Title
The dynamics of preschoolers' categorization choices

Journal
Child Development, 67

Authors
Deak, Gedeon O
Bauer, Patricia J.

Publication Date
1996

DOI
0009-3920/96/6703-0001

Peer reviewed

The Dynamics of Preschoolers' Categorization Choices

Gedeon O. Deák and Patricia J. Bauer

University of Minnesota

DEÁK, GEDEON O., and BAUER, PATRICIA J. *The Dynamics of Preschoolers' Categorization Choices.* CHILD DEVELOPMENT, 1996, 67, 740–767. The present research explored the effects of stimulus and task factors on preschoolers' (Experiments 1 and 3) and adults' (Experiment 2) tendency to categorize according to taxonomic relations, when those relations conflict with appearances. In Experiment 1, we examined the effects of and interactions among (*a*) available information, operationalized by using more- or less-informative stimulus types (objects vs. line drawings) and by the presence or absence of labeling, and (*b*) task constraints, operationalized by comparing sorting questions with inductive inferences questions. When provided with information that constrained the categorization decision, either through the availability of labels or a combination of enhanced physical informativeness of objects and an inference question, preschoolers reliably based their categorization decisions on taxonomic relations between physically dissimilar items. In Experiment 2, stimulus type (objects vs. line drawings) was shown to have a similar effect on adults. In Experiment 3, we examined the effects of stimulus type on preschoolers' inductive inferences and accuracy of naming. The effects in the two tasks were closely related, suggesting that the amount of available information affects different responses in similar ways. These data demonstrate the interactive effects of available information and task constraints on categorization decisions.

Traditional theories of cognitive development posit a striking age-related change in the criteria used to categorize objects: preschoolers group or sort objects on the basis of similarity of salient physical features (e.g., overall shape and color), whereas older children and adults categorize on the basis of more abstract, symbolic, or conceptually important attributes, even if those attributes are nonobvious (e.g., Bruner, Olver, & Greenfield, 1966; Flavell, 1985; Inhelder & Piaget, 1964; Kendler & Kendler, 1975; Vygotsky, 1934/1986; Werner, 1957). The argument that preschoolers cannot categorize on any basis other than overt perceptual similarity supported the characterization of preschoolers as "perceptually bound" relative to adults. This belief in preschoolers' perceptual boundedness suffused early work on conceptual development and has persisted in more recent work (e.g., Fenson, Cameron, & Kennedy, 1988; Gentner, 1978, 1989; Melkman, Tversky, & Baratz, 1981; Tomikawa & Dodd, 1980; Tversky, 1985).

In spite of its prevalence, the traditional view has both empirical and conceptual shortcomings. Empirically, many studies challenge this view of development. Carey (1985), Deák and Bauer (1995), Deák and Pick (1994), Gelman and her colleagues (e.g., Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Kalish & Gelman, 1992), Kemler-Nelson (1991), Massey and Gelman (1988), and Miller (1973) have demonstrated that preschoolers are not limited to categorizing according to overall physical similarity. Other findings indicate that adults categorize according to overall physical similarity in certain circumstances (e.g., Smith & Kemler-Nelson, 1984; Smith & Shapiro, 1989). These results cast doubt upon

This research was supported by a University of Minnesota graduate school fellowship to the first author, NICHD FIRST award (HD28425) to the second author, and NICHD training grant (HD-07151) to the Center for Research in Learning, Perception, and Cognition at the University of Minnesota. Parts of this research were presented at the Annual Symposium of the Jean Piaget Society, Philadelphia, June 1993, and at the biennial meeting of the Society for Research in Child Development, Indianapolis, April 1995. The authors thank Louise Hertsgaard, Charles Nelson, Anne Pick, Lisa Travis, Leah Welch, and several reviewers for their helpful comments, Kirsten Condry for drawing the stimuli, and the adults, children, and parents who took part in this research. Address correspondences to the first author at the Department of Psychology and Human Development, Box 512 Peabody College, Vanderbilt University, Nashville, TN 37203 or by e-mail to deakgo@ctrvax.vanderbilt.edu.

the developmental progression asserted by the traditional view. Conceptually, the traditional view is no less problematic. The view assumes a dichotomy between perception and cognition, since categorization decisions are conceived as being made *either* on the basis of perception *or* on the basis of cognition: young children categorize using perceptual attributes, whereas older children and adults categorize using conceptual attributes. The assumption that perception and cognition are separable is difficult to support (Deák, 1994), a point sufficiently important to merit discussion.

A major reason that the perception versus cognition dichotomy is questionable is that many behaviors cannot easily be labeled "perception" or "cognition." It is true that some behaviors, such as judging the validity of logical syllogisms or solving calculus problems, seem predominantly abstract and symbolic and can be classified as cognition. Other behaviors, such as categorical perception of color or orienting toward a loud, unexpected sound, seem relatively fast and automatic and can be classified as perception. However, attempting to describe many behaviors as cognition or perception per se is wrought with difficulty. Object recognition (Beiderman, 1988), depth perception via pictorial cues (Yonas, Arterberry, & Granrud, 1987), subitization (Starkey & Cooper, 1980), infants' discrimination of physically "impossible" events (e.g., Baillargeon, 1987), and perception of invariants (Bahrick, 1983) are behaviors that do not fit neatly into a framework in which perception and cognition are dichotomous. Thus, some behaviors might easily be classified as cognition and others as perception, but a wide range of behaviors cannot easily be classified.

