been documented by Annette Schmidt in a remarkable study, *Young People's Dyirbal: An example of language death from Australia* (Schmidt, 1983). Dyirbals who are roughly 45 or older speak traditional Dyirbal. Speakers who are under 35 speak some more simplified form of the language. Schmidt's study of the overall demise of Dyirbal included two degenerate stages of the categorization system: an intermediate stage (from a 35 year old speaker) and a very simple system (from much younger speakers). These two degenerate stages provide evidence that the analysis given above, or something close to it, is correct for traditional Dyirbal.

The intermediate stage lacks some of the hypothesized links that allow things that are not males and females to be categorized in classes I and II.

-Danger and fighting are eliminated from class II. Thus, the dangerous fish (stonefish and garfish) are in class I with the other other fish. The stinging nettle is in class IV with the other trees. Fighting implements -- spear and shield -- are no longer in class II; they are also in class IV. Fishing implements -- fish line and fish spear -- are no longer in class I with fish; they too are in class IV with "others".

But other parts of the traditional system are kept:

-The mythical model links are preserved: the moon is still in class I with males, while the sun, stars and birds are still in class II with females.

-Fire and water are kept in class II with females and the sun.

In the simple stage of the youngest speakers, the system has broken down almost completely and only the central cases of classes I and II survive, while class III is lost completely. Here is the simple system:

(I) *-human males*

(II) *-human females*

(IV) *-everything else*

This kind of breakdown of the system indicates that the traditional system is not just a random list. It suggests that the hypothesized central members and chains are real, since in the intermediate stage parts of the chain appear, as wholes, to be split off -- in particular, danger, fighting, and fishing. And in the final stage, only central members of classes I and II are left.

The analysis of the Dyirbal classifier system given above can be diagrammed as follows:

Here the universe is divided up into four clearly defined mutually exclusive domains, represented by the boxes. These form what we will be calling a *base model*. All the base model tells us is that there are four distinctions. Three of them have an internal structure, with elements at the center. The centers are indicated by squares in the diagram. The fourth, being made up of what is left over from the first three, has no internal structure. The centers of the domains in the base model are also structured -- by what we will call a *basic opposition model*.

human males vs. human females, or: center of I vs. center of II

people vs. edible plants, or: centers of I and II vs. center of III

Finally, and perhaps most significantly, there is the chaining structure inside the base model.

The clearly defined mutually exclusive domains of the base model are consistent with the classical theory of categories. But that does not make it a classical system. This system differs from the classical theory in that it does not have any defining characteristics that are shared by the members of the categories.

To describe this system, we need:

-a *base model*, which in this case is very simple. It specifies that there are four distinct mutually exclusive categories, and that the fourth is made up of what is not in the first three.

-a specification of which subcategories are central, or *most typical*, of the first three categories.

-a basic opposition model, which structures the centers with respect to one another.

-a specification of chaining principles.

-a short list of exceptions, which are distributed according to the basic opposition model.

The Dyirbal classifier system exhibits many of the basic mechanisms used in human categorization. I would now like to turn to another case that exhibits other fundamental mechanisms used in categorization. This example is based on field research done by Pamela Downing (personal communication).

Japanese *Hon*

The Japanese classifier, *hon*, in its most common use, classifies long, thin, rigid objects: sticks, canes, pencils, candles, trees, 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 what are presumably less representative cases:

-martial arts contests, with staffs or swords (which are long and rigid)

-hits (and sometimes pitches) in baseball (straight trajectories, formed by the forceful motion of a solid object, associated with baseball bat, which is long, thin and rigid)

-shots in basketball, serves in volleyball, and rallies in ping pong

-judo matches (a martial arts contest, but without a staff or sword)

-a contest between a zen master and student, in which each attempts to stump the other with zen koans.

-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 by Reddy (1979) and Lakoff and Johnson (1980).

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

-letters (another instance of communication; moreover, in traditional Japan, letters were scrolls and hence stick-like);

-movies (like radio and tv; moreover, they come in reels like rolls of tape).

-medical injections (done with a needle, which is long, thin, and rigid)

These cases, though not predictable, are nonetheless not arbitrary. They do not have anything in common with long, thin rigid objects, but it *makes sense* that they might be classified in the same way. Let us ask exactly what kind of sense it makes.

Let us begin with martial arts contests using staffs or swords. Staffs and swords are long, thin, rigid objects, which are classified by *hon*. They are also the principal functional objects in these matches. A win in such a match can also be classified by *hon*. That is, the principal goal in this domain of experience is in the same category as the principal functional object.

Baseball bats are central members of the *hon* category. They are one of the two most salient functional objects in the game, the other being the ball. Baseball is centered on a contest between the pitcher and the batter. The batter’s principal goal is to get a hit. When a baseball is hit solidly, it forms a trajectory -- that is, it traces a long, thin path along which a solid object travels quickly and with force. The image traced by the path of the ball is a *hon* image -- long and thin.

The extension of the *hon* category from bats to hits is another case of an extension from a principal functional object to a principal goal. It is also an extension from one principal functional object with a *hon* shape to a *hon*-shaped path formed by the other principal functional object. Incidentally, in the small amount of research done on *hon* to date, it appears that while base hits and home runs are categorized with *hon*, foul balls, pop flies, ground balls and bunts appear not to be. This is not surprising, since these are not principal goals of hitting, nor do their trajectories form a *hon* shape.

Such a relationship among image schemas-- between long, thin things and trajectories -- is a common relationship that forms the basis for the extension of a category from a central to a noncentral case. Let us consider three examples from English.

-The man ran into the woods.
-The road ran into the woods.

In the first case, *run* is used for a case where there is a (long, thin) trajectory. In the second case, *run* is used for a long, thin object, a road.

-The bird flew over the yard.
-The telephone line stretched over the yard.

In the first case, *over* is used for a (long, thin) trajectory. In the second case, *over* is used for a long, thin object, a telephone line.

-The rocket shot up.
-The lamp was standing up.

In the first case, *up* is used for a trajectory. In the second case, *up* is used for a long, thin object, a standing lamp.

Such relationships are common and suggest that there exists which might be called an *image-schema transformation* of the following sort:

TRAJECTORY SCHEMA <-----> LONG, THIN OBJECT SCHEMA.

This image-schema transformation is one of the many kinds of cognitive relationships that can form a basis for the extension of a category.

Some speakers of Japanese extend the *hon* category to baseball pitches, as well as hits -- again on the basis of such an image-schema relationship within the same domain of experience. Some speakers extend *hon* to pitches using both the trajectory and the contest-perspective, in which the hitter and pitcher are engaged in a contest. These speakers use *hon* only for pitches seen from the point of view of the hitter. There are also speakers who classify pitches with *hon* only if they achieve the principal goal of pitching. Since getting strikes is the principal goal of pitching, such speakers can classify strikes, but not balls, with *hon*. No speakers have been found who use *hon* to classify balls but not strikes. Similarly, no speakers have been found who classify bunts and foul balls with *hon*, but not home runs and base hits.

