The Development of Induction within Natural Kind and Artifact Categories

SUSAN A. GELMAN

University of Michigan

Recent studies have shown that children as young as age 3½ use category membership as the basis of their inductive inferences. The present studies examine how children determine which category-based inductive inferences are warranted and which are unwarranted. Preschool and elementary school children learned various facts (e.g., "This apple has pectin inside") and reported whether they thought the facts generalized to other items varying in similarity to the target (e.g., other apples, a banana, and a stereo). Categories included both natural kinds and artifacts and varied as to how similar category members were to one another (category homogeneity, as rated by adults). Results indicate that even the youngest children placed certain constraints on their inferences. However, the preschoolers made few principled distinctions among categories, basing their inferences primarily on category homogeneity. In contrast, older children made several distinctions that seemed based on domain-specific knowledge. Most importantly, they drew more inferences within natural kinds than within artifact categories, at times even overextending the distinction. Comparison with other research suggests that increasing scientific knowledge exerts powerful effects on patterns of induction within basic-level categories. © 1988 Academic Press, Inc.

This paper focuses on children's inductive inferences and presents evidence that constraints on category-based inductions emerge early in development and undergo marked changes with age. Much of human reasoning involves induction, or the process of making inferences that ex-

tend beyond the available evidence (Holland, Holyoak, Nisbett, & Thagard, 1986; Skyrms, 1975). Inductive inferences are often based on the use of categories. That is, we often expect facts that are true of one category member to hold true of other category members as well. For example, we expect many categories to capture facts concerning internal parts (all tigers have the same number of bones), chemical structure (all water is composed of hydrogen and oxygen), behaviors (all fish breathe through gills), or other nonobvious properties. Importantly, these inferences sometimes conflict with superficial appearances. Even among outwardly dissimilar objects (e.g., typical lizards and legless lizards), common category membership can be critical for guiding inductive inferences.

Given the importance of categories for induction, a large body of developmental data and theory would predict that young children would not be able to use categories to make inferences about novel cases (either as they confront them or as they have to make predictions). Young children typically depend on superficial perceptual and functional properties to make judgments of similarity on classification tasks, free recall, free associations, and word definitions (Flavell, 1985; Mansfield, 1977). Furthermore, much of preschoolers' scientific knowledge is quite limited (Carey, 1982, 1985; Keil, 1979, 1986). An appreciation for the inductive power of categories could require a mature understanding of science, since scientific knowledge and theory provide the rationale for mature category groupings.

Despite widespread assumptions that children cannot form mature categories or look beyond the obvious (Wellman & Gelman, in press), there is recent evidence that preschool children expect categories to promote inductions (Gelman, Collman, & Maccoby, 1986; Gelman & Markman, 1986, 1987). Young children use category membership to predict underlying similarities among objects—even when perceptual similarity would lead to a different prediction. Gelman and Markman conducted a series of studies with 3- and 4-year-olds, examining children's inductive inferences with items that pitted category membership against perceptual similarity. On each of a series of problems, children were shown two objects and were taught 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. For example, on one problem children saw a tropical fish and were told that it breathes underwater. They saw a dolphin and were told it pops out of the water to breathe. They then had to decide how a second fish, a shark that looked like a dolphin, would breathe.

Children relied on category membership to draw inductions even though perceptual similarity would lead to a different conclusion. For

example, 4-year-olds usually inferred that the two fish breathe the same way. The use of identical labels cannot account for the results. In a later study category members were given synonymous (nonidentical) labels (e.g., "rabbit" and "bunny"), and children again drew inferences most often from one category member to another (Gelman & Markman, 1986). Even when no labels were used, children often drew inferences based on category membership rather than appearances (Gelman & Markman, 1987).

These studies demonstrate that preschool children expect certain categories to have a richly correlated structure that goes beyond superficial appearances. Thus, the problem of children's use of categories for induction must be construed an alternative way. Without constraints, inductive reasoning is too powerful and will produce unwarranted generalizations (Goodman, 1955; see also Sternberg, 1982). Instead of asking when children first begin to use categories as the basis of inductive inferences, the question becomes how children *limit* their inductive inferences from categories. In other words, when do children distinguish among categories in the inferences they draw? The present studies examine this question, focusing on two concerns: What are children's early inferences based on? For example, do certain categories have a special status for promoting inductions? And how do patterns of category-based induction change with development?

To answer these questions, it helps to consider some ways in which categories differ from one another. Below I discuss two category variables that have been proposed to be significant in constraining inductions: category homogeneity and object naturalness. These variables represent an important contrast: homogeneity is defined in terms of similarity, without reference to the actual content of the category; naturalness is defined specifically in terms of category content. In addition to these category variables, it is also necessary to consider the kinds of properties involved. Nisbett, Krantz, Jepson, & Kunda (1983) point out that inferences about properties from categories vary as a function of both the category and property in question. For example, whether we generalize a new property from one bird to another of the same species will depend on whether the property concerns nesting pattern or color. Accordingly, property variables are discussed in relation to each category variable. Together, these distinctions provide the framework for Study 1.

Category homogeneity. At minimum, a category that promotes inductions is structured so that category members are similar to one another on certain important dimensions. Objects that appear to be very similar in their known properties (thus forming "homogeneous" categories) may often share underlying similarities as well. For example, zebras apparently share many obvious similarities; appropriately, the class of zebras

promotes many novel inductions. Conversely, various "striped things" differ radically from one another on all dimensions besides stripedness and do not promote novel inductions. There are deep problems with precisely and objectively pinning down the notion of similarity (see Quine, 1969; Murphy & Medin, 1985). Nonetheless, gross differences in intuitive similarity, on a purely subjective basis, could predict how well a category promotes novel inferences.

A set of studies by Rips (1975) suggests the importance of similarity for inductive reasoning. Rips examined how the patterns of people's inferences reflect the "family resemblance" structure of categories. In his task subjects were told a novel fact, that a certain species of animal had a particular disease. They then were asked to estimate what proportion of various other animal species had the same disease. Rips found that adults generalized from one species to another based on how similar the species are to each other. For example, they were more likely to generalize from eagles to hawks than from eagles to ducks; they were more likely to generalize from geese to ducks than from geese to hawks.

The importance of category homogeneity leads to the straightforward prediction that superordinate-level categories will not promote as many inferences as less inclusive categories within the same hierarchy (see Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, for discussion of different levels within a hierarchy). More specific categories within a hierarchy are more homogeneous; for example, "rocking chair" is a more homogeneous category than "chair," which is a more homogeneous category than "furniture." It would be reasonable to expect children to draw inferences differently, depending on category level. Basic-level categories are especially easy for children to learn; superordinates are especially difficult (Markman & Callanan, 1984). Accordingly, the basic level may have special status for promoting children's inferences.

Beyond the predictions based on category level, however, there are serious limitations to the idea that perceived homogeneity predicts whether a category promotes inductions. Most troublesome is that inductions depend partly on theoretical considerations, beyond what is directly observable (see Carey, 1985; Goodman, 1955 for more detailed discussion). Inductions are influenced by how much variation is considered *possible* within a category, a judgment that is determined by our theories. For example, the fur color of giraffes will reliably promote inferences far into the future, whereas car color in the 1920s (then reliably black) projected only a few years ahead. Furthermore, a distinction must be made between outer, observable properties and those that are more internal or less obvious. As discussed earlier, outer properties are not always a reliable guide to those that are less obvious.

In short, adults' inductive inferences are partly dependent on a more

complex set of considerations than homogeneity alone. Nonetheless, homogeneity could serve as a useful guide when different levels within a hierarchy are compared.

Naturalness. A second distinction concerns whether a category consists of naturally occurring objects (*natural kinds*) or human-made *artifacts*. An object is natural if it was not constructed by humans.[1] The distinction between natural kinds and artifacts is not a clear-cut one, as there are many borderline cases (e.g., domesticated animals; cultivated fruits; mineral alloys). Furthermore, as discussed below, many artifacts seem to function like natural kinds in how readily they promote inferences (e.g., computers are as complex as natural kinds; plastics have predictable chemical structures). Thus, there is probably a continuum rather than a strict dichotomy between natural kinds and artifacts, in the number of inferences they promote.

With these caveats in mind, however, naturalness may be an important predictor of the extent to which a category permits inductive generalizations. One reason is that natural kinds are susceptible to scientific study (e.g., biology, chemistry, geology), whereas artifacts have traditionally not been (Simon, 1981). One purpose of a scientific theory is to discover and organize sets of nonobvious information.