The inability to easily classify a wide range of behaviors is probably related to the problem of identifying substantial bases for the perception versus cognition dichotomy. For example, one hypothetical basis for the dichotomy is impermeability of information, or modularity, between processes deemed perceptual and those deemed cognitive. At the very least, this criterion demands either (*a*) a one-way flow of information, so that processing of one type affects the other, but not vice versa; or (*b*) relatively less permeability *between* perceptual and cognitive processes than *within* either process. The first possibility is clearly false, as demonstrated by "top-down" processing (e.g., Warren, 1970). The second possibility seems indefensible upon objective consideration of the empirical literature: although there is some degree of impermeability between, for example, visual information and knowledge (e.g., perceiving the Müller-Lyer illusion despite knowing that the lines are the same length), there is just as clearly some impermeability *within* nominally perceptual processes (e.g., swaying in a swinging room, denoting an inability to reconcile visual and vestibular information, Lee & Aronson, 1974) and within nominally cognitive processes (e.g., semantic processing without conscious awareness; MacKay, 1973). Currently there is no way to determine whether there is less permeability of information *between* processes deemed perceptual and cognitive than *within* each of these processes. Other potential bases for the perception versus cognition dichotomy—including neurological explanations—are similarly problematic (Deák, 1994). In general, it is extremely difficult to justify the dichotomy.

The dichotomy between perception and cognition has profoundly affected the way researchers frame developmental phenomena and study development. Its influence is apparent in the traditional research agenda of charting preschoolers' release from perceptual boundedness; however, its influence is also obvious in the work of investigators critical of the characterization of preschoolers as perceptually bound. For example, some researchers have suggested that infants' event discriminations and expectancies are driven by symbolic reasoning (e.g., Baillargeon, 1993; Leslie, 1988; Spelke, 1991), and others suggest that age-related changes in domain-specific judgments are due to changes in children's theories (e.g., Carey, 1993; Wellman & Gelman, 1988, 1992). To take a more specific example, Gelman and Markman (1986, 1987) asked whether preschoolers use labels (deemed conceptual information) or overall appearance (deemed perceptual information) to make categorical inferences. This distinction is somewhat artificial, because a spoken label is, among other things, perceptual information, so providing identical labels for to-be-classified items may obviate thinking about the conceptual relations between the items.[1] Furthermore, even if la-

[1] The distinction also can be challenged on the grounds that the use of overall appearance for categorization choices might result from learning about the usefulness of overall appearances in such decisions.

bels do not provide a direct basis for categorization, they may alert subjects to look for *physical* details that support inferences about objects' nonobvious similarities. Such operationalizations of the perception versus cognition dichotomy imply that subjects use *either* one kind of information *or* the other. These frameworks do not readily accommodate evidence that preschoolers use many kinds of information for categorization (e.g., Deák & Bauer, 1995; Deák & Pick, 1994; Jones, Smith, & Landau, 1991; Kalish & Gelman, 1992; Kemler-Nelson, 1991; Miller, 1973).

Given these concerns, instead of asking when and how preschoolers overcome perceptual boundedness, we seek to determine the circumstances in which subjects of different ages use different kinds of information to make categorization decisions. We begin with the (empirically derived) assumption that children can use many kinds of information to make categorization decisions, and the information used will depend on the goals of a particular categorization decision. Accordingly, rather than asking whether children use, for example, labels *or* appearance, we ask under what circumstances labels, appearance, or both are used by the child to make a categorization decision. This question is particularly important in situations, such as those used in experimental studies, in which there are several legitimate ways to categorize a set (e.g., either according to overall appearances or according to theoretically driven taxonomic relations). This approach allows for a more veridical account of what children are doing, and why they are doing it, when they make a categorization decision. To generate such an account, we will refer to a descriptive taxonomy including three nonorthogonal factors: the *information available* to the subject, constraints of the *experimental task*, and characteristics of the *subject*. We will use this taxonomy to specify the dynamics of preschoolers' categorization choices and their use of various types of information. Three variables representing two of these factors—*information* and *task*—are discussed below.

One critical aspect of the information available to subjects is the informativeness of the stimuli. It is striking that most studies of preschoolers' categorization have used two-dimensional stimuli (e.g., colored line drawings) that are impoverished compared to the objects that occupy our environment. As a result, subjects in these studies have limited access to the subtle information that often is diagnostic of principled relations between complex objects. For example, Gelman and Markman (1987) presented preschoolers with pictures of objects drawn so that same-category objects (e.g., two cats) sometimes looked less similar than different-category objects (e.g., a cat and a skunk). In the absence of labels, preschoolers were more likely to extend categorical inferences to similar-looking pictures than to dissimilar-looking pictures, even when the similar-looking pictures depicted different kinds of things (cats and skunks). Although cats may be drawn to look like skunks, or bats like blackbirds, real cats and skunks and bats and birds have physical similarities and dissimilarities that are diagnostic of their relations (e.g., blackbirds and flamingos, but not bats, have feathers). These similarities are not usually captured in line drawings. As a result, preschoolers might categorize drawings on the basis of overall appearance, which is a perfectly reasonable response because overall appearance (i.e., shape and color) is usually highly diagnostic of other, nonobvious similarities. In contrast, providing more informative real objects or detailed three-dimensional representations might allow preschoolers to attend to the subtle physical regularities that denote taxonomic relations and, thereby, categorize in accordance with these relations.[2]

[2] The term "taxonomic" (or "taxonomically related") has both a more general and a more specific meaning, which often causes confusion. Under the more general meaning, an inclusion relation is defined by any attribute shared by several entities. Thus, any predicate, such as "has a stripe on its back," "is a liquid," or "tends to keep a messy desk," specifies a taxonomic category. However, this general meaning allows an infinite number of taxonomic categories that are useful only in specific situations and that tend not to be encoded as count nouns in natural languages. The more specific meaning (the modal meaning in cognitive psychology) defines categories that tend to be more generally useful and are encoded as count nouns in natural languages. It rests on specific relations within a scientific taxonomy for natural kinds (e.g., lines of evolutionary descent in zoology), and similar (intended) functions for artifacts. It is this more specific meaning that is the basis for taxonomic relations in the stimuli for the studies reported here.

