In the next chapter we will see that the correlational structure is important for the implications and consequences of categorization, and not just for the acquisition of categories. Moreover, we will see that children not only use correlated features to acquire categories, but in fact may start out assuming that categories have a correlated structure that goes well beyond what can be determined by visual inspection of exemplars. That is, children may assume that category members have properties in common that are not apparent from their perceptual features.

Chapter 5
Natural Kinds

In this chapter about natural kind terms I pursue the suggestion that children might expect categories, or at least categories of a special status, to have a correlated structure.

Recent arguments in philosophy about natural kind terms provide another challenge to traditional theories of concepts (Kripke 1971, 1972; Putnam 1970). According to these analyses, natural kind terms do not have meaning in the traditional sense.

The term "natural kind" has various uses in philosophy and psychology, but to begin with some clear cases, natural kinds are categories that are found in nature, such as various categories of plants and animals. For example, the categories "mammal," "gold," and "water" are all natural kinds.

One of the most distinctive characteristics of natural kinds is the remarkable richness of their correlated structure. According to Mill (1843),

Some classes have little or nothing in common to characterize them by, except precisely what is connoted by the name. (p. 135)

Consider the category of white things.

White things are not distinguished by any common properties except whiteness or if they are—it is only by such as are in some way dependent upon or connected with whiteness. . . . But hundreds of generations have not exhausted the common properties of animal or plants. . . . If anyone were to propose for the investigation the common properties of all things which are of the same color, the same shape, or the same specific gravity, the absurdity would be palpable. . . . Of these two classifications, the one answers to a much more radical distinction in the things themselves than the other does. One classification is made by nature, the other by us. (p. 136)

Thus, some categories allow us to infer a great deal about their members. A related point has been made about basic level categories,
which also have a correlated structure (Rosch et al. 1976); for example, if we know that something is a bird, we know that it has feathers and wings, that it is quite likely to have a beak and to fly, and so on. Mill's claim is that some categories have such an extraordinarily rich correlated structure that even after extensive study of the category, we have not exhausted all there is to learn about it. We have sciences devoted to discovering and understanding different animal species such as mammals. Although a category such as "mammal" may have originally been delimited by a few superficial features, such as having fur or hair, members of the category have unforeseen properties in common. We have continued to discover properties that are characteristic of mammals, so much so that separate scientific disciplines are needed to account for the accumulation of knowledge. The properties of mammals figure in scientific laws or generalizations from various sciences, including physiology, anatomy, genetics, and ethology. The fact that several scientific disciplines continue to discover new facts about this category is testimony to how richly structured it is.

Now, to take Mill's contrasting example, consider a category defined as "white things." If it seemed important or useful, we could certainly define such a category; yet we could not have a science devoted to determining what these diverse objects that happen to be white have in common. This category has a minimal correlated structure. It includes a white cloud but excludes a grey one. It includes a white car but excludes cars of any other color. It includes white paper but excludes yellow paper, and so on. Being white does not tell us very much about an object beyond the fact that it is white. Whiteness does not correlate with or predict much else.

The Causal Theory of Reference

The new theory of reference for natural kinds poses several problems for traditional theories of meaning. According to the traditional view (see chapters 1 and 3), each category term has both an intension (the set of necessary and sufficient properties that define the concept) and an extension (the set of objects to which the term applies). The intension specifies the criteria that determine which objects will make up the extension. That is, in order for an object to be in the extension of the term, it must satisfy the relevant criteria. To take a concrete example, consider the concept "square." The intension of the term "square" is an equilateral quadrangle with four right angles. In order for an object to be a square, it must fulfill these criteria; that is, it must have four sides, four right angles, and so on. The extension of the word "square" is the set of objects that meet these criteria. That is, the extension of the word "square" is composed of the objects that are squares.

One view of category terms that contrasts with this traditional view has already been discussed in chapter 3, under family resemblances. On this view, there may be no features that are necessary and sufficient; instead, the intension of the term consists of a set of features some number of which must apply to an object for it to qualify as an instance of the category. As long as an object has enough of the relevant features, it will qualify; it is not required to fulfill all of them.

What these two views have in common, however, is that they are both criterial accounts of category terms; that is, both claim that the meaning of a category term derives in some way or another from a set of criteria associated with that category. In this regard, the causal theory of reference contrasts with both the classical view and the family resemblance view of categories. In a major departure from the typical way of looking at category terms, Kripke (1971, 1972) and Putnam (1977) suggest that category terms be treated on analogy with proper names. This is a radical view, because on their treatment, at least, proper names do not have criterial meaning in the ordinary sense. (For a good summary of the issues, see Schwartz 1977.) As an example, consider the proper name "Shakespeare." We know a fair amount about Shakespeare—for example, that he was the author of Hamlet, Romeo and Juliet, and so on. Imagine a list of such facts about Shakespeare that we would use to identify him or describe him to someone who does not yet know who he is. It is possible, however, that we are mistaken about these facts. It could turn out that Shakespeare never really wrote those plays and that our beliefs about him were therefore erroneous. Nevertheless, the name "Shakespeare" would still apply to the same person. We wouldn't decide that we were wrong and that that person isn't really Shakespeare. We would conclude instead that Shakespeare did not write those plays. Thus, the facts or descriptions that we might use to identify Shakespeare do not function as criteria. They are not necessary and sufficient features of what it takes to be Shakespeare. They are beliefs about Shakespeare that are helpful in identifying him, but they do not constitute criteria in the traditional sense.

Kripke argues that instead of being based on criteria, the "meaning" of a proper name derives from the causal history of its use, going back to its first introduction, on analogy with a baptismal ceremony. Moreover, he claims, proper names refer to the same individual, even in counterfactual situations. When we ask, "If Johnson had not declined to run again, would Nixon have been elected president?"
there is no question that we are still referring to Nixon, even though he might not have been president. We would still be using the proper name “Nixon” to refer to Nixon. The name refers to the same individual even when used in hypothetical discussions where important facts about the world and about the individual in question are changed.