Of course, it is possible to study artifacts scientifically (in computer science, archeology, etc.). However, the scientific study of artifacts differs fundamentally from the scientific study of natural kinds (Simon, 1981). Artifacts cannot be studied in and of themselves; they depend crucially on their environment for their functioning. In other words, what scientists can study are not properties inherent to the artifact object itself, but rather properties that emerge from the interaction between an object and its environment.[2]

---

[1] For philosophical discussion of the natural kind–artifact distinction, see Schwartz (1977, 1979). Psychological studies of the distinction have been rare (see Barr & Caplan, 1983; Caplan & Barr, 1985; Keil, 1986; Keil & Batterman, 1984; Malt, 1985). Naturalness is ascribed to objects taken individually; however, categories of naturally occurring objects need not be natural kinds. For example, the set of all objects named by a five-letter word in English includes naturally occurring objects (such as whales), but otherwise does not function like a natural kind. Here, the term "natural kind" is used to refer to (at least) basic-level categories of naturally occurring objects.

[2] One possible objection at this point is that even simple artifacts are studied by experts, albeit nonscientists. (For example, a dress designer knows many things about dresses.) Therefore, it certainly is the case that experts have knowledge about artifacts that extends beyond what the average person knows. It remains to be seen whether expertise concerning artifacts is as cohesive as that concerning natural kinds. Artifacts are inherently varied (see below), and knowledge of this variation must comprise much of the expertise in the field. It may be that artifact experts learn complex sets of distinctions, whereas natural kind experts learn more general laws describing entire categories.

Another hallmark of natural kinds is that they imply a particular internal structure (genetic or chemical), whereas single artifact terms do not. For example, an apple cannot be made of wood or plastic (unless it is fake), whereas a doll can be made of wood, plastic, cloth, or even dried apple. As long as the substance or structure remains the same, a particular natural entity retains its identity across varied forms (e.g., a silver dollar is still silver when melted) whereas an artifact does not (e.g., a silver dollar is no longer a dollar when melted). Although some artifact categories are constrained to include only a particular substance (e.g., all cotton blouses are made out of cotton), such categories are specially marked by language, in that they do not receive basic-level names. Importantly, internal substance and structure are nonobvious to unaided perception and often must be inferred.

Rather than capture internal structure, artifacts tend to capture functions. This is significant for two reasons: first, functions are generally well known and obvious, compared to the internal structure of an object. Second, as long as the function is carried out, then all else can vary, including shape, material, and internal parts. With new technology, artifacts can be made of new materials (such as plastic), can change size (for example, tape recorders have gotten smaller), or can change overall shape and internal parts (e.g., airplanes). With shifting fashions, artifacts can change style and color (e.g., hats, cars). Someone can invent an entirely new subtype of an artifact (witness the dramatic changes in computers over the past 30 years). In contrast, natural kinds are constrained by their genetic or molecular structure to reach a certain size, to have a certain color, to have a certain set of parts. Although natural kinds do undergo evolutionary changes, they are much more gradual and are often considered to signal a change in kind (e.g., birds are descended from dinosaurs but are not themselves considered to be dinosaurs).

The present experiments were designed to discover when (if ever) children draw principled distinctions among categories and among properties, in the inferences they draw. In Studies 1 and 2, children were given an induction task modeled on Rips (1975) and Carey (1985). For each of a series of problems, children were taught a new fact (for example, that a particular rabbit has a spleen) and had to decide whether the fact applied to other objects as well (e.g., other rabbits, a dog, a telephone). Both the type of category and the type of property were varied and included both domain-general and domain-specific distinctions. *Domain-general* factors include category homogeneity (how similar category members are to one another, relative to their dissimilarities) and property generalizability (whether a property is accidental or enduring). *Domain-specific* factors include object naturalness, object complexity, and property content (e.g.,

internal parts versus function). In order to provide a basis of comparison with Studies 1 and 2, Study 3 examines children's understanding of the literal distinction between natural kinds and artifacts (that only natural kinds are naturally occurring).

Preschoolers and second graders were studied. Preschoolers were chosen primarily because they have not yet undergone formal schooling and therefore have relatively untutored scientific beliefs. Second graders provide an interesting comparison age, given findings by Carey (1985) and Keil (1986), indicating that there is a marked reorganization in scientific knowledge between preschool and early elementary school.

STUDY 1: MAPPING OUT CHILDREN'S INDUCTIVE INFERENCES

Method

Subjects

One hundred twenty-eight children (64 preschoolers, 64 second graders) participated in the study, 96 of whom also took part in Study 3 (see below). Sixteen children of each age participated in each condition (see Properties, below). Subjects were randomly assigned to conditions,[3] with the constraint that conditions were roughly equated for age and distribution of boys and girls. Mean ages of each age group were: equally generalizable condition (4–8 and 7–10); generalizable for natural kinds condition (4–8 and 7–11); generalizable for artifacts condition (4–8 and 8–0); and nongeneralizable condition (4–7 and 8–0).

Categories and Exemplars

Table 1 presents the categories and exemplars used in all conditions of Study 1. Each child was tested on 18 sets of items: 9 natural kinds and 9 artifacts. To ensure that we sampled from a wide range of items, we included three categories at each of three levels of complexity, for both natural kinds and artifacts. Complexity ratings were gathered from 16 adults. Subjects were instructed to rate each of a series of items on a scale of 1 to 9 (1 = simple; 9 = complex). "Complex" was defined as "has many different parts and complicated internal workings"; "simple" was defined as "has few or no different parts and a more uniform internal structure." The mean complexity ratings and standard deviations of the target items at each level are, for natural kinds: animals ($M = 6.29$, SD = 2.04), vegetation ($M = 4.19$, SD = 2.02), and substances ($M = 2.25$, SD = 1.33); for artifacts: complex ($M = 6.27$, SD = 1.65), medium–complex ($M = 3.23$, SD = 1.46), and simple ($M = 1.65$, SD = 0.86).

On every set, the experimenter taught a new fact about one picture, the target picture. The child was then tested on four probe pictures, to determine how far he or she generalized the new piece of information. The four probe pictures included (with examples for the target picture *daffodil*):

    (a) same category and appearance as target (another daffodil),
    (b) same category but different appearance than target (rose),

---

[3] The generalizable for artifacts condition was conducted separately, a few weeks following the other conditions. However, the schools were drawn from a comparable population of children in the same city and socioeconomic class.

TABLE 1
Target and Probe Items Used in Study 1

Natural kinds
Rabbit (brown rabbit, white rabbit, dog, phone)
Fish (yellow fish, red fish, turtle, couch)
Spider (small spider, big spider, bee, spoon)
Apple (red apple, yellow apple, banana, stereo)
Carrot (small carrot, big carrot, peas, shoe)
Flower (daffodil [flower], rose [flower], plant, bowl)
Gold (bar of gold, gold nugget, silver, airplane)
Diamond (polished diamond, uncut diamond, pearl, flute)
Water (ocean water, cup of drinking water, milk, block)
Artifacts
Ball (golf ball, football, doll, pepper)
Cup (teacup, plastic cup, plate, lettuce)
Hammer (claw hammer, ball-peen hammer, saw, ant)
Coat (raincoat, fur coat, pants, snow)
Chair (easy chair, wooden chair, table, lemon)
Guitar (acoustic guitar, electric guitar, drum, frog)
Clock (clock with a pendulum, digital clock, watch, salt)
Bike (10-speed bike, child's bike, car, tree)
TV (color TV, black-and-white TV, radio, mouse)

Note. Actual labels used by experimenter are underlined.

    (c) same superordinate category as target (houseplant), and
    (d) unrelated category from a different ontological category (mixing bowl).

This set enables us to see just how far a child's generalization has gone: within the subordinate level (daffodils only), within the basic level[4] (flowers only), or within the superordinate level (plants only). It also allows us to distinguish orderly patterns from random guessing (such as generalizing to just the rose and the mixing bowl). Pictures included realistic color drawings and photographs, mounted on 3 × 5-inch cards.