What features of real objects are present in or absent from line drawings? Line drawings tend to preserve the overall two-dimensional contour (shape) of objects and, if colored, the approximate dominant color(s). Internal lines might also suggest internal contours or markings. Thus, line drawings tend to portray those attributes that underlie judgments of overall similarity, namely, shape and color. Line drawings typically exclude features such as texture (although some texture can be suggested by the sharpness and straightness of sides and edges), reflectance, and translucence. Similarly, they cannot convey information about weight or density, nor can they convey information about smell, taste, or sound. Finally, because line drawings are static representations, no dynamic information is available from drawings, and drawings cannot be viewed from different perspectives, as can objects. Notice that some features absent in drawings may be inferred by calling upon preexisting knowledge: for instance, a picture of a dinosaur may be judged to represent something with scales. However, this is based on the subject's knowledge of dinosaurs, not on information in the drawing.

The relative lack of information in drawings might affect categorization judgments in at least two ways. First, the absence of particular features will make some categorization judgments difficult or impossible. This might occur when a question implies attributes that are unavailable in the stimuli. For example, if drawings of a blackbird and a flamingo do not portray feathers, than the feature of having feathers cannot play into a child's inductive inferences (unless, of course, the child makes separate inferences that both drawings depict birds, and birds have feathers). Considering another example, questions implying material composition may be difficult to answer because drawings typically lack most features (e.g., texture, reflectance, weight, density) that are diagnostic of substance.

Second, line drawings might be more difficult to identify or name, or to identify with great specificity, than objects. For example, if the bat and the blackbird cannot be differentiated as members of distinct species, children cannot base decisions on this difference. Note that these two effects are not unrelated—for example, children might have difficulty identifying drawings because drawings lack features that also play a role in children's inferences about those items.

Whereas the informativeness of stimuli used in categorization tasks might be a major determinant of performance, other available information might compensate for less-informative materials. Even with impoverished stimuli, researchers have found that preschoolers can use labels to classify items. Gelman and Markman (1986, 1987; see also Markman & Hutchinson, 1984; Waxman & Kosowski, 1990) report that labels are readily used to resolve ambiguous categorization decisions. Yet presenting impoverished stimuli might have compelled subjects to rely on labels to make categorization decisions. This possibility is illustrated by a thought experiment suggested by Smith and Heise (1992): place several closed boxes in front of a preschooler and label their (hypothetical) contents so that a subset of the boxes is given the same name. If asked to categorize the boxes, the child certainly would group the boxes that were given the same label. The point of the thought experiment is that subjects can match labels without regard to any other properties of the actual items, and presenting labels makes it difficult to determine what other information preschoolers can use to make categorization decisions (however, see Gelman & Markman, 1986, Study 2). Of course, the ease with which children use labels implies the importance of labels in categorization. Names are a salient source of information about the world—children request names for novel objects, and hearing the same label for several objects might lead children to look for similarities among those objects. Thus, it behooves us to examine the role of labels relative to other factors affecting categorization and to explore the effectiveness of labels relative to other kinds of information.

The availability of (sometimes subtle) physical information may be necessary for making a taxonomic-based categorization decision when labels are unavailable. However, such information may not be sufficient to make a principled decision when the subject does not know what information to weight heavily and attend to in a given situation. For example, a preschooler must attend to certain features to distinguish cetacean mammals from fish, but different features to distinguish reptiles from amphibians. Individuals need some direction or principle to determine *how* to prioritize information, and children might be more likely to attend to specific kinds of information, no matter how subtle, when a task is sufficiently specific to provide such direction. In most everyday situations this direction is implicit in the goal of the task at hand, but it might not be evident in experimental tasks, particularly those that present two or more alternative categorization choices. Perhaps differences in the goal directedness or specificity of experimental tasks affect subjects' abilities to select pertinent information.

As an illustration, consider two variations of the often-used triad oddity task. In both variants subjects are shown a trio of stimuli consisting of a Standard, a Target, and a Distracter. Their task is to match either the Target or the Distracter to the Standard. The Target and Standard are related taxonomically (i.e., in some abstract, principled way, e.g., by function or by evolutionary descent), but are dissimilar in overall appearance. The Distracter and Standard are similar in overall appearance, but their conceptual relation is more distant and less compelling (we use the term "Distracter" for simplicity, not to imply an incorrect choice). In the first variation, a taxonomic categorization task, a child might be shown a round white seashell (Standard), a round white rock (Distracter), and a spiral striped seashell (Target) and asked a general question designed to elicit a taxonomic-based response, such as, "Which one of these [Target or Distracter] is the same kind of thing as this one [Standard]?" In this case the child is presented with two reasonable choices and has difficulty determining which information to use, since the question provides few clues about what features are relevant to the categorization task. In the second variation of the oddity task, inductive inference questions such as those used by Carey (1985) and Gelman and Markman (1986, 1987) might constrain the categorization decision by providing clues about which aspects of the stimulus are important. In the inference task, children are told a fact about the Standard and asked whether that fact applies to the Target or the Distracter. For example, the child might be told, "This one [round white seashell] has calcium in it," and asked, "Which one of these [spiral striped shell or round white rock] also has calcium in it?" In this case the child might observe that the shells share subtle physical similarities involving material composition (e.g., both are smooth, lightweight, slightly translucent, and have similar reflectance) and use this information to make an inference linking the two shells.[3] This inference is a categorization choice because it entails grouping together two discriminably different items by virtue of some shared similarity. If the inference task effectively allows the child to narrow down the features relevant to a categorization decision, then children in this task might be expected to make more categorization choices consistent with taxonomic relations, compared to children tested in a more traditional (i.e., taxonomic question) task. (Notice, in this example, that