There are similar motivations behind the extensions of *hon* to other concepts in sports. Thus, *hon* can classify shots and free throws in basketball, but not passes. And it can classify serves in volleyball and rallies in ping pong. These are cases where there is both a trajectory and a possibility of scoring (achieving a principal goal).

There are several morals to be drawn from these examples:

First, what are taken to be the central cases for the application of *hon* appear to be concrete basic-level objects: sticks, pencils, bamboo staffs, baseball bats, etc. The direction of extension appears to go from concrete basic-level objects to other things, like hits and pitches.

Second, a theory of motivations for the extension of a category is required. Among the things we need in such a theory are image-schema transformations and conceptual metonymies -- cases where a principal object like a staff or bat can stand for a principal goal like a win or hit.

Third, hits in baseball and long, thin rigid objects do not share common properties. The relationship between them is given by an image-schema transformation and a metonymy. Hence, the classical theory, which requires that categorization be based on common properties, is inadequate.

Fourth, the application of *hon* to baseball pitches may make sense, but it is not predictable. It is a matter of convention -- not an arbitrary convention, but a *motivated* convention. Thus, the traditional view that everything must be either predictable or arbitrary is inadequate here. There is a third choice: motivation. In this case, the independently needed image-schema transformation and the object-for-goal metonymy provide the motivation.

Ideally, each instance of the use of a classifier outside the central sense should have a motivation. The motivation cannot be ad hoc -- one cannot just make up a metonymy or image-schema just to handle that case. It must be justified on the basis of other cases. This imposes a criterion of adequacy on the analysis of classifier languages.

Some investigators have suggested that such a criterion of adequacy is too strong; they have claimed that some classifications simply are arbitrary and that no non-ad-hoc motivation exists. That is an empirical question, and the facts are by no means all in. But arbitrariness is a last resort. Even if there are some completely unmotivated cases, one can still apply a slightly weakened criterion of adequacy. Find out which extensions "make sense" to speakers and which extensions seem "senseless", and account for those that make sense. Each sensible extension of a category needs to be independently motivated. No analysis of a classifier system is complete until this is done.

So far, we have seen that metonymies and image-schema transformations can provide motivation for the extension of a category. Another important kind of motivation comes from conventional mental images. Take the example of a roll of tape, which can be classified by *hon*. We know what rolls of tape look like, both when they are rolled up and when they are being unrolled. That is, we have conventional mental images of tape, both when it is in storage form and when it is being put to use. We also know that we unroll tape when we are about to use it, and that the tape is functional when it is unrolled. A conventional image of tape being unrolled has two parts -- the rolled part and the unrolled, functional, part. The image of the unrolled, functional part fits the long, thin object image-schema associated with the central sense of *hon*. The image of the nonfunctional rolled part does not fit the central *hon* image-schema. Metonymy is involved here; the functional part of the conventional image is standing for the whole image, for the sake of categorization. The functional part fits the *hon* schema. This is, presumably, the motivation for the use of *hon* to classify rolls of tape. Again, we cannot predict the use of *hon* for rolls of tape; but we can do something that is extremely important. We can show why it makes sense. Making sense of categorization is no small matter. And doing so in a manner that shows in detail how basic cognitive mechanisms apply is anything but trivial. If the cognitive aspects of categorization are to be understood, it will require attention to detail at this level.

Medical injections are another case where the principal functional object (the needle) is long and thin; the needles can be classified with *hon*, and by metonymy, so can the injections.

So far we have seen how image-schema transformations, conventional mental images, and metonymy all enter into categorization by a classifier. Let us turn to a case that involves all of these plus metaphor. Recall that *hon* can be used to classify telephone calls. The conventional mental image of engaging in a telephone call involves using the most functional part of the telephone -- the receiver, which is a long, thin rigid object and fits the central image-schema for *hon*. The other principal conventional image related to telephone calls involves telephone wires. These are understood as playing a principal functional role in telephonic communication. These fit the long, thin object image-schema. They also fit the CONDUIT of the CONDUIT metaphor -- the principal metaphor for communication. In short, there are two related but different motivations for the use of *hon* for telephone calls. That is, there are two ways in which this fits the conceptual system, and where motivation is concerned, the more kinds of motivation, the better. That is, it is not a matter of finding which is right; both can be right simultaneously.

So far, we have seen that extended senses of *hon* can be based on the central sense of *hon*. But extended senses may themselves serve as the basis for further extensions via category chaining. Recall that letters are classified with *hon*. There are a number of considerations that motivate such a categorization. First, letters were originally in the form of scrolls, often wound around long thin wooden cylinders. They have been categorized with *hon* ever since, and that image remains very much alive in Japanese culture through paintings and the tradition of calligraphy. Second, the conventional image of writing a letter involves the use of a pen, which plays a principal functional role, and is also a long, thin object. Third, letters are a form of communication, and therefore an instance of the CONDUIT metaphor. These diverse motivations allow *hon* to fit the ecology of the Japanese classifier system.

Letters and telephone calls are intermediate steps in a chain. Radio and tv programs are also classified with *hon*. They are forms of communication at a distance, like letter-writing and and telephone communication. They too are motivated by the CONDUIT metaphor for communication. Given that letters and telephone calls are classified by *hon*, radio and tv programs constitute a well-motivated extension. Movies are also classified by *hon*. They are also instances of communication at a distance; in addition, one of the principal conventional images associated with movies is the movie reel, which looks like a spool of tape, which is classified with *hon*.

The phenomenon of category-chaining shows very clearly that the classical account of categorization is inadequate. Sticks and tv programs are both in the *hon* category, but they share no relevant common properties. They are categorized in the same way by virtue of the chain structure of the *hon* category.

Finally, let turn our attention to judo matches and contests between zen masters and students. Judo matches are in the same domain of experience as martial arts contests with staffs or swords. A win in judo match can also be classified as a *hon*. Similarly, zen contests are, in Japanese culture, in the same experiential domain as martial arts contests, and a win there also can be classified as a *hon*.

Incidentally, the noncentral cases of the *hon* category vary in some cases from speaker to speaker. Thus some speakers do not include baseball pitches and some do not include wins in zen contensts. But to my knowledge, every speaker of Japanese includes the central members -- the candles, staffs, baseball bats, etc. Moreover, many of the extensions have become conventionalized for speakers in general: letters, telephone conversations, home runs, spools of thread. The variation just displayed involves chaining that has not yet stabilized but which shows the same principles at work as in the stable cases.