Analysis of Natural Kind Terms

According to Kripke and Putnam, there are important analogies in the way that natural kind terms and proper names function. First, Kripke and Putnam argue that many of the properties that we may have taken to define a category term do not really do so. The claim is that the superficial properties that we use to identify natural kind categories do not function as necessary and sufficient criteria and that in fact each such property could be violated and yet we would still agree on the classification of the object. Putnam uses the existence of abnormal members of categories to demonstrate this point. As a way of describing the natural kind term “lemon,” one might specify that lemons have an oval shape, yellow skin, and a sour taste. Yet one could imagine a change in some chemical in the atmosphere that would modify the pigment in the skin, causing lemons to become green. Similarly, a change in the nutrients in the soil could effect a change in the taste of lemons, causing them to be bitter, and so on. Yet, Putnam argues, despite all of these changes in the properties we have listed, we would still agree that these (green, bitter) objects are lemons. That is, the word “lemon” would still refer to lemons despite the changes that we would need to make in the description of the properties. To further illustrate this point, Putnam notes that what we currently take to be normal members of the category might, in fact, be abnormal members. Perhaps lemons were originally green and bitter but alterations in the chemical composition of the atmosphere and land have caused them to be yellow and sour.

On this causal theory of reference, the descriptions that we give for natural kinds function just like the descriptions we might give for a proper name. They are useful in identifying the objects in question, and they may effectively describe the stereotype we have of the object, but they do not qualify as criteria for the category. Even if the superficial properties change, the natural kind term will continue to refer to the same category.

Given this rejection of the traditional account of meaning, two questions naturally arise: How is the referent for a term established? and How can we tell whether or not the term is being used cor-

rectly—for example, how do we know that an object we call a “lemon” is in fact a lemon?

According to this theory, the reference of a term may be established by a causal chain, by analogy to the way the reference for proper names is established. While the object is present, someone labels the object, and that provides the first link in a causal chain in which subsequent speakers continue to use the label to refer to the same kind of object. For natural kind terms, speakers will typically select a prototypical exemplar to label at first, and they may provide a description of the natural kind to identify the relevant category. The description does not define the category, however, but serves as a way of helping speakers to fix the referent, that is, to identify what is being labeled. Tigers might be described as large, striped, ferocious, wild cats to enable someone to identify what objects are being referred to as “tigers.” Yet under some circumstances we might readily agree that a small, albino, tame cat was, nevertheless, a tiger. First, Putnam points out that there is a “division of linguistic labor.” Not everyone need acquire the methods for determining whether or not something is gold, for example. Instead, we may often rely on experts to make the final determination. But what do the experts rely on? Experts base their judgments on the best scientific knowledge that is available at the time—on the most well established empirical theory. In the case of gold, its atomic number will be the deciding factor; in the case of water, its molecular formula (H₂O); and in the case of lemons, their chromosomal structure. Here Putnam embraces a kind of essentialism. The assumption is that there are some “deep” properties, or hidden structural properties, that account for or determine what the more superficial properties will be. These structural properties establish the “essential nature which the thing shares with other members of the natural kind. What the essential nature is is not a matter of language analysis but of scientific theory construction” (Putnam 1977, 104).

Contrast this analysis of natural kind terms with a standard analysis of more arbitrary “one-criterion” terms, such as Mill’s example of “white things.” Here the properties we would give to describe the meaning of the term would function as a definition and not just as a useful aid to identifying the category. If the properties changed—for example, if pollution turned all white things a dingy grey—they would no longer be white things. We do not have a scientific theory of white things to rely on, nor do we expect there to be some deep, hidden structure that is common to all white things.

Each of these points—the unlimited richness of the categories, the search for more theory-relevant explanatory properties, the reliance
on authority to distinguish exemplars of a category from nonexemplars, the acceptance of abnormal members, and the corrugibility of beliefs about categories—distinguishes natural kinds from other types of categories.

**Natural Kinds and Induction**

Susan Gelman and I (Gelman and Markman 1986, 1987) have argued that this analysis of natural kinds, especially the emphasis on the richness of the structure and the belief that unobservable properties are common to members of a natural kind, predicts that natural kinds will often be used to support inductive inferences from one category member to another. That is, if categories are structured so as to capture indefinitely rich clusters of information, then new features learned about one category member will often be projected onto other category members as well. In this way, natural kind categories should promote inductive inferences. Moreover, there are two ways in which the inductions are made without perceptual support. First, even if an object does not look much like other members of a given natural kind, knowing what kind it belongs to should lead people to assume that it will share relevant properties with other members of the category. Second, these properties, such as internal organs or chemical structure, are often unobservable by the average person.

Certainly, only some types of inferences within a kind are justified. Whether the inference is reasonable or not depends in part on what type of property is attributed to what type of natural kind. Among animal species, for example, we expect members of the same species to share methods of reproduction, respiration, and locomotion. We do not expect other kinds of properties to be common even to members of the same species. For example, if one poodle is 2 years old, we should not expect another poodle also to be 2 years old. Thus, at least implicitly, people have embedded natural kind categories in scientific or prescientific theories that limit what classes of properties are expected to be common to a given natural kind category.