Homogeneity Ratings

Ratings of category homogeneity were gathered from adults to serve as a basis of comparison with children's inductive inferences. In a preliminary study, 20 adults were asked to rate the homogeneity of each of 52 categories as a whole (with each subject rating 26 categories). Subjects were given exactly 1 min to look over the entire test booklet to familiarize themselves with the range of categories. Then the subject rated each category one at a time. For each category, the subject had 15 s to think about the category (all the different subtype) and how these subtypes differed from one another) and then 5 s to indicate how

---

    [4] The substances have not been studied by Rosch and so could not be classed as basic level. The present studies provide an especially conservative test of a natural kind–artifact difference, because *tree, flower, bird,* and *fish* were included as basic-level natural kinds— all of which are fairly general categories that may at times function as superordinates; see Rosch *et al.* (1976).

similar category members are to one another, on a scale of 1 ("very similar") to 9 ("very different").[5]

Properties

As discussed earlier, whether an inference is reasonable or not depends not only on the category but also on the property. Similarly, children's inferences may depend on which facts they learn. In order to obtain a broad sample of children's inferences, facts were varied systematically in terms of how readily they should promote inferences from the target picture. Four sorts of properties were used: equally generalizable, generalizable for natural kinds only, generalizable for artifacts only, and nongeneralizable. These are given in Table 2. Each type is described below.

Equally generalizable. These properties should legitimately promote inferences to members of the same basic-level category. They are modeled on familiar (true) properties, but are themselves unfamiliar. For example, children already know that a rabbit likes to eat carrots; the unfamiliar property is, "This rabbit likes to eat alfalfa." Each property contains one unfamiliar word so that children could not have known the facts beforehand. Nonetheless, the properties deal with domains that are quite familiar and sensible to children (what rabbits like to eat, what coats are used for, what one can do with cups).

Generalizable for natural kinds. Properties in this condition characterize the natural kind–artifact distinction by referring to substance, internal structure, and previous form or developmental stage. Four different sentence frames were used: "has (an) X inside," "made out of X," "has parts made out of X," and "used to be an X." "X" was always an unfamiliar word.

Generalizable for artifacts. For some properties, most notably those describing function, artifact categories may promote at least as many inferences as natural kinds. Facts in this condition were designed to stress function and use. To accomplish this, three different sentence frames were used throughout: "used for X," "you can X with it," and "used to make X." "X" was always an unfamiliar word.

Nongeneralizable. These properties should not generalize from one category member to another. Each involves one or more of the following: inherent variation within the individual (e.g., "dirty"), historical accident (e.g., "fell on the floor"), or a temporal aspect (e.g., "a year old"). They consist entirely of familiar words, so as to provide an especially clear case of nongeneralizable properties.

Procedure

Children were tested individually by a female experimenter. Each child was tested on all 18 picture sets, but heard only one type of property throughout, equally generalizable, generalizable for natural kinds, generalizable for artifacts, or nongeneralizable. The experimenter explained that she was going to show the child some pictures and ask some questions about them. The child received a sticker after each set of questions, regardless of his or her answers. The first item was a set of three easy warm-up questions, designed to help

---

[5] These ratings were based on an unconstrained task that did not specify the basis of subjects' ratings. A second study was conducted to validate that subjects were rating homogeneity and to control for the sort of properties that were being rated. One group of subjects generated properties (see Rosch & Mervis, 1975), which were then amended to remove idiosyncratic facts and falsehoods. A second group of subjects rated the homogeneity of each category with respect to each property that was generated. Such ratings correlated significantly with the ones reported here (at the basic level).

TABLE 2
Target Properties Used in Study 1

Equally generalizable condition
  Rabbit: likes to eat alfalfa
  Fish: needs branchias to breathe
  Spider: catches bestiolas
  Apple: can use it to make a sauce called "chutney"
  Carrot: if you eat it, it's good for your oculis
  Flower: needs CO_2 to grow
  Gold: if you put it in flamma, it melts
  Diamond: sharp enough to cut vitrum
  Water: can turn into vapor
  Ball: canines (they're a kind of animal) like to play with it
  Cup: procyons (they're a kind of animal) can drink out of it
  Hammer: used for putting in brackets
  Coat: protects you from precipitation [pluvia]
  Chair: a vole (that's a kind of animal) can sleep on it
  Guitar: you can play cantatas on it
  Clock: measures punctis
  Bike: needs petroleum
  TV: has a knob for verticality
Generalizable for natural kinds condition
  Rabbit: has a spleen [omentum] inside (from Carey, 1985)
  Fish: has parts made out of calcium
  Spider: used to be a deutovum
  Apple: has pectin inside
  Carrot: made out of cellulose
  Flower: used to be a gemma
  Gold: made out of A–U
  Diamond: used to be a piece of carbon
  Water: has hydrogen inside
  Ball: has glycerine inside
  Cup: made out of ceramic [argillo]
  Hammer: used to be a piece of chalybs
  Coat: made out of Gore-tex
  Chair: has urethane inside
  Guitar: used to be a larch
  Clock: has a pondus inside
  Bike: used to be a piece of bauxite
  TV: has parts made out of quercus
Generalizable for artifacts condition
  Rabbit: you can rollick with it
  Fish: used to make fumet
  Spider: used for expunction
  Apple: you can repast with it
  Carrot: used to make a proboscis
  Flower: used for dizening
  Gold: used to make chaplets
  Diamond: used for riving
  Water: you can depurate with it

TABLE 2—*Continued*

Ball: you can roister with it
Cup: used for imbibing
Hammer: used to make apertures
Coat: used for occulting
Chair: you can loll with it
Guitar: used to make diapason
Clock: used for horology
Bike: you can traverse with it
TV: used to make tintamarre
Nongeneralizable condition
  Rabbit: is cold
  Fish: has a piece of grass stuck to it
  Spider: is a year old
  Apple: has a little scratch on it
  Carrot: came from Nevada
  Flower: was in somebody's bedroom yesterday
  Gold: is dirty
  Diamond: belongs to a doctor
  Water: some sugar fell in it this morning
  Ball: is a little dirty
  Cup: has gum stuck on the bottom
  Hammer: is a year old
  Coat: smells yucky
  Chair: came from Oregon
  Guitar: was in sombody's kitchen yesterday
  Clock: has a little scratch
  Bike: belongs to a teacher
  TV: fell on the floor this morning

Note. Words in brackets were used with the second graders.

children feel confident and comfortable in answering questions, for example, "Is this [green bowl] green or is it yellow?"

For each of the remaining 18 sets of items, children were taught a new fact about one object or substance and then were asked whether it was true of the other four items in that set. For example, children in the equally generalizable condition heard: "This [brown] rabbit likes to eat alfalfa. See this telephone? Do you think it likes to eat alfalfa, like this rabbit? See this [white] rabbit? Do you think it likes to eat alfalfa, like this rabbit? See this dog? Do you think it likes to eat alfalfa, like this rabbit? See this [brown] rabbit? Do you think it likes to eat alfalfa, like this rabbit?" The target picture, on which the initial fact was taught, was placed on the table first and remained on the table throughout the four probe questions in each set. The pictures corresponding to the four probe pictures were placed on the table one at a time, directly below the target. Each probe picture was removed from the table before the next one was shown.

The order of the sets was randomized separately for each subject, as was the order of probe pictures within each set (e.g., whether the child saw the telephone first or the dog first). Before showing the child each picture card, the experimenter announced what it was, e.g., "Now I'm going to show you a telephone" or "Now I'm going to show you another

rabbit." As each picture was named, the experimenter pointed to it, to focus the child's attention on the picture as well as the question. Care was taken not to stress either the category names or the word "this." Throughout the session, children's spontaneous comments were written down.

Children were encouraged to give an answer of "yes" or "no" even when unsure. However, children rarely expressed uncertainty (e.g., "don't know," "maybe," or "I'll take a guess"), doing so on only 0.3% of all trials. Over 84% of the children never expressed uncertainty at all. For the youngest group, if a child gave only "yes" or only "no" answers during the first two sets of items, the experimenter gave the child his or her sticker and then said, "OK, ready for the next one? Remember, sometimes the answer might be 'yes,' and sometimes the answer might be 'no.'" This reminder was given to make certain that children realized that they were free to use either answer.

Items were blocked, with natural kinds in one block and artifacts in the other. The second graders received both blocks in a single session, with the order of blocks counterbalanced across subjects of each sex within each condition. The preschool children were tested in two sessions, usually within 2–3 days of each other. For the preschoolers, the order of blocks was also counterbalanced across subjects of each sex within each condition.