A Cognitive Model Approach to Prototype Theory

Language is among the most characteristic of human cognitive activities. To understand of human beings categorize in general, one must at least understand human categorization in the special case of natural language. The two cases we have just discussed are quite typical of the way human categorization functions in natural language. What is it about the human mind that allows it categorize in this way? Is there some general cognitive apparatus used by the mind that gives rise to categorizations of this sort?

The analyses given above suggest that such categories can be characterized using cognitive models of four types:

Propositional models: These specify elements, their properties, and the relations holding among them.

> Much of our knowledge structure is in the form of propositional models. Thus, a model of a domain (like fighting in Dyirbal) would include elements that occur in that domain (like fighting spears). A propositional model characterizing our knowledge about fire would include the fact that fire is dangerous. A taxonomic model, like the base model for Dyirbal given above, would include four elements corresponding to each of the categories, and a condition stating that anything not a member of the first three categories is in the fourth.

Image-schematic models: These specify schematic images, such as trajectories or long, thin shapes or containers.

> Our knowledge about baseball pitches includes a trajectory schema. Our knowledge about candles include a long, thin object schema.

Metaphoric Models: These are mappings from a propositional or image-schematic model in one domain to a corresponding structure in another domain.

> The conduit metaphor for communication maps our knowledge about conveying objects in containers onto an understanding of communication as conveying ideas in words.

Metonymic Models: These are models of one or more of the above types, together with a function from one element of the model to another.

> Thus, in a model that represents a part-whole structure, there may be a function from a part to the whole that enables the part to stand for the whole. In Dyirbal knowledge about the hairy Mary grub, knowledge about its sunburn-like sting may stand for the grub itself in determining that it is a member of the same category as the sun.

Such models can characterize the base structure, indicate what the central members are, and characterize the links in the internal chains.

The distinction between base structure and internal structure is important to bear in mind. In Dyirbal, the base model sets up clear boundaries but is not sufficient to characterize membership in the categories. However, there are categories where the base model is sufficient to distinguish members from nonmembers of the category, and where there is still an internal structure.

Take, for example, the category corresponding to the English word *bird*. There are criteria for distinguishing birds from nonbirds. Birds have wings, feathers, lay eggs, have beaks, have two legs, etc. However, as Eleanor Rosch has shown, people view some birds as better, or more central examples of the category than other birds. Thus, robins and sparrows are judged as better examples of birds than eagles, pelicans, chickens, ostriches and penguins. As Rosch discovered, there seems to be a representation of central members that includes a mental image of a prototypical bird, plus various properties that are not shared by all birds: flying, not predatory, not domestic, singing, etc. And Rips has shown that this internal structure is used in certain forms of reasoning. For example, in one experiment some subjects were told that all the robins on an island got a certain disease, and were asked if they thought the ducks would get it. They tended to say yes. In the reverse situations, when subjects were told that all the ducks on an island got a certain disease, they tended not to think that the robins would necessarily get it. In other words, certain inferences go from central to non-central members, but not vice versa.

The category *bird*, like the categories *bayi*, *balam*, and *balan* in Dyirbal, has both a base and an internal structure. The category *bird* happens to have a base structure that distinguishes birds from nonbirds. The internal structure plays a different role in cognition in this case.

Many cases have been isolated in which central members of categories exhibit different cognitive characteristics than noncentral members. These are called *prototype effects*. Central members are recognized faster, are learned earlier, are used more frequently, facilitate matching tasks, and are used in what Rosch (1981) refers to as "reference point reasoning" -- cases like the robin-duck example cited above. On the whole, central members seem to be used in comprehending the category as a whole. They therefore both aid in recognition, recall, and learning, and form the basis from which people generalize in certain situations.

Prototype effects are superficial phenomena. They arise when some subcategory or member or submodel is used (often for some limited and immediate purpose) to comprehend the category as a whole. In other words, these are cases where a part (a subcategory or member or submodel) stands for the whole category -- in reasoning, recognition, etc. Within the theory of cognitive models, such cases are represented by metonymic models.

More specifically, a metonymic model has the following characteristics:

-There is a "target" concept A to be understood for some purpose in some context.

-There is a conceptual structure containing both A, and another concept B.

-B is either part of A, or is closely associated with it in that conceptual structure. Typically, a choice of B will uniquely determine A, within that conceptual structure.

-Compared to A, B is either easier to understand, easier to process, easier to recognize, or more immediately useful for the given purpose in the given context.

-A metonymic model is a model of how A and B are related in a conceptual structure, together with a function from B to A.

When such a conventional metomymic model exists as part of a conceptual system, B may be used to stand, metonymically, for A. If A is a category, the results is a metonymic model of the category.

Just as there are many types of metonymy in general (see Lakoff and Johnson, chapter 8), there are many types of categorial metonymy, where a part of a category stands for the whole in order to serve some cognitive purpose. As a result there are many types of prototypes -- as many are there are possible categorial metonymies. I have by no means isolated the full range of such cases. Here are the types I have compiled so far.

Typical Examples

These are cases like the following:

Robins and sparrows are typical birds.
Apples and oranges are typical fruits.
Saws and hammers are typical tools.

And presumably, in Dyirbal,

Men are typical *bayi*.
Women are typical *balan*
Edible plants are typical *balam*

The use of typical category members is usual unconscious and automatic. They are not the subject of public discussion, and they seem not to change

Social Stereotypes

Social stereotypes are usually conscious and the subject of public discussion. They are subject to change over time, and they may become public issues. They are used in reasoning and especially in what is called "jumping to conclusions". Their use in reasoning may be overtly challenged. Here are some examples:

The stereotypical politician is conniving, egotistical, and dishonest.
The stereotypical mother is a housewife.
The stereotypical bachelor is macho, dates a lot of different women, is interested in sexual conquest, hangs out in bars, etc.
The stereotypical Japanese is industrious, polite, and clever.

Social stereotypes are most commonly used to make quick judgements about individuals. They are exploited in advertising, and in most forms of popular entertainment -- especially news stories.

Ideals

Many categories are understood in terms of abstract ideal cases -- which may be neither typical nor stereotypical. For example:

The ideal husband: a good provider, faithful, strong, respected, attractive.

The stereotypical husband: bumbling, dull, pot-bellied, ...

Ideals are used in making judgments and in making plans.

Paragons

We also comprehend categories in terms of individual members who represent either an ideal or its opposite. Thus, we have institutions like the ten-best and ten-worst lists, the Halls of Fame, Academy Awards, the Guiness book of World Records, etc. We have baseball paragons: Babe Ruth, Willie Mays, Sandy Koufax, etc. Paragons are made use of in constructions in the language: *a regular Babe Ruth, another Willie Mays, the Cadillac of vacuum cleaners, etc. Scientific paradigms are also characterized by paragons. Thus, for example, The Michaelson-Morley Experiment is the paragon of physics experiments -- and is used by many people to comprehend what a great experiment in physics is.*

Generators

There are cases where the members of a category are defined, or "generated", by the central members plus some general rules. Numbers are the best-known example. The natural numbers are, for most people, characterized by the integers between zero and nine, plus addition and multiplication tables and rules of arithmetic. The single-digit numbers are central members of the category *natural number* which happen to generate the entire category, given certain principles. In our system of numerical representation, single digit numbers are employed in comprehending natural numbers in general. The properties of large numbers are understood in terms of the properties of smaller numbers, and ultimately in terms of the properties of simgle-digit numbers.