[3] We define categorization as treating two or more discriminably different entities as identical in some way or for some purpose. This definition, while somewhat broad, avoids arbitrary distinctions between different kinds of generalizations. By this definition, inductive inference is a categorization decision, as is sorting or naming. In inductive inferences, similarities between entities are used to make judgments about the likelihood that they share other, unknown attributes. These similarities may be shared names, physical characteristics (however subtle), roles in events, etc. Subjects choosing the Target item may draw inferences on the basis of shared (subtle) attributes or may explicitly consider the taxonomic connection between the Target and Standard item; we do not assume that these processes are qualitatively different, and furthermore it is extremely difficult in almost any circumstances to distinguish them. If children use texture, weight, and reflectance to infer that two different-looking pieces of soap are both "made from lye," it may be because the children conceptualize a taxonomic relation among different kinds of soap, or it may simply be because the children apprehend that the items are made from the same substance. However, because the function of soap depends on its substance, the taxonomic relation depends on material. Thus children are either using their apprehension of similar material to make an inference about the taxonomic relation or making an inference on the basis of the attribute (material) that determines the taxonomic relation. This distinction is interesting, but it is not central to determining the circumstances in which children make a categorization decision *consistent* with taxonomic relations, when those relations conflict with appearances. However, it seems unlikely that a participant would categorize on the basis of a single feature such as texture, rather than overall appearance (which is usually a good indicator of taxonomic relations), unless the participant held some belief about the importance of the feature with regard to the decision at hand. Thus, it is likely that when subjects base their choices on such a feature, they believe the feature is diagnostic of a more pervasive relationship, such as shared membership in a natural-language taxon.

the features a child might use to make judgments about material are largely absent from line drawings.)

In the experiments reported here, we investigated the dynamics of preschoolers' categorization decisions. In Experiment 1, we examined the separate and joint effects of variables representing two of the factors discussed above—available information (operationalized by varying stimulus type and labeling) and experimental task (operationalized by varying the questions asked of children)—on preschoolers' categorization choices. Significantly, the above principles are not limited to preschoolers. For example, the information available for categorization decisions should affect older subjects, including adults. In Experiment 2, we tested this possibility by examining the role of stimulus type on adults' categorization decisions (the variable of age represents the third factor, subject characteristics). In both experiments, we used a triad oddity task to investigate subjects' categorization choices. As described above, subjects in this task are asked to match either a Target or a Distracter to a Standard item. In addition to "conflict" trios, in which the Standard is related to the Target but looks like the Distracter, we included a set of "no-conflict" trios, in which the Standard item was both more closely related *and* more similar in appearance to the Target than to the Distracter. An example is a bull (Standard), a cow (Target), and a coyote (Distracter).

Employing no-conflict trios as a within-subjects control is important to our design for several reasons. First, it allows us to determine whether subjects are choosing randomly. This will be particularly important if the percentage of conflict trio Target choices is not different than chance (50%), because we would then wish to determine whether children were choosing randomly from *all* trios. Children should choose Target items from the no-conflict trios at ceiling, no matter how they perform on the conflict trios; if they do not (i.e., if they also choose from the no-conflict trios at chance levels) it would suggest that they were inattentive or did not understand the task. Furthermore, including no-conflict trios allows us to determine whether children are using an unexpected response strategy, in the event that they mostly choose Target items from the conflict trios. For example, if children choose mostly conflict-trio Target items, it might be because they used a rule such as "choose the one that looks different." However, if children primarily choose the Target items from the conflict trios *and* the no-conflict trios, they cannot be following such a strategy. Finally, no-conflict trios mitigate the artificiality of the experiment, since naturally occurring conflicting categorization problems might be somewhat unusual.

Experiment 1

We examined the effects of three variables on preschoolers' categorization choices. First, we tested half of our child subjects using three-dimensional objects (either real objects such as seashells or detailed representations such as plastic dinosaurs); we tested the other half using pen and watercolor drawings of the same objects (see "Method" below). This allowed us to assess the effect of availability of stimulus-based information on categorization choices. Because many studies of children's categorization abilities have used line drawings, it will also inform our interpretation of past research as well as the design of future research. Finally, it will provide general information about the possible effects of using line drawings and other impoverished stimuli to assess children's abilities to make complex decisions.

In order to make valid comparisons between objects and drawings, it is necessary to equate their psychophysical properties. Specifically, it is necessary to demonstrate that each conflict trio's Standard and Distracter are perceived as more similar in overall appearance that the Standard and Target. Similarly, each no-conflict trio's Standard and Target should be more similar in appearance than the Standard and Distracter. Furthermore, the magnitude of these inequalities should be similar for object trios and drawing trios, so that children do not make appearance-based choices in response to drawings simply because the Standard-Distracter similarity was greater for drawings than for objects. To demonstrate these patterns our stimulus trios were pretested on adult participants. In addition, it is desirable to demonstrate that our drawings were not systematically misleading. That is, because two items in every trio were selected to be similar in appearance, it is important to show that each drawing actually looks more like the particular object it was designed to represent than the similar-appearance object from the same trio. To demonstrate this, a separate group of adults was asked to match pictures to the objects they represent.

The second factor varied in this experiment was labeling: half of the children in each group heard the items labeled, the other half heard no labels. This variable was included because labels constitute important, socially provided information that might guide children's attention, and because of the possibility that labels minimize the need to actively compare items within a set.

Third, in each group we gave half of the children an inductive inference task and the other half a taxonomic question task. This is an important comparison because inductive inference questions tend to imply specific kinds of information that are relevant, whereas a more generic categorization question (e.g., "Which ones are the same kind of thing?") typically does not allow a very precise narrowing-down to determine which information is relevant. An inductive inference question, then, potentially constrains the information subjects will consider, and consequently guides categorization decisions, more than typical categorization questions. There is relatively little data, however, on young children's ability to utilize this task-based constraint (but see Kalish & Gelman, 1992). This is pertinent because many everyday categorization choices are specific and constrained; they are not as ambiguous as typical experimental categorization questions. In evaluating preschoolers' categorization abilities, then, we should explore the possible effects of question specificity. This comparison also bridges a gap in the categorization literature between studies using one or the other task: no study has compared children's performance on general sorting questions to performance on inductive inference questions.