Little is known about how expectations about natural kinds originate. How much exploration of categories or even explicit scientific training is needed before children come to expect that categories reflect more than superficial perceptual similarities? There is a large developmental literature suggesting that young children rely on superficial perceptual properties on cognitive tasks, including those involving classification, free recall, free association, and word definitions (see Flavell 1963, 1977; Mansfield 1977). Young children have often been characterized as “concrete” and as “perceptually bound,” meaning that their cognition is captured by the appearances of things. A well-known example is the Plagian conservation problems. In a task involving conservation of number, for example, two equal rows of objects—say, pennies—are lined up in one-to-one correspondence. Children judge that both rows have the same number of pennies. Then, while the child watches, one of the rows is spread out. Children now judge that the lengthened row has more pennies. One interpretation of this is that the children are unable to overlook the misleading perceptual cue of the length of the row. Their judgment of equality or inequality is presumably based on the available perceptual information rather than on the actual quantity. To take an example concerning categorization, Tversky (1985) has found that young children prefer to group objects together on the basis of color or shape rather than on the basis of common category membership. On one of her tasks, for example, a 4-year-old typically groups a fire engine with an apple because both are red rather than grouping a fire engine with a car because both are vehicles. In this task and others, children seem unable to override what are sometimes misleading perceptual cues.

Based on these findings, one might expect that young children would rely heavily on perceptual characteristics of objects for judgments of category membership. Young children may have no means of appreciating the rationale for grouping perceptually dissimilar objects together. Even for natural kinds, children might represent category members as sharing superficial properties and only later come to realize that they have deeper properties in common. Thus, according to this view, children, with their reliance on perceptual features and their limited scientific knowledge, should not rely on natural kind categories to support inductive inferences about objects.

On the other hand, given the importance of natural kind categories for human cognition, children might quite early on expect categories to have a richly correlated structure. Even with only rudimentary scientific knowledge, children might believe that natural kind categories are united by many unobservable properties. Children could be biased from the start to expect that categories they learn will share clusters of features, or such an expectation could be derived from experience. Even with only limited scientific knowledge, children could notice that natural kinds have many observable features in common. They could then extend this belief and expect natural kinds to be united by many unobservable properties as well. Any appreciation of natural kinds at this early age would probably reflect an unsophisticated, undifferentiated belief in the richness of categories. Children would lack the requisite scientific knowledge that
could limit their inductions to categories and attributes that are appropriate.

One piece of evidence that children are not solely dependent on perceptual similarity for drawing inferences comes from work by Carey (1985). In one study she showed several groups of children between ages 4 and 10 a mechanical monkey, one that could move its arms to bang cymbals together. The mechanical monkey looked much like a real monkey. Children, who knew that real monkeys breathe, eat, and have baby monkeys, were asked whether the mechanical monkey could breathe, eat, and have babies too. All but one group of 4-year-old children denied that the mechanical monkey possessed these animate properties. In other words, despite the striking perceptual similarity of these two types of objects—mechanical and real monkeys—children did not generalize facts about one to the other. These children had differentiated living things from nonliving things and therefore refused to impute properties that characterize living things to nonliving things. It could be, however, that it is only at the level of such basic ontological distinctions (Kell 1979) such as living versus nonliving that children treat categories as natural kinds.

Gelman and I (Gelman and Markman 1986) questioned children about much more specific natural kind categories, from both biological (e.g., squirrel and snake) and nonbiological (e.g., gold and salt) domains. To determine whether children would induce new information from natural kind categories, rather than from perceptual appearances, we pitted category membership against perceptual similarity. Children were shown two objects and told a new fact about each. They then had to infer which of the facts applied to a third object that looked very much like one of the first two objects but was given the same category label as the other one. These experiments test whether very young children are sensitive to the richness of natural kind categories and whether they use these categories, in the absence of perceptual support, to justify inductive inferences.

One of the main assumptions about natural kind categories that motivated our developmental work is that adults expect members of a natural kind to share many properties and will therefore use the natural kind category to support inductive inferences. In particular, adults should rely on the natural kind membership of an object more than on its superficial perceptual appearance to make inferences about its internal structure, behavior, and other theoretically relevant properties. Gelman and I conducted a preliminary study (reported in Gelman and Markman 1986) to establish that adults would use the natural kind category to support inductions for those categories that would later be used with children.

Undergraduates were presented with 20 problems. Each problem consisted of a set of three pictures. New information was given about two of the pictures, then a question was asked about the third picture. The correct answer could not be determined from the picture. The subject could arrive at one of two answers: either by making an inference based on category membership or by making a different inference based on perceptual similarity.

The three pictures for each problem were arranged on a card. Two pictures were at the top of the card, and the third (the target) was directly underneath, centered below the first two. Directly below each of the two topmost pictures was a sentence that labeled the picture and provided some new factual information about it. Directly below the target picture was a question asking which of two new attributes applied. For example, on one problem subjects saw a flamingo and a bat at the top of the card. Underneath the flamingo was written, "This bird's heart has a right aortic arch only." Underneath the bat was written, "This bat's heart has a left aortic arch only." Below these two pictures was a picture of a blackbird (which looked more like the bat than the flamingo). Underneath the blackbird was written, "What does this bird's heart have?" For each item, subjects were to choose one of the two answers. After each choice they were asked to justify their selection by responding to the question, "Why is this the best answer?" Finally, they were asked to rate their confidence in their answer on a scale from "1" (very unsure) to "7" (very sure).

As predicted, adults based their inferences on the common natural kind membership of the objects. Overall, they concluded that the target picture had the same property as the other similarly labeled object an average of 86% of the time. Apparently subjects expected slightly more variation within a category for the biological categories than for the nonbiological categories. They inferred properties on the basis of common category 92% of the time for the nonbiological categories and 80% of the time for the biological categories. In addition, subjects were highly confident that their choices were correct (the mean rating was 5.8 on the 7-point scale). However, they were significantly more confident about their judgments for the nonbiological categories (mean = 6.0) than biological categories (mean = 5.5).