Results

Overview

Initial analyses considered overall generalization tendencies (Fig. 1) and the extent to which these differed from chance (Table 3). As can be seen in Fig. 1, children were sensitive to the generalizability of the prop-
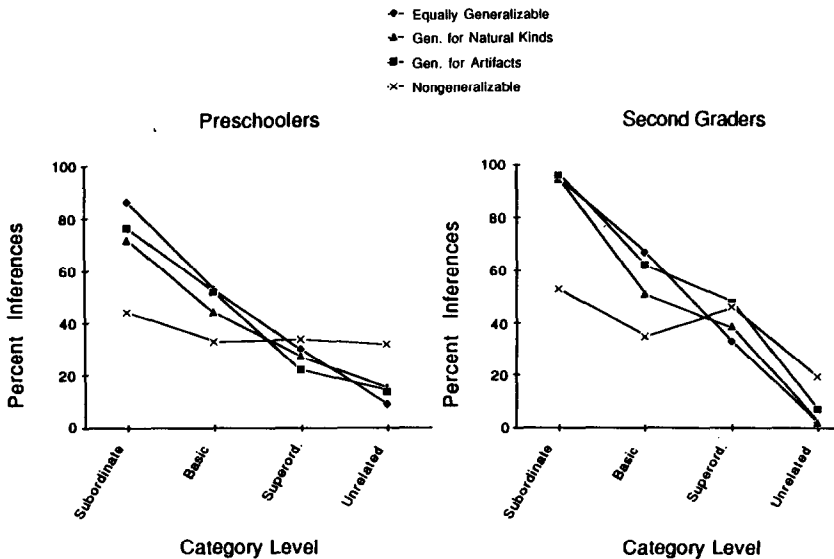

FIG. 1. Study 1, mean percentage inferences drawn as a function of property type, category level, and age.

erty; the three generalizable conditions yielded similar findings and can be distinguished from the random responding in the nongeneralizable condition. Furthermore, children attended to category level; they drew more inferences within the subordinate level (e.g., easy chairs) than within the basic level (e.g., chairs), within the basic level than within the superordinate level (e.g., furniture), and within the superordinate level than to an unrelated picture (e.g., from a chair to a lemon). Children were sensitive to these several layers of potentially classifying a single object.

To compare performance against chance, I classified children's answers into response patterns, taking into account how children answered for all questions in a given set of four probe questions. Three response patterns are considered "sensible," as they are the only ones for which children drew more inferences, the greater the similarity between test and target pictures: (a) the child drew an inference only to the picture of the same category and similar appearance as the target (e.g, from one daffodil to another), (b) the child drew inferences to both pictures of the same basic level as the target (e.g., from a daffodil to a daffodil and a rose), and (c) the child drew inferences to all three pictures of the same superordinate level as the target (e.g., from a daffodil to a daffodil, rose, and houseplant). Since children were asked to make one of two responses ("yes" or "no") to each of four items in a set (all the inferences from a given target), the total number of possible responses to a set was $2^4 = 16$. Thus, the probability that a child's performance would be classified into a sensible response pattern is 3/16 or 0.19.

Table 3 shows the proportion of response patterns within each age, property type, and naturalness type that fell within one of the sensible response patterns. These proportions far exceeded chance in the three

TABLE 3
Study 1, Percentage Sensible Response Patterns (Chance = .19)

	Preschool	Grade 2
Equally generalizable		
Natural kinds	.77	.90
Artifacts	.69	.87
Generalizable for natural kinds		
Natural kinds	.53	.87
Artifacts	.58	.74
Generalizable for artifacts		
Natural kinds	.60	.75
Artifacts	.62	.82
Nongeneralizable		
Natural kinds	.09	.30
Artifacts	.09	.26

generalizable conditions. In contrast, response patterns in the nongeneralizable condition were essentially random. Thus, children clearly distinguished between generalizable and nongeneralizable properties.

I turn next to a detailed consideration of the results, focusing on each of the following in turn: (a) category homogeneity, (b) object naturalness, and (c) object complexity. Finally, I examine children's justifications.

Category Homogeneity

As discussed in the overview, all children were sensitive to category homogeneity as defined by level (e.g., subordinate versus basic level). To test the possibility that children were sensitive to the homogeneity of each category *within* the basic level, rank–order correlations were performed comparing children's inferences with adults' ratings of homogeneity, for each category collapsed across subjects. The adult homogeneity ratings were taken from the preliminary study; children's inferences were those to pictures of the same basic-level category but appearance different from that of the target. These pictures were chosen because they show more variability than pictures of the same category and similar appearance.

There are obvious limitations to comparing children's inferences with adults' homogeneity ratings, since children's category representations may be biased, incomplete, or otherwise at variance with those of adults (e.g., Nelson, 1974). Nonetheless, this analysis yielded some suggestive results (see Table 4). The most interesting finding is that 4-year-olds evidenced moderate to high correlations with artifact categories. In other words, 4-year-olds drew more inferences within artifact categories that adults rated as homogeneous (such as hammers or TVs) than within artifact categories that adults rated as heterogeneous (such as coats or balls). All other correlations were low and nonsignificant.

TABLE 4
Correlations between Adult Homogeneity Ratings and Children's Inferences in Study 1

	Preschool		Grade 2	
	NKS	ARTIFS	NKS	ARTIFS
Equally generalizable	.22	.74*	.28	.17
Generalizable for natural kinds	− .08	.72*	− .07	.64
Generalizable for artifacts	.13	.36	.16	.29
Nongeneralizable	.15	.58	− .06	− .16
N	9	9	9	9

* Significantly greater than chance, $p < .05$.

Object Naturalness

To examine the effects of naturalness, eight composite scores were calculated for each subject: four for natural kinds and four for artifacts, with a score for each category level (subordinate, basic, superordinate, unrelated). Each score was the sum of total "yes" responses (i.e., inferences drawn) in each cell. These scores were entered into a 3 (property type) × 2 (age) × 2 (naturalness) × 4 (category level) ANOVA. Only the three generalizable conditions were included, since the nongeneralizable condition showed mostly random responding (see Table 3).[6]

As anticipated, category level strongly affected children's inferences (as shown in Fig. 1), min-$F'(3,116) = 151.33, p < .001$. (See Clark, 1973, for discussion of min-F' analyses). Furthermore, older children drew more inferences overall than younger children (min-$F'(1,102) = 4.30, p < .05$), although the age effect varied with category level, min-$F'(3,279) = 8.45, p < .01$. All of the remaining significant differences involved object naturalness. There was a naturalness × level interaction, min-$F'(3,65) = 3.85, p < .05$, a naturalness × age interaction, min-$F'(1,69) = 5.14, p < .05$, and a naturalness × level × age interaction, min-$F'(3,129) = 2.86, p < .05$. Post hoc comparisons using Scheffe's test revealed that the natural kind–artifact difference was significant only among the second graders and emerged only at the basic level (e.g., when drawing an inference from a raincoat to a fur coat), $p < .05$. Overall, there were no significant natural kind–artifact differences among preschoolers, or at the other category levels. Figure 2 shows children's inferences to natural kinds and artifacts in each condition.[7]

There was also a nonsignificant trend toward a property type × age × naturalness × level interaction, min-$F'(6,287) = 2.09, p > .05$. Thus, a further analysis was conducted to examine the effect of specific property content on children's inferences. Of interest is whether children appropriately draw most inferences to natural kinds when taught properties that are generalizable for natural kinds, and most inferences to artifacts when taught properties that are generalizable for artifacts. By Scheffe's test, preschoolers showed no effects for condition (property type), for either natural kinds or artifacts. In contrast, second graders did draw inferences differently across conditions. When generalizing within basic-level artifact categories (e.g., chairs), second graders appropriately drew

---

[6] A separate analysis of the nongeneralizable condition confirmed that there were no significant natural kind–artifact differences at either age in this condition.

[7] Preschoolers' roughly 50% performance is primarily due to individual differences: 63% of the preschoolers gave either more or fewer inferences than would be expected by chance on a binomial test, $p < .05$, for items of the same category and different appearance.
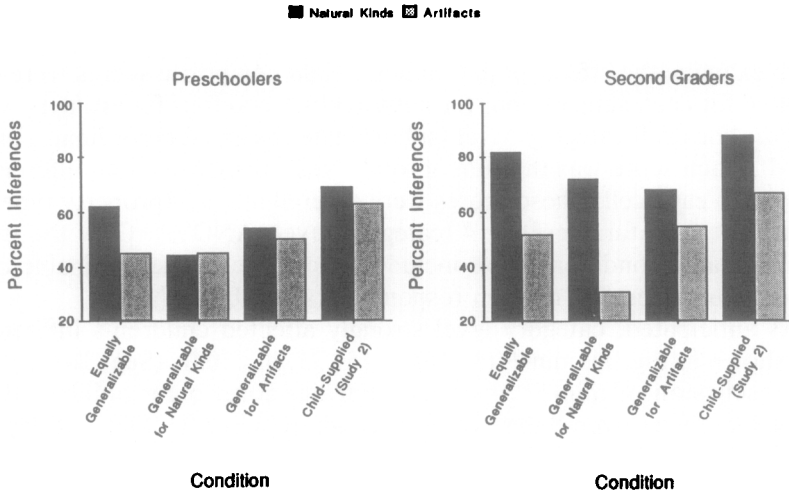
FIG. 2. Studies 1 and 2, generalizable conditions. Mean percentage inferences drawn to natural kinds and artifacts of the same basic-level category, different appearance.

fewest inferences for properties that are generalizable for natural kinds (e.g., "has urethane inside"), $p < .05$. The only other significant difference occurred for second graders when generalizing within natural kind superordinates (e.g., from a spider to a bee). There, children drew more inferences for properties that are generalizable to artifacts (e.g., "used for expunction") than for properties that are equally generalizable (e.g., "catches bestiolas"), by Scheffe's $p < .05$. This last result was unanticipated, but suggests that functional properties (those generalizable to artifacts) may be especially likely to characterize an entire superordinate category.