The single-digit numbers, together with addition and multiplication tables and rules of arithmetic, constitute a model that both generates the natural numbers and is metonymic in our sense: the category as a whole is comprehended in terms of a small subcategory.

The natural numbers, in addition, have certain models that subdivide the numbers according to certain properties -- odd and even, prime and nonprime, etc. These models are not metonymic. They work by classical Aristotelian principles. But they only define *subcategories* of the natural numbers. The category as a whole is defined metonymically and generatively by the single-digit numbers plus rules of arithmetic.

To make matters more complicated, other kinds of numbers are also defined by metonymic generative models: the rationals, the reals, the imaginaries, the transfinite cardinals, etc. Thus rational numbers are understood as ratios of natural numbers, and real numbers are understood as infinite sequences of natural numbers. In other words, the rationals and the reals are understood metonymically in terms of the natural numbers -- a subcategory used to generate the larger categories.

Submodels

Another way to comprehend a category is via a submodel. Take the category of natural numbers again. The most common submodel used is the subcategory of powers of ten: ten, a hundred, a thousand, etc. We use this submodel to comprehend the relative size of numbers. The members of the submodel are among what Rosch refers to as *Cognitive Reference Points*, which have a special place in reasoning.

Some submodels have a biological basis: the primary colors, the basic emotions, etc. Others are culturally stipulated, e.g., the Seven Deadly Sins.

Salient Examples

It is common for people to use familiar, memorable, or otherwise salient examples to comprehend categories. For example, if your best friend is a vegetarian and you don't know any others well, you will tend to generalize from your friend to other vegetarians.

Prototype effects can have a wide variety of sources. Such effects exist for all of the above cases where a category is understood metonymically in terms of a subcategory, member, or submodel. It is thus extremely important to try to isolate the sources of those effects. A theory of cognitive models that permits one to characterize types of metonymic models seems to be necessary if one is to understand how such effects come about.

Experiential Clusters

Metonymic models are by no means the only sources of prototype effects. Let us use the term *base model* to refer to models that are "ordinary" rather than metonymic. It commonly happens that a number of base models intersect to form a complex cluster that is psychologically more basic than the models by themselves. We will refer to these as *experiential clusters*.

Mother

An example is the concept *mother*. According to the classical theory, it should be possible to give clear necessary and sufficient conditions for *mother* that will fit all the cases and apply equally to all of them. Such a definition might be something like: *a woman who has given birth to a child*. But as we will see, no such definition will cover the full range of cases. *Mother* is a concept that is based on a complex model in which a number of individual base models converge, forming an experiential cluster. The models in the cluster are:

-The birth model: the person giving birth is the *mother*.

The birth model is usually accompanied by a genetic model, although since the development of egg and embryo implants, they do not always coincide.

-The genetic model: the female who contributed the genetic material is the *mother*.

-The nurturance model: the female adult who nurtures and raises a child is the *mother* of that child.

-The marital model: the wife of the father is the *mother*.

-The geneological model: the female of the first ascending generation is the *mother*.

The concept *mother* normally involves a complex model in which all of these individual base models converge to form a cluster. But because of the complexities of modern life, they have come to diverge more and more. Still, many people feel the pressure to pick one base model as being the right one, the one that "really" defines what a mother is. But although one might try to argue that only one of these characterizes the "real" concept of *mother*, the linguistic evidence does not bear this out. As the following sentences indicate, there is more than one criterion for "real" motherhood:

-I was adopted and I don’t know who my real mother is.

-I am not a nurturant person, so I don’t think I could ever be a real mother to any child.

-My real mother died when I was an embryo, and I was frozen and later emplanted in the womb of the woman who gave birth to me.

-I had a genetic mother who contributed the egg that was planted in the womb of my real mother, who gave birth to me and raised me.

-By genetic engineering, the genes in the egg my father’s sperm fertilized were spliced together from genes in the eggs of twenty different women. I wouldn’t call any of them my real mother. My real mother is the woman who bore and raised me, even though I don’t have any single genetic mother.

In short, more than one of these models contributes to the characterization of a *real mother*, and any one of them may be absent from such a characterization. Still, the very idea that there is such a thing as a *real mother* seems to require a choice among models where they diverge. It would be bizarre for someone to say:

-I have four real mothers: the woman who contributed my genes, the woman who gave birth to me, the woman who raised me, and my father’s current wife.

When the cluster of models that jointly characterize a concept diverge, there is still a strong pull to view one as the most important. This is reflected in the institution of dictionaries. Each dictionary, by historical convention, must list a primary meaning when a word has more than one. Not surprisingly, the human beings who write dictionaries have varied on their choices. Dr. Johnson chose the birth model as primary, and many of the applied linguists who work for the publishers of dictionaries, as is so often the case, have simply played it safe and copied him. But not all. Funk and Wagnall’s Standard chose the nurturance model as primary, while the American College Dictionary chose the geneological model. Though choices made by dictionary-makers are of no scientific importance, they do reflect the fact that, even among people who construct definitions for a living, there is no single, generally accepted base model for such a common concept as "mother".

When the base models for *mother* diverge, we get compound expressions like *stepmother*, *surrogate mother, adoptive mother, foster mother, biological mother, donor mother*, etc. Such compounds, of course, do not represent simple subcategories, that is, kinds of ordinary mothers. Rather, they describe cases where there is a lack of convergence of the various models.

And, not surprisingly, different models are used as the basis of different extensions of *mother*. For example, the birth model is the basis of the metaphorical sense in

-Necessity is the mother of invention

while the nurturance model is basis for the derived verb in

-He wants his girlfriend to mother him.

The geneological model is the basis for the metaphorical extension of *mother* and *daughter* used in the description of the tree diagrams that linguists use to describe sentence structure. If node A is immediately above node B in a tree, A is called the *mother* and B, the *daughter*. Even in the case of metaphorical extensions, there is no single privileged base model for *mother* on which the extensions are based. This accords with the evidence cited above which indicates that the concept *mother* is defined by a cluster of converging base models.

This phenomenon is beyond the scope of the classical theory. The concept *mother* is not clearly defined, once and for all, in terms of common necessary and sufficient conditions. There need be no necessary and sufficient conditions for motherhood shared by normal biological mothers, donor mothers (who donate an egg), surrogate mothers (who bear the child, but may not have donated the egg), adoptive mothers, unwed mothers who give their children up for adoption, and stepmothers. They are all mothers by virtue of their relation to the ideal case, where the base models converge. That ideal case is one of the many kinds of cases that give rise to prototype effects.