Young children might, like the adults, infer a new property of an object from its category, or they might instead be governed by the perceptual appearances of the objects. In study 1 from Gelman and Markman 1986 children were tested on the same categories as the adult subjects in the preliminary study. They were asked about dif-
ferent attributes, however, because they would not understand most of the ones that adults were questioned about.

Preschool children ranging in age from 4:0 to 4:11 (with a mean age of 4:5) participated in the study.

There were three conditions in this study. In the experimental condition children were taught information about each of two objects. They were then shown a third object that looked like one of the two training objects but was given the same category label as the other. Children were asked to infer which piece of information applied to the third object. This condition was designed to reveal whether children's influences are influenced by their knowledge of an object's category or by perceptual similarity.

A second condition, the no-conflict control, was designed to demonstrate that when perceptual similarity and category membership coincide, children readily draw the correct inferences. In this condition, like the first, children were taught properties of two objects. However, the third object not only looked like one of the training objects but also was given the same label as that object. This task provides a baseline measure for how often children will draw the correct inference when both perceptual similarity and category membership lead to the same conclusion.

A final condition, the attributes control, was designed to make certain that children did not already know the information we would be teaching them. Children saw only one picture at a time—the third item in the other two conditions—and were asked which of the two properties applied. Children were expected to perform at about chance level in this condition.

Since the experiment proper involved teaching children new properties about two objects and asking them to judge which of the properties applied to a third object, it was important to ensure that children did not already know the information to be presented. A preliminary study was conducted to select questions about various animals, plants, and substances to which children would not yet know the answers.

An item was selected for inclusion in the experiment proper only if children were unable to answer the question significantly above or below chance. Table 5.1 shows examples of items that were selected. Half of the items concerned biological natural kinds and half concerned nonbiological natural kinds.

In the experimental condition, children saw 20 sets of three pictures each. Information was given concerning two of the pictures in each set, and children were asked a question about the third picture. The third picture looked like one of the first two pictures but was given the same category label as the other. Children could answer on the basis of either perceptual similarity or category membership. For example, children were shown a tropical fish, a dolphin, and a shark. In this case the shark was perceptually similar to the dolphin but was given the same label as the tropical fish. The experimenter labeled the three pictures, “fish” for the tropical fish, “dolphin” for the dolphin, and “fish” for the shark. Children were asked to repeat the names until they could name all three pictures correctly. (On 88% of the trials, children repeated all three names correctly on their first try.) The experimenter then pointed to the tropical fish and said, “This fish stays underwater to breathe.” She pointed to the dolphin and said, “This dolphin pops above the water to breathe.” Finally she pointed to the shark and said, “See this fish. Does it breathe underwater, like this fish, or does it pop above the water to breathe, like this dolphin?” Comparable questions were asked for each item in table 5.1. Figure 5.1 presents an example of another triad: bird, bat, bird.

In the no-conflict control condition, the picture triads and procedure used were identical to those used in the experimental condition, with one exception. In this case the labels of the two similar objects were made to agree rather than conflict. For example, children heard the tropical fish labeled “fish,” the dolphin labeled “dolphin,” and the third picture (shark) labeled “dolphin” (instead of “fish” as in the

<table>
<thead>
<tr>
<th>Table 5.1</th>
<th>Sample items and attributes used in Gelman and Markman 1986</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological items</strong></td>
<td><strong>Nonbiological items</strong></td>
</tr>
<tr>
<td>This squirrel eats bugs. (gray squirrel)*</td>
<td>If you put this gold in a hot oven, it melts. (gold bar)</td>
</tr>
<tr>
<td>This rabbit eats grass. (brown rabbit)</td>
<td>If you put this clay in a hot oven, it burns. (reddish blob)</td>
</tr>
<tr>
<td>(Target): squirrel (Kaibab, looks like rabbit)</td>
<td>(Target): gold (brown blob; looks like clay)</td>
</tr>
<tr>
<td>This dinosaur has cold blood. (brontosaurus)</td>
<td>This pearl comes from inside a sea animal. (seed pearl)</td>
</tr>
<tr>
<td>This rhinoceros has warm blood. (gray rhinoceros)</td>
<td>This marble comes from a big piece of rock. (round, pink)</td>
</tr>
<tr>
<td>(Target): dinosaur (triceratops, look like rhinoceros)</td>
<td>(Target): pearl (round, pink; looks like marble)</td>
</tr>
</tbody>
</table>

*Descriptions of the objects used are given in parentheses. These descriptions were not mentioned to subjects.
previous condition). The children in this condition were provided with the same information about the two initial pictures—that is, that the fish breathes underwater and the dolphin pops above the water to breathe. The experimenter then pointed to the third picture and said, “See this dolphin. Does it breathe underwater, like this fish, or does it pop above the water to breathe, like this dolphin?”

In the attributes control condition children viewed only one picture at a time—the third item from each of the triads in the experimental condition. Without hearing any prior information, children were asked to judge which of the two properties applied. For example, children were shown the picture of the shark and asked, “See this fish. Does this fish breathe underwater or does it pop above the water to breathe?”

The main question that this study was designed to address is whether preschool children are willing to infer properties of an object based on its natural kind category. In particular, when category membership and perceptual similarity are in conflict, will children show any sensitivity to category membership, or will their inferences be based on the appearance of the objects? If children do use category membership as a basis for induction even in the absence of perceptual support, it is likely that they will do so for some conceptual domains before others. Therefore, we included both biological and nonbiological natural kinds, to test for generality. Table 5.2 presents the data according to condition and item type.

To address these questions, we first needed to establish that children’s inferences in the experimental condition were based on the information provided to them in the experiment rather than on preexisting knowledge. The results of the attributes control condition indicate that children were in fact unaware of the correct answers. When simply given the test question, with no extra information to guide their answer, children performed at chance level, answering a mean of 53% of the questions correctly.