Object Complexity[8]

Three levels of object complexity were defined, for both natural kinds and artifacts (see "Categories and Exemplars," above). The different levels of complexity for natural kinds represent different ontological categories (substances versus vegetation versus animals; see Keil, 1979) and different content domains. Given recent evidence that children clearly distinguish between animate and inanimate objects (Gelman & Spelke, 1981; Gelman, Spelke, & Meck, 1983), and to some extent between ob-

---

[8] On an analysis including the items used in Study 1, complexity is not significantly correlated with homogeneity for natural kinds separately, artifacts separately, or all items combined (all $p > .05$, by Spearman's ρ).

jects and substances (Keil, 1979; Smith, Carey, & Wiser, 1985), we may expect children to draw most inferences to animals and fewest to substances. Complexity differences for artifacts, although as fully pronounced as those for natural kinds as measured by adult ratings, do not correspond to ontological differences. The complexity analyses were conducted primarily in order to elucidate the natural kind–artifact distinction. Thus, the analyses focused on the level that yielded a natural kind–artifact distinction to begin with: pictures of the same category as the target but of different appearance (e.g., a digital clock, where the target was an analog clock). Only the three generalizable conditions were included.

A 3 (property type) × 2 (age) × 2 (naturalness) ANOVA was conducted. Effects for age, naturalness, and property type were discussed in the previous section, so this analysis will report only those effects involving complexity. There was a significant main effect for complexity, $F(2,180) = 22.40, p < .0001$, and a complexity × naturalness interaction, $F(2,180) = 24.72, p < .0001$. (Min-$F'$ analyses were not conducted because there were too few items per complexity level to make such an analysis meaningful.) Complexity did not interact with either age or property type. The complexity × naturalness interaction is shown in Fig. 3.

As can be seen, substances promoted the fewest inferences among natural kinds; complex artifacts promoted the most inferences among arti-
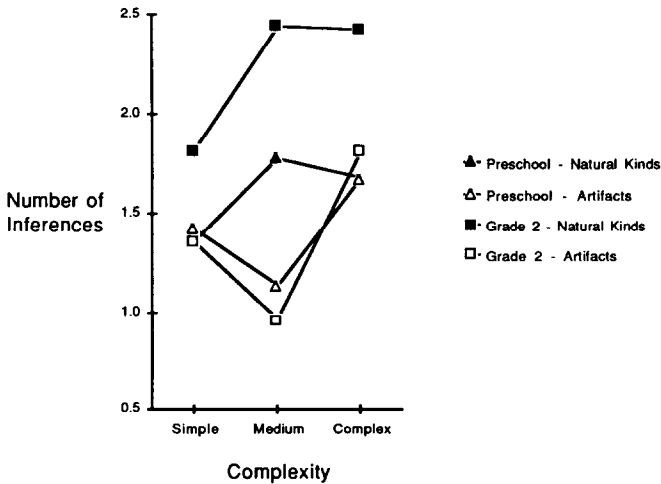

FIG. 3. Study 1, generalizable conditions. Mean number of inferences (out of 3) drawn to pictures of the same basic-level category, different appearance as a function of complexity and age.

facts.[9] In contrast to expectations, there was no evidence of a strict animate–inanimate distinction, at either age; children did not treat animals any differently from vegetation (at either age) or complex artifacts (at preschool age). Despite the significant effects for complexity, complexity per se cannot account for children's answers, since inferences did not increase monotonically with object complexity. This was especially evident with the artifacts, where children consistently drew fewer inferences to the medium–complex artifacts than to either the simple or the complex artifacts. Furthermore, only one rank-order correlation conducted between adult complexity ratings and children's inductions was significant: that of second graders drawing inferences to natural kinds in the generalizable for natural kinds condition (Spearman's ρ = .73, p < .05).

Justifications

All of children's spontaneous comments were coded as to their content (as detailed below). A second coder rated 25% of the set, yielding agreement of 84%. The first analysis examined whether children translated the unfamiliar words they heard into something more familiar. We coded every time a child translated an unfamiliar word (e.g., for alfalfa, "That's carrots"), gave a partial translation (e.g., "That's probably something like carrots or lettuce"), gave a questioning translation (e.g., "Like a carrot?"), or outright asked what the word meant (e.g., "What is alfalfa?"). Nontranslations were coded as well, e.g., "Some rabbits like to eat carrots, too," which by their wording imply that the child did not translate the novel word into a familiar one. Finally, we coded each time the child claimed to know the unfamiliar word. Altogether across both ages and all generalizable conditions, 172 comments fell into these categories. Most typically, children treated the novel words as unfamiliar; they claimed to know the meaning only 6 times and translated the word directly only 18 times (with 15/18 translations made by the preschoolers).

[9] Given that substances were the only nonobject categories and did seem to function differently from other natural kinds, two additional analyses were performed excluding substances. (a) The homogeneity ratings were recalculated for natural kinds. Although all correlations increased when substances were removed, none reached significance. (b) A property type × age × naturalness × level ANOVA was conducted excluding the substance items. On this analysis, preschoolers showed more of a natural kind–artifact difference than in the main analysis, for pictures of the same basic level as the target but of different appearance. However, the natural kind–artifact difference among the preschoolers still failed to reach significance across the three property types by Scheffe's test, and there was no significant interaction with property type. Moreover, the naturalness effect was still significantly more pronounced among the second-grade children than among preschoolers, as shown by a naturalness × level × age interaction, min-F' (3, 118) = 3.27, p < .05.

Forty-three of the comments were partial translations or questioning translations, and 100 of the comments were nontranslations (33) or outright asking what the word meant (67).

It is also important to determine how often children explicitly invoked the category name when drawing an inference. Children may have invoked category names to explain or justify their answers, in effect saying, "I drew [did not draw] that inference *because* that object is a [category name]." If so, children should have shown the following patterns of naming: When children made an inference (by saying "yes"), they should have named the level at which the two pictures share a category; e.g., when drawing an inference from a golf ball to a football, they should have said "It's a ball" or "It's a toy," *not* "It's a football." When drawing an inference from a golf ball to a doll, they should have said, "It's a toy," not "It's a doll." In contrast, when children refused to draw an inference (by saying "no"), they should have named the level at which the two pictures were members of contrasting categories below a common superordinate: "It's a football" when comparing a football to a golfball, or "It's a ball" or "It's a football" when comparing a football to a doll. In other words, if children were basing their answers on category membership, then they should have adjusted their level of naming to conform to the appropriate contrast.

Because of the small number of comments, the data were collapsed across the three generalizable conditions, across naturalness, and across age of child. The data are quite clear. When children said "yes," 84% of their labels (59 out of 70) referred to a category that was common to both target and test pictures. When children said "no," 92% of their labels (68 out of 74) referred to a level at which the pictures were members of contrasting categories. Chance scores were calculated for each of the three levels of word provided (subordinate, basic, and superordinate) and for each level of generality of item.[10] Overall, chance alone would predict that 99 of children's labels (69%) would be appropriate. However, subjects gave 127 appropriate labels (88% of the total). Therefore, it appears that children supplied category labels in order to justify their answers. The nongeneralizable condition provides an important control. Only

_____

[10] Chance levels were calculated for each of the three types of names (subordinate, basic, and superordinate) and for each level of item (subordinate, basic, superordinate, and unrelated). The total number of names of a given type supplied at each item level was multiplied by the proportion of answers to that item level that would be consistent with a label of that sort (e.g., the number of basic-level names given to basic-level items was multiplied by the proportion of times subjects said "yes" on basic-level items). A calculation was then performed of how many labels were predicted to conform to the appropriate pattern and how many labels actually did conform to the pattern.

three justifications in that condition included a category label at all. Thus, children spontaneously justified their answers by invoking the category label and only where appropriate.