Working Mothers

The category *mother* has two levels of prototype structure. One level is defined by the experiential cluster described above. A second level of prototype structure has to do with the fact that, even in the ideal case, where all the above models converge, there is also a stereotype: the mother as housewife. We have no term *housewife mother* -- we don't need one since that is taken to be the norm. But the housewife-mother subcategory, though unnamed, exists. The named category *working mother* is defined in contrast to the stereotypical housewife-mother.

The housewife-mother stereotype arises from a stereotypical view of nurturance, which is associated with the nurturance model. According to the stereotypical view, mothers who do not stay at home all day with their children cannot properly nurture them. There is also a stereotypical view of work, according to which it is done away from the home, and housework and childrearing don't count. This is the stereotype that the bumpersticker "All Mothers Are Working Mothers" is meant to counter.

The housewife-mother stereotype is therefore defined relative to the nurturance model. This may be obvious, but it is not a trivial fact. It shows that metonymic models like stereotypes are not necessarily defined with respect to an entire cluster. In this case, the metonymic model is characterized relative to only one of the models in the cluster -- the nurturance model. Here is some rather subtle evidence to prove the point:

Consider an unwed mother who gives up her child for adoption and then goes out and gets a job. She is still a mother, by virtue of the birth model, and she is working -- but she is not a *working mother*!

The reason is that it is the nurturance model, not the birth model, that is relevant. Thus, a biological mother who is not responsible for nurturance cannot be a working mother, though an adoptive mother, of course, can be one.

This example shows the following:

-A social stereotype (e.g., the housewife-mother) may be defined with respect to only one of the base models of an experiential cluster (e.g., the nurturance model).

-A subcategory (e.g., working mother) may be defined in contrast with a stereotype (e.g., the housewife-mother).

-When both of the above occur, it is only the relevant base model (e.g., the nurturance model) that is used as a background for defining the subcategory (e.g., working mother).

Thus, only those mothers for whom nurturance is an issue can be so categorized. Stepmothers and adoptive mothers may also be working mothers, but biological mothers who have given up their children for adoption and surrogate mothers (who have only had a child for someone else) are not working mothers -- even though they may happen to be holding down a job.

Such models of stereotypes are important for a theory of conceptual structure in a number of ways. First, as we have seen, they may be used to motivate and define a contrasting subcategory like *working mother*. This is important because, according to the classical theory, such cases should not exist. In the classical theory, social stereotypes have no *cognitive* function -- that is, no role at all in defining *concepts* and *conceptual categories*. But the fact that the conceptual category *working mother* is defined by constrast with the housewife-mother stereotype indicates that stereotypes do have a role in characterizing concepts.

Secondly, stereotypes define a normal expectation which is linguistically marked. For example, the word *but* in English is used to mark a situation which is in constrast to some model that serves as a norm. Stereotypic models may serve this function:

NORMAL: She is a mother, but she isn't a housewife.
STRANGE: She is a mother, but she's a housewife.

The latter sentence could only be used if stereotypical mothers were not housewives.

In summary, we have seen two kinds of models for *mother*:

-An ideal model, consisting of a cluster of converging base models.
-A stereotypic model, which is a metonymic model, not a base model.

Both models are prototypical, but in different ways. Together, they form a cluster with a composite prototype: the best example of a mother is a biological mother who is a housewife, principally concerned with nurturance, not working, a generation older than the child, and married to the child's father. This composite prototype imposes what is called a *representativeness structure* on the category: the closer an individual is to the prototype, the more representative a mother she is.

Representativeness structures are linear. They concern nothing but closeness to the prototypical case, and thus they hide most of the richness of structure that exists in the cognitive models that characterize the category. Representativeness structures, though real, are mere shadows of cognitive models.

It is important to bear this in mind, since prototype theory is sometimes thought of as involving only such linear representativeness structures and not cognitive models. This is an extremely impoverished view of prototype theory.

The study of representativeness structures has played an important role in the history of prototype theory -- largely in demonstrating that prototypes do exist and in making a bare first approximation to finding out what they are and what properties they have. But a full study of category structure must go well beyond just isolating a prototype and giving a linear ranking of how close nonprototypical cases are. At the very least, it must provide an account of the details of the cognitive models that give rise to the representativeness structure.

Radial Structures

Here are some kinds of mothers:

-The central case, where all the models converge. This includes mothers who are and always have been female, and who gave birth to the child, supplied her half of child's genes, nurtured the child, is married to the father, is one generation older than the child, and is the child's legal guardian.

-Stepmother: She didn't give birth or supply the genes, but she is currently married to the father.

-Adoptive mother: She didn't give birth or supply the genes, but she is the legal guardian and has the obligation to provide nurturance.

-Birth mother: There is a system of adoption in which a woman who gives birth and doesn't want to raise the child can choose adoptive parents for the child. In that system, the woman who gives birth is called the *birth mother*.

-Foster mother: She is being paid by the state to provide nurturance.

-Biological mother: She gave birth to the child, but is not raising it and there is someone else who is and who qualifies to be called a mother of some sort.

-Surrogate mother: She has contracted to give birth and that's all. She may or may not have provided the genes, and she is not married to the father and is not obligated to provide nurturance. And she has contractually given up the right to be legal guardian.

-Unwed mother: She is not married to the father at the time of the birth.

-Genetic mother: This is a term I have seen used for a woman who supplies an egg to be

> planted into someone else's womb, and has nothing else whatever to do with the child. It has not yet to my knowledge become conventional.

These subcategories of mother are all understood as deviations from the central case. But not all possible variations on the central case exist as categories. There is no category of mothers who are legal guardians but who don't personally supply nurturance, but hire someone else to do it. There is no category of transsexuals who gave birth but since had a sex-change operation. Moreover, some of the above categories are products of the twentieth century, and simply did not exist before. The point is that the central case does not productively generate all the these subcategories. Instead, the subcategories are defined by convention as variations on the central case. There is no general rule for generating kinds of mothers. They are culturally defined and have to be learned. They are by no means the same in all cultures. In the Trobriands, a woman who gives birth often gives the child to an old woman to raise. In traditional Japanese society, it was common for a woman to give her child to her sister to raise. Both of these are cases of kinds of mothers that we don't have an exact equivalent of.

The category of *mother* in this culture has what we will call a *radial structure*. A radial structure is one where there is a central case and conventionalized variations on it. Categories that are generated by central cases plus general principles -- say, the natural numbers -- are ruled out. We are limiting radial structures only to cases where the variations are conventionalized and have to be learned. We are also ruling out cases where the central case is just more general than the noncentral case -- that is, where the noncentral cases just have more properties than the central case, but no different ones.