On the other hand, when the perceptual similarity and category label coincided in the no-conflict control condition, children were capable of drawing the correct inference. When both the label and the appearance of the object led to the same conclusion, children were correct on 88% of the items.

In the experimental condition, where perceptual similarity and category membership were opposed, children preferred to use the category information 68% of the time, which is significantly better than chance, p < .001. Thirty-seven percent of the children consistently based their judgments on common category membership; that is, they inferred the property of the new object based on category membership on at least 15 out of 20 items. In contrast, no child showed a consistent preference for basing inferences on the perceptual similarity of the objects.

Thus, the children in the experimental condition were taking account of the training information in deciding about the properties of the new objects, as performance in that condition was better than in the attributes control condition. Although children in the experimental condition often based their judgments on the natural kind category of the object, they were in some conflict because of the divergence between category membership and perceptual appearances. Also, giving the category name may not be a perfect way of establishing category membership. That is, this procedure tests the power of the category in governing inductive inferences only insofar as the label successfully conveys the natural kind category. Some failure to respond to the category may reflect some degree of weakness in this for children, and not necessarily weakness in natural kind in-

| Table 5.2 |
| Percent correct (category choices) from Gelman and Markman 1986 |

<table>
<thead>
<tr>
<th></th>
<th>Experimental condition</th>
<th>No-conflict control condition</th>
<th>Attributes control condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological</td>
<td>67</td>
<td>89</td>
<td>59</td>
</tr>
<tr>
<td>Nonbiological</td>
<td>69</td>
<td>87</td>
<td>48</td>
</tr>
</tbody>
</table>
ductions per se. When category labels and perceptual similarity coincided (the no-conflict control condition), children answered more questions correctly.

Children were able to use category membership as often for the biological as for the nonbiological categories. Overall the mean correct was 67% for the biological and 69% for the nonbiological categories. However, there was a sex difference in the experimental condition. Boys performed better on the biological categories than on the nonbiological categories, whereas girls showed the reverse pattern. We could not explain this difference, although it is possible that boys and girls differ in which categories they find more familiar. Boys may be more familiar with such things as snakes, worms, bugs, and leaves (that is, the biological categories we tested), whereas girls may be more familiar with such things as sugar, salt, diamonds, and pearls (the nonbiological categories we tested).

In summary, even 4-year-olds realize that natural kind categories such as “squirrel” and “diamond” promote a rich set of inductive inferences. These young children have already come to expect new knowledge to be organized in accord with the categories named by their language, even in the stringent test case where the label conflicts with perceptual appearances. Using a simplified procedure, Gelman and Markman (1987), determined that this conclusion extends to 3-year-olds as well. For these young children, the procedure used in Gelman and Markman 1986 was too demanding. Instead of pitting objects against each other, we taught children a new fact about an object (as before) and then determined which objects it would generalize to: an object that (a) looked like the original, (b) had the same label as the original, (c) looked like the original and had the same label, or (d) differed from the original in both respects. The findings were that children drew more inferences based on category membership than inferences based on perceptual appearances. Thus, even 3-year-olds assume that categories named by their language will include more than superficial features.

Study 3 of Gelman and Markman 1986 addressed two questions raised by this finding. One question was whether identity of the linguistic information is necessary for children to use the common category as a basis for inductive inferences, or whether other means of indicating common category membership would be sufficient. To address this question, we designed a synonyms condition in which category membership was conveyed by means of synonyms rather than identical labels. For example, one triplet consisted of a target rabbit, another rabbit with a different appearance, and a squirrel that had long ears and looked like the target rabbit. The two rabbits in the synonyms condition were called “rabbit” and “bunny.” If children infer that objects named by synonyms share the same properties, then their inferences cannot be based simply on identity of the labels.

Another question that was addressed in study 3 of Gelman and Markman 1986 was whether children would use the common category to make arbitrary decisions for which common category membership is not relevant. To test this, we designed an arbitrary decision condition in which children were asked to decide what color chip should go on a picture after witnessing the experimenter place a chip of one color on a perceptually similar picture and a chip of another color on a dissimilar picture with the same category label. For example, the experimenter placed a red dot on the picture of one rabbit (called “rabbit”) and a yellow dot on the picture of the squirrel (called “squirrel”). The child then had to pick which color dot to put on the third picture, the other rabbit (called “rabbit”) that looked like the squirrel. If children are distinguishing between the induction task, where category labels are relevant, and this arbitrary task, where they are not, then they should be at chance in this condition, having no real basis on which to make a decision.

The third condition in this study was the standard condition that replicated the procedure of the experimental condition in the first study using the items that appeared in the synonyms and arbitrary decision conditions.

In the standard condition children once again drew inferences to category members at an above-chance level (68%). As before, even when the category conflicted with perceptual appearances, children tended to base their inductive inferences on the common category membership. Moreover, as the results from the synonyms condition indicated, children do not need to hear common labels to use common category membership to draw inferences. When children heard synonyms to indicate common category membership, they still based their inferences on category membership at a level greater than chance (63% of the time). The results of this study further indicate that children have begun to differentiate between inferences where category membership is relevant and arbitrary decisions where it is not. When children were asked about arbitrary decisions such as what color chip should go on a given picture, they were no more likely to base that decision on the color chip they had seen placed on a common category member than to base it on a chip placed on a perceptually similar picture.

In summary, children's inferences were based on the category membership of the objects and not simply on how the pictures were labeled. Children drew inferences based on common category mem-
bership even when category members were not given identical labels. Moreover, when category members were given identical labels, children relied on these labels only when the task required them to draw inferences.

From these studies, we know that young children will infer that an object shares properties with another object from the same category, even when these inferences do not have perceptual support. With the exception of the completely arbitrary property, the properties that were examined were ones that in fact would be largely determined by the natural kind category of the objects. That is, we asked about the internal organs, method of respiration, feeding habits, behavior, and so on, of animals—all properties that typically are common to a species. Similarly, we asked about the internal structure and chemical and physical properties of the nonbiological categories—again, all properties that typically are common to a mineral, metal, or other substance.