In the inference condition each child was asked to make two judgments about each trio. In addition to the inductive inference question, each child was asked a second question, similar in format to the inductive inference question, but involving a fact known of the Target item. For example, for the bar of soap Target, the children were asked a question based on a known fact, "Which one do you use to wash your hands?" in addition to the inductive inference question, "Which one is made of lye?" We will refer to the two kinds of questions as "inductive inference" and "known." The merit of asking both kinds of questions is as follows: if children know that bars of soap are used for washing and they are told that the Standard is used for washing, they may make a decision merely by looking for another item that can be used for washing (i.e., another piece of soap)—they do not need to consider the similarities between the Standard and the other items. For this reason, we cannot be certain that responses to the known-fact questions involve categorization; the known questions thus eliminate task demands imposed by inductive inference questions while maintaining a similar format. Furthermore, a relevant comparison will be between the effects of labels versus known questions, since both may minimize or eliminate the need for active comparison between physical attributes of the items in a set and, therefore, might be expected to cause a similar change in performance from the no label/categorization question condition.

In contrast to the known fact questions, if children hear that the Standard is made of lye—necessitating an inductive inference—they must look for some connection between the Target and one of the other items and cannot use preexisting knowledge to decide between the Target and the Distracter. In this example, they must look for information specifying similarity of substance. This active search for theoretically relevant information with which to make an inference constitutes a categorization decision.

Consistent with the preceding discussion, we expected children to choose more Target items when viewing objects than when viewing line drawings. In all cases we expected labels, as well as known inference questions, to increase the probability of choosing Target items to very high levels compared to the probability of choosing unlabeled Target items in response to a categorization question. We also expected children to choose more Target items in the inductive inference condition than in the taxonomic question condition. Finally, we predicted that making a constrained judgment (inductive inference) about unlabeled, more-informative stimuli (objects) would elicit a rate of Target choices (those based on subtle information specifying taxonomic relations) substantially higher than that found in the drawings/no-label condition, and higher than that found in no-label conditions in related studies that used drawings (e.g., Gelman & Markman, 1987).

Method

Subjects.—Twenty-four adults (students and staff at a large metropolitan university) participated in the stimulus selection pretest, an additional 12 adults participated in the line drawing similarity rating pretest, and six adults participated in a drawing-to-object matching pretest (described below). Two groups of 12 4-year-olds each (mean ages 4-4 and 4-5, respectively) participated in the inference question selection pretest.

Ninety-six 4-year-olds (41 female, 55 male; mean age 4-8; range 4-5 to 4-11) were tested in one of eight experimental conditions ($n = 12$ per group) derived from a complete crossing of three between-subjects variables: stimulus type (object or line drawing), labeling (labels or no labels), and task (taxonomic question or inductive inference). One additional preschooler failed to complete the study. Children were recruited from an existing pool whose parents had responded to solicitation at the time of their children's births. Data on race, ethnicity, and SES were not collected. None of the children had participated in pretesting. All children were given a small gift for their participation.

Design.—Three factors were varied in training and testing (see "Procedure"). The first variable was stimulus type. Half of the children were shown *object* trios, and the other half were shown *line drawing* trios. The second variable was labeling. Half of the children in each stimulus-type group heard *labels* for all items in a trio before they made a judgment, the other half of the children heard *no labels*. The third variable was experimental task. Half of the children in each of the four labeling-by-stimulus-type groups answered a *taxonomic* question, whereas the other half answered *inference* and *known* questions.

Materials.—Experimental stimuli were trios of objects or pictures, each one consisting of a Standard, a Target, and a Distracter item (see Table 1). Children in the object conditions saw 19 trios (three training trios, 16 test trios) presented in 32 × 40 cm trays with the Target and Distracter closer to and on either side of the child, and the Standard in the center and closer to the experimenter. Object trios consisted of real objects except when it was not feasible to present a real token (e.g., cow). In these cases, detailed three-dimensional representations of real objects were used. Children in the line drawing conditions saw laminated pen and watercolor drawings mounted on 38 × 58 cm posterboard in a spatial arrangement similar to that of the object trios. Line drawings were executed by an artist instructed to accurately depict the stimulus objects. Photographs of sample object and line drawing conflict trios are shown in Figure 1.

Stimulus selection procedure.—Twenty-four adults, in two groups of 12, were asked to rate similarity of appearance for 28 trios of objects (this procedure is also reported in Deák & Bauer, 1995). One group of 12 subjects rated 24 candidate trios, 14 of which were conflict trios in which the Standard and Target were relatively closely taxonomically related (according to lines of evolutionary descent or intended function) but dissimilar in appearance, and the Standard and Distracter were more distantly related but similar in appearance. The remaining 12 candidate trios were no-conflict trios in which the Standard was both more closely related and more similar looking to the Target than the Distracter. In order to generate a sufficient number of conflict trios, a second group of 12 subjects rated an additional set of seven candidate conflict trios (as well as the no-conflict trios).