Dyirbal categories I, II, and III have radial structures. Category IV, the "other" category does not. Japanese *hon* has a radial structure. Radial category structures are extremely common in natural language.

Fuzziness: Graded and Nongraded Sources

There is a big difference between prototypicality and fuzziness. Fuzziness arises when category membership is a matter of degree. In all the examples we have considered so far, there is prototypicality of various kinds, but no fuzziness. Examples may be more or less central without being fuzzy. Thus a mother may diverge from the housewife-stereotype and still be no less a mother. Ducks and vultures are not prototypical birds, but they are no less birds.

Fuzziness and prototypicality have often been confused -- and there is a good reason why. Fuzziness results in prototype effects. And there are two principal sources of fuzziness. Graded concepts are the most obvious. These have been discussed by Lotfi Zadeh for many years. They are concepts like *tall, rich, middle-aged,* etc. Such concepts have scales built into them. People are rich to degrees, they may be more or less middle-aged. Consider the category *middle-aged man*. A fifty-year-old man is a better example of a middle-aged man than a thirty-five-year-old man. When one reaches seventy, one is clearly past middle-age. The concept *middle-aged man* is graded, defined by a scale. The fuzziness is built into the concept itself. The prototypical cases are those which are clearly in the category.

Fuzziness may also arise from nongraded concepts -- concepts defined by models that have no scales built into them. Fillmore (1982) gives as an example the time-honored case of *bachelor*. He observes that *bachelor* is defined relative to an idealized model of the world, one in which there is a social institution of marriage, marriage is monogamous, and between people of opposite sexes. Divorce is not considered in this idealized model, nor are priests, monks, homosexuals, bigamists, eunuchs, nor long-term unmarried couplings. And in this idealized model, it is clear when men become mature and every mature man has the means to support a wife. Such an idealized model provides the background against which *bachelor* is defined. Every mature male in the model is clearly either married or unmarried. The unmarried ones are the *bachelors*.

This idealized model fits the classical theory of categories. Within the model, *bachelor* is a very clearly defined Aristotelian category. But this idealized cognitive model, or ICM, does not fit the world as we know it very well. When this model is placed within the context of the rest of our knowledge, fuzziness arises -- not because of what is in the model but because of discrepancies between the background assumptions of the model an the rest of our knowledge. Here are some cases where the background conditions fail, and as a result it is difficult to give clear, unequivocal answers:

-Is Tarzan a bachelor?
-Is the Pope a bachelor?
-Is a Moslem who has only three wives and is eligible to marry a fourth a bachelor?
-Is a gay man a bachelor? If same-sex marriage is allowed and he seeks to marry another gay man, is he then a bachelor?
-If a priest leaves the priesthood at age fifty, does he become a bachelor?

The answers to such questions are not clear-cut, and the reason is that the idealized model with respect to which *bachelor* is defined may not fit well with the rest of our knowledge. The source of fuzziness here is not within the model, but in the interaction of the model with other models characterizing other aspects of our knowledge.

Fuzziness of the above sort leads to prototype effects -- cases of better and worse examples of bachelors. Here *better* and *worse* refer to the degree to which one can say that the person is a member of the category at all. *Bachelor* is a particularly interesting case, because it has a second source of prototype effects -- one where there is no fuzziness at all and every case is clearly a member of the category. Such cases arise from our social stereotype of bachelors. Let us take situations where all the background conditions for the *bachelor* model apply. Among the peole who are clearly bachelors, some fit the stereotype and some do not.

> The stereotypical bachelor is macho, dates a lot of different women, is interested in sexual conquest, hangs out in bars, doesn't like to do housework or take care of children, etc.

There is no question that Joe Namath fits the bachelor stereotype. A nonmacho man who likes to take care children, works in a daycare center, has no interest in sexual conquest, and would never go near a bar or nightclub may be every bit a member of the *bachelor* category, but he simply isn't as good an example as Joe Namath because he doesn't fit the stereotype. Thus, the stereotype is another source of prototype effects.

The Sources of Nonclassical Effects

According to the classical theory, a category has clear, well-defined boundaries and membership in the category is defined by necessary and sufficent conditions shared by all the members of the category. Since these conditions are shared by all the members, no members should be better examples, or more central members, than any others. Of the examples we have considered, none work like this. But that does not mean that the Aristotelian account is entirely wrong.

Let us return to Fillmore’s *bachelor* example. The base model for *bachelor* is an idealized model which excludes many cases. If one sticks only to the cases included within the model, then that base model is a classical model. This base model -- and many others -- does include classical conditions. A nonclassical effect arises only when one goes outside the model to consider other things one knows about.

Why should a cognitive model that meets the classical conditions give rise to nonclassical effects? The reason is that the classic theory of categories is not a cognitive theory. It is a theory about things as they are in the world, not about how the mind makes sense of the world. If there is something correct about the classical theory, it is certainly not what Aristotle had in mind. What is correct about the classical theory is that it is used in part -- and only in part -- in cognitive models that are used to make sense of the world.

In other words, there are idealized models that do contain classical conditions. But these are *not* classical categories. They yield prototype effects, which classical categories should not do. Cognitive models with classical portions are designed to do a much more modest job than classical categories -- to make sense of a humanly restricted and humanly conceived portion of human experience. In the classical theory, they were to provide necessary and sufficient conditions for partitioning everything in the universe independently of any human mind. Human classification simply doesn’t fit this account, and there is no reason to think that the world does.

As we have seen, human categories are *very* nonclassical. Nonclassical effects arise from various sources: metonymic models of many kinds, clusters of base models, models of graded concepts, and the interaction between idealized classical-style models and knowledge those models weren’t designed to accommodate.

Categories of Mind, Or Mere Words

A possible objection to the kind of analyses we have been discussing is that classifiers are mere linguistic devices and do not reflect conceptual structure. That is, one might object that, say, the things categorized by *hon* in Japanese do not form a single conceptual category. Thus, one might suggest that the analysis of *hon* may show something about rules of language, but that it shows nothing about our conceptual system.

Let us, for the sake of argument, consider such a suggestion. Whatever their precise cogntive status is, rules of language are some part or other of our cognitive apparatus. Just what would such "rules of language" involve? In particular, they would involve all the things we discussed above in the analysis of *hon*:

-central and peripheral members
-basic-level objects at the center
-conventional mental images

-knowledge about conventional mental images
-image-schema transformations
-metonymy applied to mental imagery
metonymy applied to domains of experience
-conceptual metaphors (which map one conceptual domain into another)

These mechanisms are needed, no matter whether one calls them linguistic or not. Moreover, they appear to be the kinds of things that one would tend to call conceptual -- mental images and image transformations do not appear to be merely linguistic. Moreover, linguistic categories can be used in nonlinguistic tasks, as Kay and Kempton (1983) have demonstrated. But whether they are used in nonlinguistic tasks or not, linguistic categories *are* categories -- and they are part of our overall cognitive apparatus. Whether one wants to dignify them with the term "conceptual" or not, linguistic categories are categories within our cognitive system and a study of *all* categories within our cognitive system will have to include them.