There are cases, however, where inductive inferences based on category membership would be unwarranted. For some properties, the perceptual similarity of objects should be used as the basis for induction rather than the common category membership. The size of an object, for example, is a better predictor of its weight than is its category.

Study 4 of Gelman and Markman 1986 examined whether young children are selective in the kinds of inferences that they make based on category judgments. Preschool children have impoverished scientific knowledge, yet they believe that category members share unobservable properties. This implies that their beliefs about categories may be fairly general ones, not modulated by specific knowledge. In other words, such young children may not have sorted out which properties legitimately do and do not promote inferences within natural kind categories. They might, then, erroneously infer information on the basis of common category membership, even when asked about attributes that are more likely to be consequences of superficial perceptual properties. If attention to category membership dominates their judgments regardless of type of property, then children should assert, for example, that a rocklike chunk of salt will blow away in the wind as does fine-grained salt, and not remain in place as does a rock.

In this experiment children were taught a fact about each of two objects and then asked which of the properties applied to a third object. The third object was from the same category as one of the objects but was more similar in appearance to the other object. Unlike the previous studies, however, the properties were predicted more by perceptual similarity than by common category membership.

This study had two conditions, an experimental condition and a control condition. Children in the control condition were given the final test questions without any extra information. For example, they were shown a shark and were asked, "This is a fish. Do you think it weighs 20 pounds or 100 pounds?"

The procedure for the experimental condition was identical to that of the earlier studies, except that different attributes were being taught. Children viewed three objects, two placed side by side and one centered beneath them. Each object was labeled. A property was attributed to each of the two topmost pictures, and the children were then asked which of the properties applied to the third (target) picture. For example, after the pictures were labeled, children were told about a tropical fish, "This fish weighs 20 pounds." They were told about an object from a different category, in this case a dolphin, that, "This dolphin weighs 100 pounds." They were then asked about the target picture, a shark that looked more like the dolphin than the tropical fish, "See this fish. Does it weigh 20 pounds, like this fish, or 100 pounds, like this dolphin?"

In this study children were asked questions about perceptually based attributes: weight, visibility at night, and so forth. With no prior information on which to make inferences, children in the control condition selected the category choice no more often than expected by chance (48% of the time). In contrast to what happened in the earlier studies, fewer children in the experimental condition based their inferences on common category membership. They selected on the basis of common category an average of 49% of the time, which is not significantly different from chance or from the control group. However, the experimental condition, unlike the control condition, was markedly trimodal. Four of the 20 children consistently selected on the basis of perceptual similarity (that is, they chose the attribute of the perceptually similar picture on at least 15 out of 20 items); and 4 children consistently selected on the basis of common category membership even though that choice was unwarranted. The remaining 12 children did not seem to know which answer to choose, with a mean of 47% categorical choices. This condition contrasts with the control condition, in which none of the 20 children had a consistent preference.

Children who based their inferences predominantly on common category membership were overgeneralizing the importance of the category label. It was unwarranted to assume, for example, that a legless lizard can run as quickly as a four-legged lizard, or that a large
pearl weighs as much as a smaller pearl. These children understood the importance of category names to promote induction but were not yet selective in their inferences.

Overall, however, children in this study who were asked about these perceptually based properties relied less on the category than did children in the earlier studies. Many of the children were sensitive to the differences between attributes such as weight that are consequences of perceptual properties and attributes such as means of respiration that are common to the members of a species. Of all the studies presented in Gelman and Markman 1986, this was the only one in which subjects chose predominantly on the basis of shared appearance. In this study of attributes derived from perceptual information, 20% of the children consistently based their judgments on perceptual similarity compared to only 1% of the children asked about attributes common to natural kind categories.

Summary of the Findings from Gelman and Markman 1986, 1987

On each of a series of problems, children had to decide whether a given object possessed one or the other of two attributes. On hearing the question alone, with no prior information to guide them, children had no basis for forming an induction and, as expected, simply guessed at which property applied. In two other conditions children were first told which attribute applied to each of two training objects. In the simple case where one of the training objects matched the target object in both appearance and natural kind category, 4-year-old children almost always drew the appropriate inference. Their performance was excellent, and establishes that the simple inferential problem is well within the capacity of young children.

The most informative condition was the one in which perceptual appearance and the natural kind category of an object led to divergent conclusions. Children relied on the shared category to promote inductions, even in this stringent case where perceptual similarity would lead to a different conclusion. Moreover, children’s inferences were based on common category membership and not just on identity of labels. When members of the same category were given synonymous labels children still preferred to draw inferences within the category, at an above-chance level.

These results are at odds with a widely held view that children’s thinking is strongly influenced by perceptual appearances. On this view, if young children draw inductions at all, they should rely more on perceptual characteristics of objects than on their category membership. Several of our findings suggest that children are not domi-

nated by appearances either in their conception of the structure of categories or in their use of categories to support inductions.

First, most children accepted our label for the third object even though it looked more like a member of the other category. For example, one object was a squirrel with very long, rabbitlike ears. Overall, it looked more like a typical rabbit than a squirrel. Some children noted the discrepancy (remarking, for instance, “funny rabbit”) and some even mildly objected to the label, but, for the most part, they accepted the labels for these abnormal category members. This finding is consistent with work by Flavell, Flavell, and Green (1983) on the development of the appearance-reality distinction. Flavell, Flavell, and Green showed 3- to 5-year-olds fake objects (for example, a sponge that looked like a rock) and then let them feel the objects to discover their true identity. On later questioning, children tended to say not only that the object was a sponge, rather than a rock, but also that it looked like a sponge. Like our subjects, these children accepted the category label even in the face of discrepant appearance.