For each trio the adults separately rated the similarity of appearance of the Standard-Target pair and the Standard-Distracter pair. Subjects were instructed to ignore their knowledge about the objects and base their ratings only on overall appearance. Subjects rated 10 of the 16 conflict trios (eight rated by the first group of subjects, two rated by the second group) as having significantly greater similarity of appearance between the Standard-Distracter pair (overall $M = 6.8$ out of 10 on a Likert scale; SD = .80) than between the Standard-Target pair (overall $M = 3.2$; SD = .56); $p < .05$ for each of the trios by separate t tests ($df = 11$). Eight of these trios were selected as test trios; the other two were used for training (see "Training and testing procedure" below). One additional trio was retained as a training trio, although the difference between Standard-Target and Standard-Distracter appearance ratings only approached significance. Ratings of the remaining five trios differed in the predicted direction but did not achieve significance; these were not used. Adults rated all 12 no-conflict trios as having significantly greater similarity of appearance between the Standard-Target pair ($M = 7.4$ out of 10, SD = .59) than between the Standard-Distracter pair ($M = 2.7$, SD = .67); $p < .05$ for each of the trios by separate t tests ($df = 23$ for both groups; results do not differ if the groups' ratings are analyzed sepa-

TABLE 1

STIMULUS ITEMS, LABELS, AND INFERENCE AND KNOWN QUESTIONS

Standard	Target	Distracter
[Known question]/[Unknown inference question]		
Training trios:		
1. White collie ("dog")	Brown dachshund ("dog")	White Persian cat ("cat")
"Which one of these barks and chases cars?"		
2. Red convertible ("car")	Motorcycle ("motorcycle")	Red speedboat ("boat")
"Which one of these drives down the road?"		
3. Polar bear ("bear")	Brown bear ("bear")	White sheep ("sheep")
"Which one of these eats meat with its sharp teeth?"		
Conflict trios:		
1. Light bulb ("light bulb")	Ornamental bulb ("light bulb")	Juice bottle ("bottle")
"Which one of these can light up?"/"Which one of these has a filament?"		
2. Brown sugar ("sugar")	Sugar cubes ("sugar")	Brown dirt ("dirt")
"Which one melts when you put it in water?"/"Which one is made of glucose?"		
3. Flat, white shell ("shell")	Spiral, striped shell ("shell")	Flat, white rock ("rock")
"Which one came from the ocean?"/"Which one has calcium in it?"		
4. White scallop soap ("soap")	Green soap bar ("soap")	White scallop ("porcelain")
"Which one can you wash your hands with?"/"Which one is made of lye?"		
5. Black panther ("cat")	Tabby house cat ("cat")	Black stallion ("horse")
"Which one has claws and whiskers?"/"Which one is a kind of feline?"		
6. Purple cedar ball ("cedar")	Unpainted cedar egg ("cedar")	Purple bath bead ("bath bead")
"Which one is made out of wood?"/"Which one comes from a conifer?"		
7. Round, white balloon ("balloon")	Long, blue balloon ("balloon")	Round, white candle ("candle")
"Which one pops if you stick it with a pin?"/"Which one is made out of latex?"		
8. Triceratops ("dinosaur")	Stegosaurus ("dinosaur")	Rhinoceros ("rhinoceros")
"Which one lived a very long time ago?"/"Which one has cold blood?"		
No-conflict trios:		
1. Steel fork ("fork")	Steel and resin fork ("fork")	Plastic spoon ("spoon")
"Which one has points?"/"Which one can spear food?"		
2. Quartz crystal ("quartz")	Quartz crystal ("quartz")	Galena crystal ("galena")
"Which one can you see through?"/"Which one is a part of granite?"		
3. French horn ("horn")	Hunting horn ("horn")	Guitar ("guitar")
"Which one do you blow into?"/"Which one is made of brass?"		
4. Small orange pliers ("pliers")	Large orange pliers ("pliers")	Hex wrench ("wrench")
"Which one is the color of a carrot?"/"Which one can pull out nails?"		
5. Brown bull ("cow")	Spotted cow ("cow")	Coyote ("coyote")
"Which one says 'moo'?"/"Which one has a thurl?"		
6. Butterfly ("butterfly")	Butterfly ("butterfly")	Goose ("goose")
"Which one used to be a caterpillar?"/"Which one comes from a cocoon?"		
7. Orange ("orange")	Tangerine ("orange")	Grapes ("grapes")
"Which one do you have to peel to eat?"/"Which one can you get zest from?"		
8. Large paper clip ("paper clip")	Small paper clip ("paper clip")	Binder clip ("binder clip")
"Which one is silver and curvy?"/"Which one can you pry open with cardboard?"		

NOTE.—Terms in parentheses are experimenter-provided labels. The first question for every trio is the known question, the second is the inductive inference question.

rately). Eight of these 12 no-conflict trios were chosen to reflect a diverse range of items similar to the conflict trios. Thus, we generated a total of 16 test trios (eight conflict and eight no conflict), and three training trios. All trios are described in Table 1.

After line drawings of these trios were created (see "Materials" above), the perceptual similarities between pairs of line drawings within a trio were rated by 12 naive adults, using a similar procedure to that used in the perceptual similarity pretesting for objects. Subjects rated all drawings of conflict trios as having significantly greater similarity of appearance between the Standard-Distracter pair (M = 6.5, SD = .66) than between the Standard-Target pair (M = 3.2, SD = .85); differences for each of the trios were significant at $p < .05$ by t test (df = 11). Similarly, adults rated drawings of all eight no-conflict trios as having significantly greater similarity of appearance between the Standard-Target pair (M = 6.7, SD = .75) than between the Standard-Distracter pair (M = 2.8, SD = 1.27); differences for each trio were significant at $p < .05$ by t test (df = 11). It is important for our hypothesis that the pattern and extent of perceptual similarity among the drawings mirror that among the objects. This was tested by converting the difference scores (between the Standard-Target pair and the Standard-Distracter pair) for each trio into z scores (to control for possible heterogeneity of variance) for analysis. All trios—conflict and no conflict—were included in this computation. The average z scores for the conflict trios were −.89 and −.87 for the objects and line drawings, respectively. The difference between the conflict trio z scores for the two stimulus types was nonsignificant, $t(7) = .12$. Similarly, the average z scores for the no-conflict trios were .89 and .87 for the objects and line drawings, respectively. This no-conflict trio difference was also nonsignificant, $t(7) = .13$. With regard to adults' judgments of the perceptual similarity structure of stimulus trios, then, these line drawings and objects are not significantly different.