Cognitive Categories, or Mere Historical Relics

Another objection occasionally raised to the kind of analysis proposed above is that the noncentral cases are "mere historical relics" not parts of a live cognitive system. Dixon has already rebutted this persuasively for the Dyirbal cases in his observation that speakers do not learn the system one case at a time but use general principles. Moreover, Schmidt's study of the decay of the Dyirbal system provides dramatic evidence that Dixon's analysis was basically correct and psychologically real.

In addition, the "mere historical relics" argument is completely invalid. There is nothing "mere" even about historical relics. When categories get extended in the course of history, there has to be some sort of cognitive basis for the extension. Chances are, the kinds of cognitive bases used are pretty much the ones we have discussed. At the time the extension of the category occurs, such mechanisms are quite real.

Experiential, Imaginative, and Ecological

Aspects of Mind

We are now in a position to see how classifier systems reflect the experiential, imaginative, and ecological aspects of mind. Let us begin with the experiential aspects.

As Denny (l976) 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.

Denny's observations fit nicely with observations by Berlin, Rosch and their coworkers on basic-level categorization. What they found was that basic level categorization depended on the nature of everyday human interaction both in a physical environment and in a culture. (See Berlin et al. (1974) and Rosch (1977). Factors involved in basic-level categorization include gestalt perception, motor interaction, mental images, and cultural importance.

Taken together, these observations support the view that that our conceptual system is dependent on, and intimately linked to, our physical and cultural experience. It disconfirms the classical view that concepts are abstract, and separate from human experiences.

The examples given above of *hon* in Japanese and Dixon's analysis of the Dyirbal classifier system display many of the imaginative aspects of mind -- especially the use of mental images, image-schema transformations, conceptual metonymies and conceptual metaphors.

Finally, the fact that extensions from the center of categories are neither predictable nor arbitrary, but instead are motivated, demonstrates the ecological character of the human mind. Motivation depends on overall characteristics of the conceptual system, not just local characteristics of the category at hand. In addition, the existence of the Dyirbal miscellaneous category indicates that categorization is not local, but at the very least depends upon contrasting alternatives.

The Simple Computationalist View

Because of the recent technological innovations in computer science, the mind-as-computer metaphor has taken hold, not only in the popular imagination, but among professional cognitive psychologists as well. According to this view, the mind is a computer with biological hardware, and runs using programs essentially like those used in computers today. I will refer to this as the *simple computationalist view* of mind. The results we have just cited are on the whole incompatible with the simple computationalist view.

To begin with, computers don't have bodies. If it is correct that our conceptual system is based on, and understood through, physical, social and other experiences, then computers simply do not now, and never will, understand anything. Moreover, in the absence of a simulation of the relevant aspects of those experiences, a computational account of mind will not be able to account for the learning and the structuring of basic-level concepts or classifier systems.

In addition, contemporary approaches to computation have not been able to even come close to adequately representing the imaginative aspect of mind. Though some limited work has been done on metaphor and the representation of mental images, nothing approaching the use of the kinds of imaginative devices used in classifier systems seems approachable using present techniques. To my knowledge, metonymy and image-schemas have not even been thought about -- and there is no obvious way at present to deal with them computationally. Moreover, the fact that metaphor, metonymy and image-schema transformations are all experientially-based makes it all that much more difficult to accomodate them computationally in an adequate way.

The ecological aspect of mind is also well beyond the simple computationalist view. Current systems have the predictable/arbitrary dichotomy built in. Prediction is equated with computability, and things are either computable or they are not. The simple computational view offers no third choice. Yet the human conceptual system has a rich third choice -- motivation -- and that is what it most commonly seems to use. Motivation is an ecological property of the conceptual system as a whole. Current computational models cannot deal adequately with motivation. Given our current understanding of what a computation is, it appears impossible to use it to construct a program that will learn, remember, and reason readily with a highly motivated case, but do so only with great difficulty with a relatively unmotivated case. In other words, there is no current or even seriously conceived of apparatus within computational systems to distinguish nonpredictable cases that fit a complex conceptual system from nonpredictable cases that don't fit the system -- and then function more efficiently with the case that fits the system. Such a capacity that would be required to take a computational view of mind beyond the simple level.

Classifier systems are among the many natural language phenomena that are beyond the simple computationalist view of mind. Formal linguistics, as it is done today, falls roughly within the simple computationalist paradigm. Formal syntax involves the algorithmic manipulation of uninterpreted symbols. Formal semantics involves the interpretation of those symbols in terms of models which are like data bases used in contemporary artificial intelligence. Neither formal syntax nor formal semantics as they are currently understood can deal adequately with the experiential, imaginative, and ecological aspects of mind. Perhaps that is why formal linguists continue to ignore the study of classifier systems.

What Prototype Theory Isn't

A number of proponents of the computational theory of mind have interpreted prototype theory, incorrectly, as fitting the simple computationalist view. A good example occurs in Osherson and Smith (1981), who characterize prototype theory as follows:"Prototype theory construes membership in a concept's extension as graded, determined by similarity to the concept's 'best' exemplar (or by some other measure of central tendency)." In other words, Osherson and Smith take prototype theory as claiming that membership in a category is completely predicted by the prototype and a computation based on similarity to the prototype.

This construal of prototype theory fits the simple computationalist view, with its focus on algorithmic computation and its predictable/arbitrary dichotomy. But this construal of prototype theory is a far cry from prototype theory as it has been developed by Berlin, Kay, Rosch, Fillmore, myself, and others. What the simple computationalist version of prototype theory ignores is what makes it most interesting to cognitive anthropologists and linguists -- the experiential, imaginative, and ecological aspects of mind. As we have seen, an adequate account of classifier systems cannot ignore these aspects of mind. Classifier systems not only disconfirm the classical theory, but they also disconfirm the simple computationalist version of prototype theory.

Epilogue

The Aristotelian Imperative

There is a common idea, at least in the West, and I would suspect in other cultures, that there is a single correct taxonomy of things in the world. A taxonomy is a cognitive model of a particular kind. They are a common feature of the mind. Taxonomies are built into languages throughout the world. They are means that human beings have used to make sense of their experience.

People have many ways of making sense of their experience -- and taxonomies of all sorts abound. Yet the idea that there is a single right taxonomy is so persistent that it puzzles me. Perhaps it arises from the relative stability of basic-level concepts (see Lakoff, 1982). Whatever the source, the idea is widespread. Taxonomies, after all, divide things into kinds, and it is commonly taken for granted that there is only one correct division into kinds. Since scientific theories develop out of folk theories, it is not at all surprising to find that folk criteria for the application of taxonomic models find their way into science. A particularly interesting example of this is discussed by Gould (1980) in his classic "What, If Anything, Is A Zebra?".