Comments about appearances were coded into three types: those that mentioned that the pictures looked alike or the same, those that mentioned that the pictures looked different, and all other descriptions of appearance. The first result of interest is that children mentioned similarity of appearance only when drawing an inference (22/22 instances). Second, when children commented on some difference in appearance between target and test pictures, they also often did so when drawing an inference (46/65 instances). That is, children often qualified a "yes" answer with some comment such as, "but they look different." Children readily drew inferences from one category member to another, despite explicitly noted differences in appearance.

Discussion

Study 1 was designed to map out children's inductive inferences across a range of categories and properties. For each of a series of problems, children learned a fact about an object and had the opportunity to generalize the fact to other objects. These problems posed a strong test of children's inference abilities, in part because each new property was taught about a single object only. This was also a strong test in that each new fact included an unfamiliar word. Consequently, children could have interpreted the properties in any idiosyncratic way. For example, "has a spleen inside" could refer to what the rabbit ate for lunch, rather than an internal organ. Therefore this study may actually have underestimated children's tendency to draw inferences within a category.

Nonetheless, children consistently drew inferences within each category. As in previous work (Gelman & Markman, 1986, 1987), inferences were not dependent on perceptual similarity. Children drew many inferences when the target and probe were markedly dissimilar (e.g., from a daffodil to a rose), especially for natural kinds. In their justifications, too, children indicated that inferences were based on category identity more than surface similarity; they often named the category that was appropriate for explaining their response and drew inferences despite remarking on perceptual dissimilarity.

The results from this study can tell us what children's early inferences are based on and how they change with age. The patterns of response were markedly different at each age. *4-year-olds* were sensitive to two domain-general factors: category homogeneity and property generalizability. They drew predictably more inferences, the more similar the probe items were to the target. For example, when children learned that a color TV has parts made out of quercus, they were most likely to gener-

alize to another color TV, less likely to generalize to a black-and-white TV, even less likely to generalize to a radio, and least likely to generalize to a mouse. To the extent that two objects were alike in known ways, children presumed they were alike in unknown ways as well. Furthermore, preschool children's inferences within basic-level categories correlated significantly with adult homogeneity ratings. That is, children were especially likely to draw inferences within those basic-level artifact categories that adults rated as very homogeneous. The results suggest that 4-year-olds may be especially sensitive to category homogeneity as they perceive it.

The patterns just described did not emerge in the nongeneralizable condition, where the properties taught were temporary or accidental (e.g., "This clock has a little scratch"). This result reveals a powerful constraint on children's inductions. By age 4, then, children expect to find deep similarities among category members, regardless of the category. They are also sensitive to the breadth of the category in drawing inferences. And they are sensitive to how transient versus enduring the property is.

Second graders also relied on category level and property generalizability to guide induction. They, too, drew fewer inferences to superordinate or unrelated than basic level categories. And again this pattern was disrupted in the nongeneralizable condition. Beyond such similarities, however, marked changes took place in the 3 years from preschool to second grade. The major developmental difference is one of refinement: 4-year-olds made few principled distinctions among categories; second graders made several. This was most evident in a shift from a small and unstable natural kind–artifact distinction among 4-year-olds to a large, consistent distinction among second graders. Second-grade children developed clearly different expectations about natural kinds (especially biological kinds such as "rabbit" and "apple") and artifact categories (such as "chair" and "hammer"). They expected to find rich similarities among natural kinds—in behavior, internal parts, and previous life form —but not among artifacts. In fact, second graders overextended the natural kind–artifact distinction, invoking it even when unwarranted, as seen in the equally generalizable condition and, to a lesser extent, in the generalizable for artifacts condition.

Second graders also showed more highly differentiated patterns based on property content; when considering artifact categories, second-grade children drew the fewest inferences from properties that are generalizable for natural kinds (such as internal parts).

STUDY 2: CHILD-SUPPLIED PROPERTIES

In Study 1, patterns of induction differed strikingly depending on

whether children had heard generalizable properties (e.g., "This apple has pectin inside") or nongeneralizable properties (e.g., "This apple has a little scratch on it"). It was argued that preschool children honor a principled distinction between enduring and more transient properties, expecting only the former to promote category-based inductions. However, the generalizable–nongeneralizable distinction in Study 1 was confounded with whether or not the property included an unfamilar word. Study 2 was conducted to determine whether similar patterns of results are obtained when the generalizable properties include only familiar properties. The procedure was identical to that of Study 1 except that children were asked to supply the properties themselves.

Method

Subjects

Thirty-two children participated in this study, 16 each at two ages: 4-year-olds from two preschools in Sunnyvale, California (range = 4–1 to 5–4; mean age = 4–9) and second graders from two elementary schools in the same area (range = 7–7 to 8–5; mean age = 8–0). Nine 4-year-olds and eight second graders were girls. None of the children had participated in Study 1.

Items

The same picture sets as in Study 1 were used, including the warm-up questions.

Procedure

The procedure was identical to that of Study 1, except that each child supplied all properties rather than learning them from the experimenter. More specifically, on each item the experimenter framed a question based on the property given in the equally generalizable condition of Study 1. The child's answer to this question became the property that was used to test his or her inductions from that item. The questions were derived from the equally generalizable condition because it concerned familiar content domains, and so the questions should be possible for children to answer. For example, on one set children in Study 1 had been told, "This [target] rabbit likes to eat alfalfa. Does this rabbit [dog, phone] like to eat alfalfa, like this rabbit?" In Study 2, the wording was, "What does this [target] rabbit like to eat?" If the child said, for example, "lettuce," the experimenter would ask, "Does this rabbit [dog, phone] like to eat lettuce, like this rabbit?" The questions used to elicit the properties presented all information but the unfamiliar word.

Results and Discussion

The major results from Study 1 were replicated; 81% of children's inferences followed sensible patterns (as defined in Table 3 of Study 1). Thus, the generalizable–nongeneralizable distinction in Study 1 could not be due to differences in property familiarity. Moreover, the relative advantage for natural kinds again emerged between preschool and second grade. A 2 (age) × 2 (naturalness) × 4 (level) ANOVA was conducted on the inference scores in this study. There was a significant age × natural-

ness × level interaction, min-$F'(3,114) = 3.28$, $p < .05$. As in Study 1, the natural kind–artifact distinction was found for pictures of the same category but different appearance than the target (see Fig. 2 for illustration). Furthermore, I conducted an ANOVA comparing the results of the present study with the results of the equally generalizable condition in Study 1 (assuming for the purpose of this analysis that subjects in the two studies were comparable). On this analysis, there were no significant main effects or interactions based on property type (whether the property was unfamiliar or child supplied); all $p > .05$. In sum, children's inferences concerning novel properties follow patterns similar to their beliefs about familiar facts.

STUDY 3: CONCEPTUAL DISTINCTION BETWEEN NATURAL KINDS AND ARTIFACTS

Given that the natural kind–artifact distinction does not emerge consistently until second grade, it is important to determine whether children grasp the distinction on a literal level. It could be that preschoolers do not know enough about the origins of things to know which objects are human made and which are not. This could contribute to their failure to make a natural kind–artifact distinction when drawing inductive inferences. On the other hand, if preschoolers do understand the literal distinction between artifacts and natural kinds, then some other developmental change is required to explain the different patterns of inferences by second grade.

Study 3 examined whether children realize that only artifacts are human made. It was possible that children might have some difficulty with the question, given previous research by Piaget (1929). In *The Child's Conception of the World,* Piaget describes young children as "artificialist," or "regarding things as the product of human creation, rather than in attributing creative activity to the things themselves" (p. 253). For example, children reported that the sun and the mountains were created by humans for human purposes. Study 3 differs from the Piagetian procedure in two ways. First, children were interviewed about a range of familiar, accessible objects (such as lemons and birds), rather than remote phenomena about which they have less direct experience. Second, children were specifically asked whether or not the objects were made by humans. In contrast, Piaget's interviews were more open ended, for example, often including the question, "How are clouds (snow, stones, mountains) made?" Such wording requires the child to report not only whether the object is human made, but also to specify what the process was—a more difficult question. Furthermore, such wording may incorrectly imply that someone actually did make the clouds and snow. The

simpler procedure in Study 3 provides a more sensitive test of children's grasp of the natural kind–artifact distinction.

Method

Subjects

Ninety-six children participated in the study: 48 preschoolers from a single preschool in Sunnyvale, California (range = 4–1 to 5–2; mean age = 4–8) and 48 second-graders from a single elementary school in the same city (range = 7–1 to 8–7; mean age = 7–11). Twenty-seven of the younger children and 30 of the older children were girls. All children participated in Study 1.