Second, only 1 of 69 children in this condition consistently generalized properties on the basis of perceptual similarity between objects. Even though the rabbit and the rabbit-eared squirrel looked very much alike, children did not assume that they both ate grass. Whatever perceptual biases children have, they are overridden by the belief that members of a natural kind share many other properties as well.

Third, in each of the studies (studies 1-3 of Gelman and Markman 1986) children in the standard condition reliably used the category of the object to support inductive inferences, even when this conflicted with the appearance of the objects.

Children drew these inferences when asked about properties that were reasonable to project from one category member to another. We asked about the eating habits, means of respiration, and internal organs of the biological categories and about chemical and physical properties of the nonbiological categories. Young children, able to answer these questions correctly, might overgeneralize the importance of the natural kind category. That is, even when it is unwarranted to do so, they may use categories more than appearances to support inductions. The results of subsequent studies indicate ways in which children have begun to limit the importance of the category. First, when the task involves making an arbitrary decision, children are not biased to infer on the basis of category membership. Second, children have at least begun to distinguish some kinds of properties from others as a basis for induction. To test this, we asked children about properties that should generalize on the basis of superficial percep-
tual similarity rather than on category membership. In answering such questions, children did not reliably use the category to support the inductions. Across several studies, those children who were asked about perceptually based properties were the only ones to reliably use the perceptual appearances of objects to support their inductions.

By age 3 and 4 children expect natural kinds to have a richly correlated structure that goes beyond superficial appearances. They use category membership to support inductions, even in the stringent test case where perceptual appearance and category label lead to different conclusions. Moreover, children have begun to differentiate between the kinds of properties that can justifiably be projected to other category members and those that cannot. Despite all these accomplishments, however, there is much left for children to learn about natural kind categories.

The Problem of Determining Which Properties Support Inductive Inferences

First, children must sort out which properties are likely to be common to members of different types of natural kinds. Although they have begun to work out this problem by age 4, their distinctions on even a crude level are imperfect. Even with properties that for adults are blatantly determined by superficial perceptual features (such as weight), some children based their inductions on the category—claiming, for example, that a large fish weighs the same amount as a little fish, because both are fish. Furthermore, there are constraints on inferences that no one has yet been able to characterize (Goodman 1955). Even for adults, we do not have good theories to explain how inferences are constrained. 1 A predicate may or may not promote inductions, depending on the level of abstraction of a category (all dogs bark, but not all animals bark) or the scientific domain (density at room temperature is important for metals but not animals). Nisbett, Krantz, and Kunda (1983) have shown that adults are quite willing to infer new information from one category member to other members of the same category, but they do so selectively, depending on the property involved. Our studies have not tested the limits of children’s abilities, but even so it is clear that children must develop more finely tuned distinctions among predicates and learn not to overgeneralize to obviously inappropriate predicates.

The Problem of Determining Which Categories Support Inductive Inferences

A related issue is how to constrain which kinds of categories support inductions. Some categories, such as “artifacts,” do not pick out objects in nature that have indefinitely many properties in common. We do not assume, for example, that all forks or all saws will have unlimited numbers of properties in common. It is possible that children could very early on notice the natural kind–artifact distinction and use categories to support inductions mainly for natural kinds. It is also possible that children would begin by expecting most categories named by language to promote inductive inferences. That is, they would assume for a variety of conceptual domains that category members share many features with each other. Only after learning more about various domains would they restrict their inferences.

To address this question, Gelman (1984) draws several distinctions between types of concepts. Natural kinds are expected to share many features besides the obvious ones, whereas artifacts are not. This is not a strict dichotomy, however, as some complex artifacts (computers and cars, for example) are probably similar to natural kinds. Building on findings from Rips (1975) and Nisbett, Krantz, and Kunda (1983), Gelman argues that in addition to the complexity of the category, assumptions about its homogeneity also may well affect how likely it is to support inductive inferences. For example, adults will infer that other samples of a metal conduct electricity, given that one sample of the metal does, but they will not infer that other people within a geographical region are obese, given that one person in that region is (Nisbett, Krantz, and Kunda 1983). Adults rely on their conception of how variable the property is within the domain being questioned (for instance, conductivity for metals) to make their inferences. Gelman suggests that natural kinds tend to be more homogeneous than artifacts and therefore support more inferences. Another possibility, however, is that it is homogeneity of the category per se, rather than its natural kind status, that predicts which categories are more likely to support certain inductions.

In a series of studies Gelman had adults rate categories in various ways to determine whether they perceive natural kinds and artifacts as differing in homogeneity. One procedure was to ask subjects to predict what percentage of the category members would be expected to have a given characteristic. Another was to ask subjects directly to

1. Sternberg (1982) has found that familiarity and complexity of a predicate affect how quickly adults can process it. However, familiarity and complexity alone cannot characterize the sorts of inferences we draw. Certainly, some simple and familiar predicates (say, "has a fever" or "is 3 days old") are not projectable. For example, just because one poodle is 3 days old, we do not expect the next poodle we see also to be 3 days old. Hence it is not clear how to extend Sternberg's findings to the problem of what confirms some inductions and not others.
rate on a scale from 1 to 9 how similar the members of a given category were to each other. The results using both of these procedures were comparable, indicating that natural kinds on the whole tend to be seen as more homogeneous than artifacts. Although on some of the measures minerals were seen as less homogeneous, natural kinds were in general thought of as homogeneous, regardless of complexity. This difference held up for superordinate level categories, as well as for basic level categories. The main exception was that, as expected, complex artifacts such as machines were perceived as being more like natural kinds.