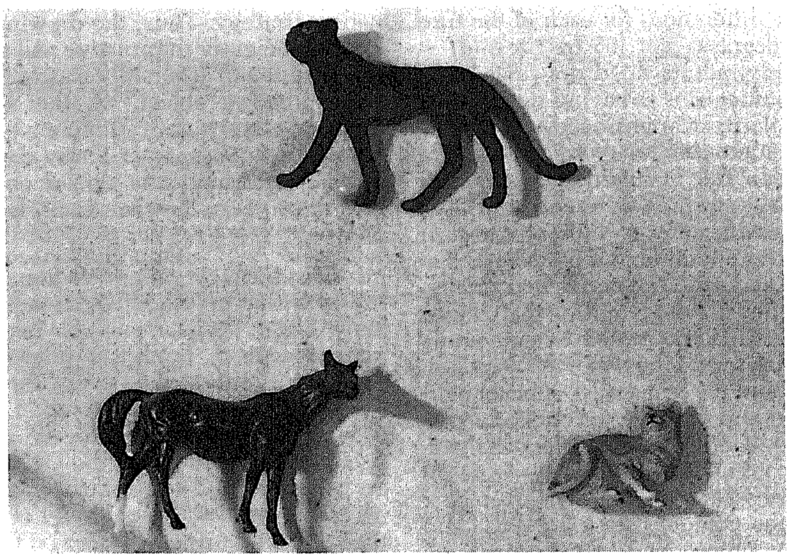
A final group of adults was tested to establish that the drawings are not systematically misleading, because if they are, the results could be biased in the direction of our hypothesis. In order to establish that our drawings were not systematically misleading, six adults were asked to match objects to drawings. Objects and drawings of every conflict-trio item (24 items), as well as four additional objects and four (different) additional drawings, were spread out in a haphazard arrangement. Subjects were asked to match objects to drawings by putting each object on top of the appropriate drawing. Adults matched an average of 23 out of 24 items correctly (96%, SD = 1.7 or 7%), which does not differ significantly from 100%, $t(5) = 1.48$, N.S. Thus, adults very accurately matched objects to drawings, in spite of the fact that they were presented with a large and confusing array, in spite of the fact that two items in each trio (Standard and Distracter) were judged to look highly similar (the few errors were switches of the Standard and Distracter within a trio), and in spite of the fact that the additional objects and drawings prevented the use of an elimination strategy. Thus, the drawings were at least sufficiently informative and accurate to allow adults to accurately match them to the corresponding object, whether or not the pictures were highly informative in and of themselves. This suggests that the drawings did not systematically distort or misrepresent the objects; they were merely less informative.

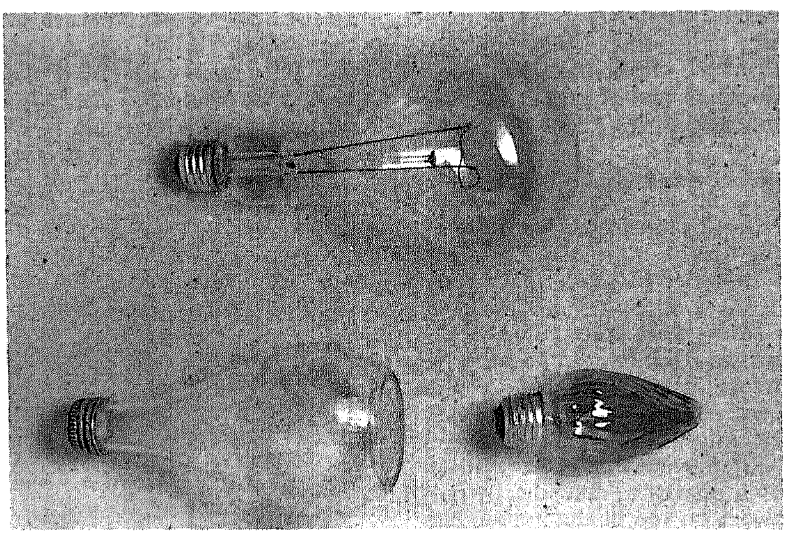
Question selection procedure.—To identify facts that preschoolers know and do not know of the Target items (for the inductive inference task), candidate inference questions were generated for every Target object. Twelve 4-year-olds (mean age = 4-4) were shown arrays of three objects. Each consisted of a Target object, the Distracter from the same trio, and a second Distracter object, which lowered the probability of randomly choosing the Target object to 33%. Second Distracter objects were chosen so that no one object from the array was very different from the others (i.e., all three were approximately the same size, were members of the same global category, etc.).

The children were shown the trios in quasi-random order and asked questions with the form, "Which one of these [*proposition*]?" Children were asked one candidate unknown fact and one candidate known fact for each trio. For example, children were shown a bar of soap (Target), a porcelain shell (Distracter), and a toy disk and asked, "Which one is made of lye?" (unknown fact for the inductive inference task). The objects then were shuffled and children later were asked, "Which one can you use to wash your hands?" (known fact question).

The criteria for using a question were as follows: (a) no more than five out of 12 children could choose the Target for an unknown question, binomial $p > .185$ (four out of 12 correct is expected by chance), and (b) at least 10 out of 12 children had to choose the Target for a known question, binomial $p < .0005$. Approximately two-thirds of the 35 necessary questions (one per training trio and two per test trio) were identified in the first round of pretesting. Additional candidate questions were generated, and 12 more 4-year-olds (mean age = 4-5) were tested with the same procedure. Using the same acceptability criteria, the remaining questions were identified. The questions are listed in Table 1.