Items

Natural kinds included two animals (tiger, bird), two types of vegetation (lemon, tree), and two substances (milk, salt). Artifacts included two categories each at three levels of complexity: simple (hat, knife), medium (doll, book), and complex (car, telephone).

Procedure

This study was conducted directly after Study 1, as part of the same session. At the end of Study 1, the experimenter said, "I'm going to ask you a few questions without any pictures. Do you think people make X's?" (where X is a category name). If the child answered, "Yes," this was followed by, "How?" If the child answered, "No," he or she was asked, "Why not?" The follow-up questions were included to clarify children's answers. Occasionally a child said, "I don't know" to the follow-up question, so another prompt was given, e.g., "Where do they come from?" Each child was tested on six categories, either all natural kinds or all artifact categories. Within each age group, condition in Study 1, and order of blocks in Study 1, roughly half the children heard natural kind questions; the others heard artifact questions. Altogether, 50 children (25 at each age) answered questions about natural kinds; 46 children (23 at each age) answered questions about artifacts. The order of categories was randomly determined for each child.

Results and Discussion

At the most literal level, children clearly grasp the natural kind–artifact distinction; they realize that humans can construct only artifacts and not natural kinds. Each child received a composite score that counted the number of times he or she said that objects of a particular type (e.g., cars) are made by people. Because every child was tested on six object categories, scores could range from 0 to 6. Then a 2 (age) × 2 (naturalness) ANOVA was conducted. On this analysis, children more often said that people make artifacts (84%) than that people make natural kinds (23%), as seen by a main effect for naturalness, min-$F'(1,16) = 36.52, p < .001$. There was also an age × naturalness interaction, indicating that second graders were more pronounced than preschoolers in their natural kind–artifact distinction, min-$F'(1,39) = 9.60, p < .01$. They were more apt to say that people make artifacts and less apt to say that people make natural kinds. Nonetheless, even the preschoolers clearly believe that artifacts tend to be human made and that natural kinds are not. A χ^2 was

calculated, based on the number of preschoolers who said "yes" no more than half the time (zero to three times out of six possible) and the number who said "yes" most of the time (four to six times out of six possible), for natural kinds and artifacts separately. Sixteen out of 23 children said "yes" most of the time for artifacts; only 3 out of 25 children said "yes" most of the time for natural kinds, χ^2 ($df = 1$) = 16.60, p < .001.

Although the overall natural–artificial distribution is quite strong, children at both ages often reported that substances are human made (mean = 50% at younger age, 41% at older age). This is reflected in a two-way ANOVA conducted on the natural kinds data, revealing a main effect for complexity (animal, vegetation, or substance) ($F(2,96)$ = 39.33, p < .001), as well as for age ($F(1,48)$ = 6.09, p < .05). It appears that children have not mastered a strict natural kind–artifact distinction, but rather a distinction between artifacts and living things. This result does need to be interpreted with caution, however. Even adults could legitimately consider milk and salt to be human made, in that both are prepared by humans before being consumed.

Justifications. After answering "yes" or "no," each child was asked either "How?" or "Why not?", respectively. These justifications were written down and coded as to their content (see Table 5). A second coder rated 25% of the justifications, yielding 91% agreement. Disagreements were resolved by discussion.

Four-year-olds had some difficulty articulating their answers, as noted by the large proportion of anomalous or nonexistent justifications. Furthermore, the details of children's answers at both ages were often incorrect. For example, one preschooler said that telephones are made "with wood and tape"; several second graders explained that salt grows or that it comes from a plant. Given children's lack of precise knowledge concerning these categories, it is all the more impressive that they are aware

TABLE 5
Study 3: Number of Justifications as a Function of Naturalness and Age

	Preschool		Grade 2	
	Natural kinds	Artifacts	Natural kinds	Artifacts
Characteristic of living thing	53[a]	0	105[a]	3
Made by God	14[a]	2	12[a]	0
Parts, substance, ingredients	10	30[a]	6	78[a]
Tool or process	1	25[a]	10	107[a]
Other	79	89	29	14

[a] Category type predicted to be appropriate.

that artifacts are made by humans, whereas certain natural kinds (animals and vegetation) are not human made.

GENERAL DISCUSSION

Children draw inductive inferences based on category membership from an early age. The fact that children made as many inferences in these studies as they did is consistent with previous research (Gelman & Markman, 1986) and is itself important, given the unfamiliarity of the properties and the fact that each property was taught about a single picture only. Yet children readily drew inferences, sometimes generalizing from a single instance to all or most of the category. For example, when shown a picture of a fish and told it needed branchias to breathe, one second grader replied, "All fish need branchias to breathe, probably." Another second grader, when told that a certain rabbit had an omentum inside, replied, "If one of 'em has it, all of 'em have to have it." In fact, when asked about a single instance, children often talked about an unspecified plurality of objects (e.g., "They [ants] bite you"). Children invoked category labels to justify their inferences and drew inferences despite marked dissimilarities in appearances (see Study 1, "Justifications"). Given the ease with which children generalize to category members from a single fact, they face the problem of induction: a need for constraints.

Two factors influence the extent of induction from the preschool age: category homogeneity and property generalizability. The importance of homogeneity implies that categories which are richly structured in their known properties (homogeneous categories) are assumed to capture many unknown properties as well. Conversely, categories that share relatively many dissimilarities on the surface are less likely to promote inductive inferences. Reliance on homogeneity may be a sensible default strategy. It is probably often the case, for many object categories, that perceived homogeneity roughly predicts homogeneity of underlying properties (such as internal parts and subtle behaviors). For a person with limited knowledge, such as a preschooler, homogeneity could serve as a useful (though fallible) heuristic. Property generalizability is also very relevant to preschoolers. Nongeneralizable properties such as transient or accidental facts are assumed not to follow "sensible" or lawlike patterns of induction. Note that the distinctions made by preschoolers indicate an early sensitivity to aspects of both the type of category and the type of property. The early acquisition of such constraints suggests that preschool children limit their inductive inferences in basically correct ways. These two factors can be considered domain general, in that they do not take into account the specifics of the category or property content. Thus, no particular type of category has a special status at this age.

Despite the importance of domain-general constraints on induction, from other research we know that preschool children override similarity in certain telling cases. One case that has already been mentioned is when category identity (at a basic level) conflicts with surface similarity (Gelman *et al.*, 1986; Gelman & Markman, 1986, 1987). A second case is presented by Carey (1985), who demonstrates that children use more than surface similarity to structure their inferences within the superordinate categories "animal" and "living thing." At times children's inferences directly conflicted with perceptual similarity; e.g., children were more likely to draw inferences from humans to stink-bugs than from bees to stink-bugs, despite the greater similarity within the latter pair. In fact, their inferences from humans to stink-bugs (a superordinate-level inference) is quite high, although presumably no higher than inferences would be within the category of humans. Carey's results do not conflict with those presented here, since these studies did not examine homogeneity within superordinate-level categories. However, her data can be used to extend our interpretation of the present results. Although preschool children appear to be more dependent on domain-general factors than are older children, they are also sensitive to the particulars of the category when drawing inferences. Thus, any theory of induction for preschoolers must include three factors: (a) attention to homogeneity and property generalizability, perhaps as default heuristics; (b) attention to category labels, such that inferences within categories are favored over inferences between categories; and (c) some reliance on domain-specific knowledge or beliefs (e.g., of humans as the prototypical animal).

Despite the initial headway that preschoolers have made, the present studies suggest that their inferences are not sufficiently constrained, nor are they sufficiently sensitive to domain-specific considerations. There is no consistent natural kind–artifact distinction at this age. To 4-year-olds, two apples are no more likely to have the same internal parts than a golf ball and a football. Likewise, preschoolers honor no consistent differences based on the content of generalizable properties. They draw inferences concerning internal parts and function equally often, for both natural kinds and artifacts.