Gelman then went on to determine what kinds of distinctions children have made between categories. Four- and 7-year-old children were taught a new fact—for example, that a rabbit has a spleen—and then had to decide whether the fact applied to a similar object (a similar rabbit), a different object from the same category (another rabbit), an object from the same superordinate category (a dog), and an unrelated object (a telephone). As evidence that the children were taking the task seriously, Gelman found that children virtually always drew inferences to a similar looking object and rarely drew inferences to an unrelated object. Only for the two intermediate levels of generality does it make sense to ask whether there is a natural kind-artifact distinction. Seven-year-olds clearly distinguish between natural kinds and artifacts. They drew more inferences from natural kinds than from artifacts at both basic and superordinate levels of categorization. The results from 4-year-olds are not so clear. These children may have begun to draw such a distinction, but it is unstable. They are more likely to draw inferences from categories that adults perceive as homogeneous. Thus, Gelman argues that the natural kind-artifact distinction is not used by preschool children but that it may evolve from an earlier distinction based on homogeneity.

Another way in which young children are likely to be limited is that they may not be able to look much beyond perceptual features of objects when they form categories on their own. In the studies reported here children were told the category labels and then asked to infer information from one member to another. This task is simpler than the converse problem of having to form the category in the first place, without knowing beforehand which properties are relevant. When initially forming a category, children are likely to be much more susceptible to perceptual appearances. Most standard classification procedures (see Inhelder and Piaget 1964) require children to divide objects into categories where many bases of classification are possible. Given the complexity of this problem (see Gelman and Bailargeon 1983; Markman and Callanan 1983), children often find the superficial perceptual appearances of objects to be an easier basis on which to organize the material.

As for natural kinds, Keil (in press) found young children to be more dependent on perceptual similarity than on deeper biological properties when they are asked to classify anomalous objects. He asked children to classify artifacts and natural kind objects, given conflicting information. For example, one object looked exactly like a skunk but its biological structures (heart, bones) and lineage (parents) were supposed to be that of a raccoon. The youngest children believed for both natural kinds and artifacts that appearance determined category membership. For example, they would say that the animal that looked like a skunk was in fact a skunk, even though they had been told it had a raccoon heart, gave birth to raccoon babies, and so forth. Not until about second to fourth grade were children willing to say that internal structure was an important criterion for categorizing natural kind objects. But these children probably have no way of knowing whether internal structure or external structure is more important. Another way of stating the difference between Keil's task and ours is that children were asked to make different sorts of comparisons in the two studies. On Keil's task children had to compare two different kinds of attributes: perceptual appearance versus biological properties. On our task children had to compare attributes (perceptual appearance) to membership in a category. The category label may be considered a more powerful source of information than a few biological properties such as having a particular heart or set of parents. Gelman, Colman, and Maccoby (1986) found that inferring properties on the basis of categories was easier than inferring category membership on the basis of properties for one category: namely, gender.

**Primitive Theories as Possible Constraints on Induction**

Finally, as Carey (1985) has argued, children must learn how natural kind categories are related to one another in a system of theory-based knowledge. Carey has found that children initially organize biological knowledge around humans as a prototype. Inferences about the biological properties of other species are based both on what children believe about humans and on how similar the other species is to humans. For example, children in Carey's study were taught that humans have an omentum and were asked whether various animals and artifacts also have an omentum. Children generalized in accord with a rough similarity gradient from most to least similar to humans. What is most striking about Carey's findings, however, is how depen-
dent children are on the category of humans to organize new biological knowledge and to trigger inferences. Children draw inferences about biological categories primarily when the property is known to be true of humans. They are more likely to infer that a biological property of *humans* will generalize to bugs than they are to infer that a property of *bees* will generalize to bugs, despite the far greater similarity of bugs to bees than to humans. This marked dependence on humans as the prototype changes with age: adults generalize from one species to another based on how similar the two species are to each other (Rips 1975).

Carey discusses this work in terms of developmental changes in the scientific theories in which these natural kind terms are embedded. One of the roles of scientific theory is to constrain the kinds of inductive inferences that are made. The marked asymmetries in projection found at age 4 (that 4-year-olds generalize more from human to bug than from bug to bee, for example) disappear by about age 10. This change reflects a major restructuring in the organization of children's biological knowledge. The biological knowledge of 4-year-olds is focused on humans. Biological properties are fundamentally properties of humans and only secondarily properties of animals. New biological knowledge must be related to humans in order for 4-year-olds to project the properties to other animals. By age 10 the special status of humans as biological creatures has diminished. Now humans are only one of many mammals as far as biological properties are concerned.

**Conclusions**

At age 4, then, children still have much to learn about natural kind categories. Yet, at an age when children are known to find perceptual appearances compelling, they nevertheless expect rich similarities among natural kind objects with the same name. Perhaps 4-year-olds have learned enough information about natural kinds for them to have reached this conclusion about the structure of categories. It is also possible that children are initially biased to interpret category terms this way, independent of experience. Other expectations about the structure of natural language categories appear quite early. When children as young as 18 to 24 months hear an object labeled with a common noun, they assume the term refers to the object as a whole rather than to one of its properties (Macnamara 1982). By the age of 3 or 4 and possibly earlier, children expect a noun to refer to objects that are taxonomically related (e.g., a dog and a cat) even though in the absence of a label they are likely to group objects on the basis of

thematic relations (e.g., a dog and a bone) (Markman and Hutchinson 1984). The assumption that categories will be structured as are natural kinds could be another early bias, one that helps children acquire category terms rapidly, organize knowledge efficiently, and induce information to novel exemplars of familiar categories. By expecting unforeseen nonperceptual properties to be common to members of a kind, children could go beyond the original basis for grouping objects into a category and discover more about the category members than they knew before. Children might start out assuming that categories will have the structure of natural kinds. With development, they would then refine these expectations, limiting them to properties, domains, and category types that are appropriate.