Gould describes the heated disputes betweeen two groups of biologists -- the cladists and the pheneticists. Each of these applies different criteria for determining the ‘one correct taxonomy’ of living beings. The pheneticists look at overall similarity in form, function, and biological role, while the cladists are primarily concerned with branching order in the course of evolution and look at *shared derived characters*, that is, features present only in members of an immediate lineage, and not in distant primitive ancestors. Ideally, overall similarity ought to converge with evolutionary branching order and yield the same taxonomy. Traditional taxonomists use both kinds of information. But in a considerable number of cases there is a divergence between the cladistic and phenetic taxonomic models.

Gould’s discussion is particularly interesting (p. 363):

> Some of our most common and comforting groups no longer exist if classifcations most be based on cladograms [evolutionary branching diagrams]...I regret to report that there is surely no such thing as a fish. About 20,000 species of vertebrates have scales and fins and live in water, but they do not form a coherent cladistic group. Some -- the lungfish and the coelacanth in particular -- are geneologically close to the creatures that crawled out on land to become amphibians, reptiles, birds, and mammals. In a cladistic ordering of trout, lungfish, and any bird or mammal, the lungfish must form a sister group with the sparrow or elephant, leaving the trout in its stream. The characters that form our vernacular concept of "fish" are all shared primitives and do not therefore specify cladistic groups.
>
> At this point, many biologists rebel, and rightly I think. The cladogram of trout, lungfish, and elephant is undoubtedly true as an expression of branching order in time. But must classifications be based only on cladistic information? A coelacanth looks like a fish, tastes like a fish, acts like a fish, and therefore -- in some legitimate sense beyond hidebound tradition -- *is* a fish.

Gould continues (p. 365):

> Phenetic similarity often correlates very poorly with recency of common ancestry. Our ideal world requires a constancy of evolutionary rate in all lineages. But rates are enormously variable. Some lineages change not at all for tens of millions of years; others undergo marked alterations in a mere thousand. When the forebears of terrestrial vertebrates first split off from a common ancestry with coelacanths, they were still unambiguously fish in appearance. But they have evolved, along numerous lines during some 250 million years, into frogs, dinosaurs, flamingos, and rhinoceroses. Coelacanths, on the other hand, are still coelacanths. By branching order, the modern coelacanth may be closer to a rhino than a tuna. But while rhinos, on a rapidly evolving line, are now markedly different from that common ancestor, coelacanths still look and act like fish -- and we might as well say so.

There are several things here worth noticing. First, both the cladists and pheneticists assume the classical Aristotelian theory of categorization. They are seeking well-defined classes based on shared characteristics. Where they differ is on which shared characteristics are to be considered. Second, the cladists, the pheneticists, and traditionalists like Gould, who try to balance both kinds of criteria, all follow the folk theory that there is only one correct taxonomy. Even though Gould recognizes the scientific validity of the cladists' views, he cannot simply say that there are two different taxonomies, equally correct for different reasons. As a traditional taxonomist, he feels forced to make a choice. Third, his choice is based on what he calls "subjective" criteria -- what a coelacanth looks like and tastes like to a human being. Of course, there is long tradition of using such humanly-based criteria in taxonomic biology.

The general point should be clear. There are two kinds of taxonomic models available to traditional biologists: the cladistic and the phenetic. Ideally, they are supposed to converge, and they do in a great many cases, but by no means in all. One may admit, as Gould does, that both have scientific validity. Still the force of the folk theory of taxonomic models is so strong that a choice must be made -- if not one or the other exclusively, then a third choice combining elements of both. Above all, there must be one correct taxonomy.

Why? Why does the Aristotelian imperative have that kind of force? Gould is no dope; he is an intelligent and reasonable man. There would be nothing wrong with saying that there are just two different taxonomic models of life forms, which are concerned with different and equally valid issues. Yet it doesn't seem like a possible alternative for him.

My guess is that we have a folk theory of categorization itself. It says that things come in well-defined kinds, that the kinds are characterized by shared properties, and that there is one right taxonomy of the kinds. It is basically an Aristotelian folk theory.

It is easier to show what is wrong with a scientific theory than with a folk theory. A folk theory defines common sense itself. When the folk theory and the scientific theory converge, things get even tougher. The folk theory of categorization and the classical Aristotelian theory of categories go hand-in-hand. But they do not correctly characterize categories of mind.

As a theory of human categorization, prototype theory is superior to the classical theory. But it doesn't exist as a folk theory, and I doubt that it ever will. Amos Tversky and Daniel Kahneman have shown repeatedly that people can (and probably do) get into trouble because they reason on the basis of what I have called metonymic models, but they don't know it.

I have not done any survey of how many sciences, or other academic fields, take the folk theory of categorization for granted. But I think the proportion must be substantial. I wonder what scientific and academic disciplines would be like if they were not limited by the classical folk theory of categorization and instead permitted their official theorizing to be done using the categories of the human mind.

References

Berlin, Brent and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: U. of California Press.

Berlin, Brent, Dennis E. Breedlove and Peter H. Raven. 1974. *Principles of Tzeltal Plant Classification*. New York: Academic.

Borges, J.L. *Other Inquisitions 1937 - 52*. New York: Washington Square Press, 1966.

Brugman, Claudia. 1981. *Story of* Over. University of California, Berkeley, M.A. Thesis. Available from Indiana University Linguistics Club, Bloomington, Indiana.

Borges, J.L. *Other Inquisitions 1937 - 52*. New York: Washington Square Press, 1966.

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

Dixon, R.M.W. 1968. Noun Classes. *Lingua*, XXI, 104-25.

Gould, Stephen Jay. 1980. *Hen's Teeth and Horses' Toes*. Norton.

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

Kay, Paul and Willett Kempton. 1983. What Is the Sapir-Whorf Hypothesis? Berkeley Cognitive Science Report No. 8, Institute for Human Learning, University of California at Berkeley.

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

Lakoff, George. 1981. Getting the Whole Picture: The Role of Mental Imagery in Semantics and Pragmatics. In *Proceedings of the Seventh Annual Meeting of the Berkeley Linguistics Society*.

Lakoff, George. 1982. Categories and Cognitive Models. Berkeley Cognitive Science Report No. 2, Institute for Human Learning, University of California at Berkeley.

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

Lindner, Susan. 1981. *A Lexico-Semantic Analysis of Verb-Particle Constructions with* Up *and* Out. University of California, San Diego, Ph.D. Dissertation. Available from Indiana University Linguistics Club, Bloomington, Indiana.

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

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

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

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

Rosch, Eleanor. 1981. Prototype Classification and Logical Classification: The Two Systems. Unpublished ms.