The conclusion that there is no consistent natural kind–artifact distinction among preschoolers does need to be viewed with caution, given the difficulty of assuming that children and adults interpreted the properties in the same way. For example, "This chair has urethane inside" could be interpreted as, "This chair has substance inside," which is in fact true of most chairs. Thus, depending on how children interpreted these properties, it is possible that a natural kind–artifact distinction would not always be appropriate. However, there are three reasons to place confidence in the developmental difference that emerged: (a) The equally generalizable properties are more specific than either the generalizable for

natural kinds or the generalizable for artifacts properties and so more highly constrain subjects' interpretations of the novel words. Accordingly, these properties should be less open to alternative interpretations. Nonetheless the same developmental pattern emerged for these properties as for the others: preschoolers did not significantly distinguish between natural kinds and artifacts; second graders did. (b) In Study 2 no unfamiliar words were used, and children generated the properties themselves, thus presumably reducing possible misinterpretations of the properties. Again, the same developmental pattern emerged. (c) In a follow-up study by Gelman and O'Reilly (in press), children were asked open-ended questions concerning internal parts (e.g., "Do all dogs [spiders, chairs, vacuums, etc.] have the same kinds of stuff inside?"). These questions were designed to eliminate unfamiliar words which could be interpreted in idiosyncratic ways. The results again were that preschool children did not distinguish between natural kinds and artifacts, whereas second graders did.

Development is rapid within the brief period from preschool to second grade. Most strikingly, second graders make principled distinctions among categories and properties that younger children do not honor. The major accomplishment is that older children draw more inferences within natural kinds (e.g., carrots) than within artifact categories (e.g., balls), even with properties that were selected to be equally generalizable for natural kinds and artifacts, and even with functional properties that were designed to generalize from artifacts only. In fact, the older children maintain that familiar properties, designed to be equally true of natural kinds and artifacts, hold true more often for natural kinds than artifacts (Study 2). Their natural kind–artifact distinction is a tenacious one. Given the above-mentioned difficulties with controlling for children's interpretations of the properties, it is all the more impressive that second-grade children do clearly draw more inferences from natural kinds than from artifacts. Furthermore, second graders are sensitive to property content. When considering artifact categories, older children draw fewest inferences concerning properties generalizable for natural kinds. And the natural kind–artifact difference is smallest with properties that are generalizable for artifacts. These distinctions are on the whole appropriate.

One can characterize the developmental shift as one from relatively domain-general to relatively domain-specific considerations. Younger children rely on domain-general strategies more heavily, perhaps because of their relative lack of knowledge on which to draw for making distinctions among categories and properties. By second grade, inductive inferences are sensitive to the nature of the category, the domain, and the property being taught. Thus, it seems that domain-specific knowledge and beliefs are underlying these patterns of inferences. It may be that children at both ages (as well as adults) use domain-specific knowledge

when available, but rely on domain-general strategies (such as homogeneity) as a default.

What is the cause of developmental change? It is not simply that natural kinds have developed a special status in children's inductive inferences. The claim is not that the literal distinction (i.e., whether or not an object is naturally occurring) is the psychologically relevant variable. Note that children at both ages draw relatively few inferences to substances. Furthermore, preschoolers grasp the literal natural kind–artifact distinction (see Study 3) but do not make use of it when drawing inductive inferences. Rather, I propose that natural kind categories promote more inferences because for second graders they tie into more richly articulated theories: (1) they are perceived as being more homogeneous in their as-yet-unknown properties, because they are more limited than artifacts in their structure and makeup; and (2) they tend to tie into richer domains of scientific knowledge. As children's theories change, so do the sophistication and complexity of their inferences.

In support of this claim, let us consider what children are learning about natural kinds by this age. Between preschool and second grade, children learn that animals and plants support rich causal explanations that artifacts do not. For example, by the age that children show a reliable natural kind–artifact distinction in their inductions, children have also worked out that members of a natural kind must have the same internal structure, whereas members of an artifact category need not (Keil, 1986). Furthermore, children are learning that living things need certain properties in order to support life and produce offspring (Carey, 1985). Artifacts certainly do not share these properties and tend not to support analogously theory-rich, category-general facts. In other words, only the older children in these studies are beginning to understand animals and plants as distinctly biological categories (thus being explainable by a distinctive set of laws). Thus, increasing knowledge yields differentiation among categories: animals and plants afford rich causal explanations; artifacts do not. Younger children instead seem to treat all categories as explainable by the same kinds of laws.

Another indication that the advantage for natural kinds emerges from their status as more highly developed theoretical constructs is that for natural kinds, fewest inferences are drawn to the substances. Correspondingly, knowledge about substances seems especially late developing (Smith et al., 1985; Study 3 of the present paper). It is interesting that substances promote so few inductions for children, given that adults draw *more* inferences within categories of substances than categories of plants and animals (S. Gelman & E. M. Markman, unpublished data).

Clearly, more research is needed beyond these examples. However, it seems reasonable to propose that one major developmental change from preschool to second grade tapped in these studies is an increased under-

standing of biology. The present data support Carey (1985) and Keil (1986) in suggesting that the very status of basic-level biological categories (e.g., rabbit, tree) is changing with increasing scientific knowledge. More generally, I am suggesting that emergence of new theoretical domains leads to more powerful, more highly differentiated inductive inferences.

In addition to what it can tell us about the organization of children's knowledge, the present work has implications for how to consider categories more generally. In the psychological literature most views of categories focus on known properties and how they are represented. Often categories are assumed to be static, unchanging, and fully formed, with meanings that are explicitly known to all adults in the community. In contrast, the present studies demonstrate that categories are open ended in the sense of allowing for new information and that they can eventually dovetail with scientific knowledge. Comparison of the present work with others (Carey, 1985; Keil, 1986) suggests that by second grade, growing scientific expertise becomes important to how children organize knowledge and draw inferences. And even earlier, children expect categories to capture more than obvious properties. A full account of what makes something a useful category and how categories function will have to deal with these phenomena.

REFERENCES

Barr, R. A., & Caplan, L. J. (1983). *Differences in semantic category representation for natural and artifactual categories*. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.

Caplan, L. J., & Barr, R. A. (1985). *Static and transformational aspects of children's understanding of categories*. Paper presented at the annual symposium of the Jean Piaget Society, Philadelphia.

Carey, S. (1982). Semantic development: The state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 347–389). Cambridge, MA: Cambridge Univ. Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.

Flavell, J. H. (1985). *Cognitive development* (2nd ed.). Englewood Cliffs, NJ: Prentice–Hall.

Gelman, R., & Spelke, E. S. (1981). The development of thoughts about animate and inanimate objects: implications for research on social cognition. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development* (pp. 43–66). Cambridge, MA: Cambridge Univ. Press.

Gelman, R., Spelke, E. S., & Meck, E. (1983). What preschoolers know about animate and inanimate objects. In D. Rogers & J. A. Sloboda (Eds.), *The acquisition of symbolic skills* (pp. 297–326). New York: Plenum.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, 57, 396–404.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183–209.

Gelman, S. A., & Markman, E. M. Young children's inductions from natural kinds: The role of categories and appearances. *Child Development,* in press.

Gelman, S. A., & O'Reilly, A. W. (1987). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development,* in press.

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard Univ. Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: Bradford Books/MIT Press.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective.* Cambridge, MA: Harvard Univ. Press.

Keil, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), *Language learning and concept acquisition* (pp. 133–153). Norwood, NJ: Ablex.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23,* 221–236.

Malt, B. C. (1985). Hedges and the mental representation of categories. *CC-AI, 2,* 13–23.

Mansfield, A. F. (1977). Semantic organization in the young child: Evidence for the development of semantic feature systems. *Journal of Experimental Child Psychology, 23,* 57–77.

Markman, E. M., & Callanan, M. A. (1984). An analysis of hierarchical classification. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 325–365). Hillsdale, NJ: Erlbaum.

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

Nelson, K. (1974). Variations in children's concepts by age and category. *Child Development, 45,* 577–584.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Piaget, J. (1929). *The child's conception of the world.* London: Routledge & Kegan Paul.

Quine, W. V. (1969). Natural kinds. In *Ontological relativity and other essays* (pp. 114–138). New York: Columbia Univ. Press.

Rips, L. J. (1975). Induction about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.

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

Rosch, E. H., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 3,* 382–439.

Schwartz, S. P. (Ed.) (1977). *Naming, necessity and natural kinds.* Ithaca, NY: Cornell Univ. Press.

Schwartz, S. P. (1979). Natural kind terms. *Cognition, 7,* 301–315.

Simon, H. A. (1981). *The sciences of the artificial* (2nd ed.). Cambridge, MA: MIT Press.

Skyrms, B. (1975). *Choice and chance: An introduction to inductive logic.* Encino, CA: Dickenson. 2nd ed.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition, 21,* 177–237.

Sternberg, R. J. (1982). Natural, unnatural, and supernatural concepts. *Cognitive Psychology, 14,* 451–488.

Wellman, H. M., & Gelman, S. A. Children's understanding of the nonobvious. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale, NJ: Erlbaum, in press.