Training and testing procedure.—Children were tested in a room outfitted with child-appropriate decorations and furniture.





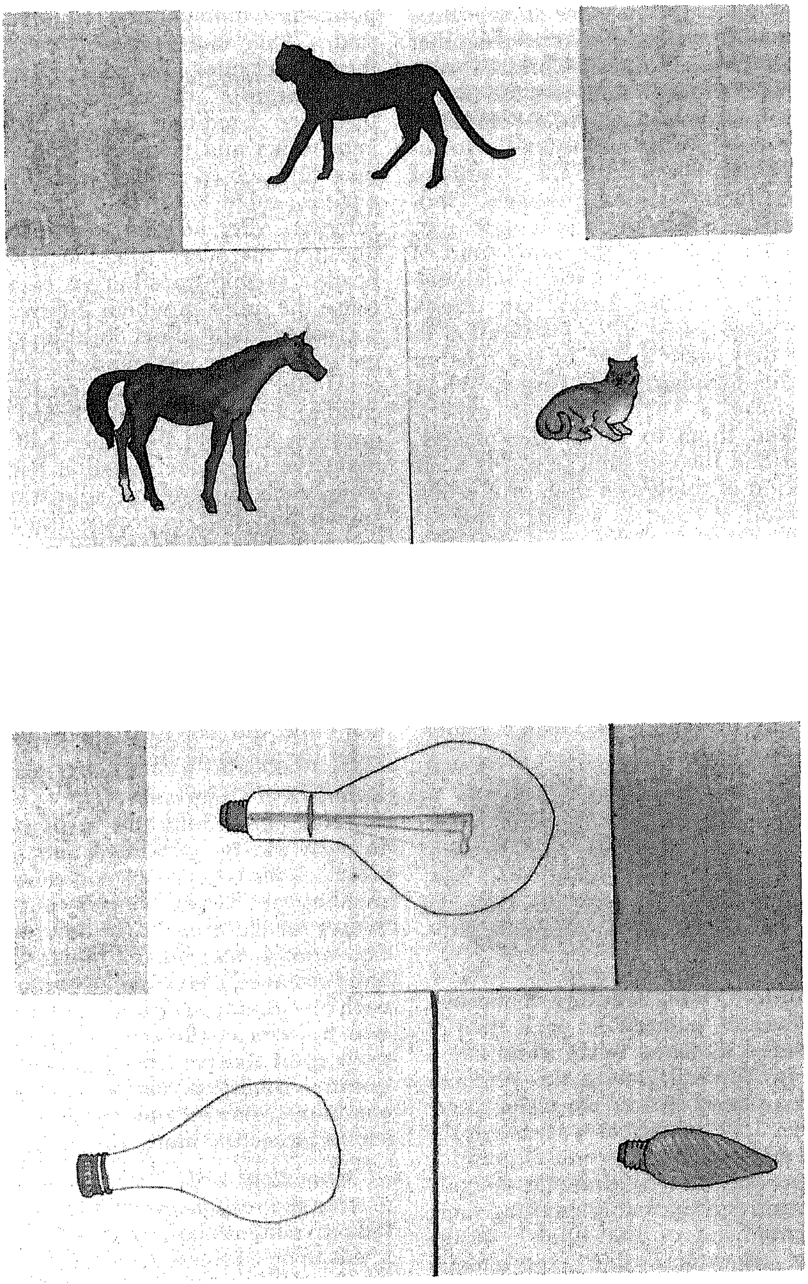


FIG. 1.—Examples of object and line drawing conflict trios used in Experiments 1–3. Each trio consists of a Standard (top), Target (lower right), and Distracter (lower left).

Prior to testing, children were shown three training trios, consisting of tokens of familiar kinds, in an invariant order. Children were first asked to name each item in a trio (in the labeling groups the labels were reiterated by the experimenter). This was intended to encourage children to attend to the identity of each item before making a judgment. Children then made a taxonomic judgment (e.g., "Which one of these is the same kind of thing as this one?") or a known inference (e.g., "Which of these barks and chases cars") and were asked for a justification to check that they were aware of the relation between the Standard and Target. When children chose a Distracter, the experimenter asked them to recall the objects' names, clarified the question (e.g., "Is a cat the same kind of thing as a dog, or is a dog the same kind of thing as a dog?") and repeated the question. This was meant to point out the disparity between the child's response and the relations between the items. Question clarification occurred up to three times, after which the experimenter proceeded to the next trio. This training procedure effectively clarifies the experimental task for preschoolers (Deák & Bauer, 1995).

Immediately following training, the 16 test trios were presented in one of 24 quasi-random orders (no more than three trios of one type, conflict or no-conflict, were presented sequentially). Left-right placement of the Target and Distracter was randomly determined.

For children in the labeling conditions, the experimenter pointed to each item in turn and stated its name twice (item labels are listed in Table 1) using the sentence frame, "This one is a(n) *x*." In order to ensure that any labeling effect was due to the label itself and not to the action of pointing to each item and thereby drawing the subjects' attention (and possibly inviting more careful comparison of the items), in the no-labeling conditions the experimenter pointed to each item in turn, after the child had examined them, and asked, "Did you see this one?"

Children in the taxonomic question conditions were shown a trio and allowed to examine each object or picture. The children were subsequently asked, "Which one of these [Target or Distracter] is the same kind of thing as this one [Standard]?" Children in the inference conditions were shown a trio and allowed to examine each item. The experimenter then pointed to the Standard and said, "This one [*proposition*]. Which of these other ones [Target or Distracter] also [*proposition*]?" Recall that children in the inference condition answered about one known fact and made one inductive inference for each trio. Thus, question type was a within-subjects variable for inference-task subjects. After hearing and answering one question for every trio, the children then heard a second question for every trio. Because the order in which the two questions was asked might affect children's responses, question order was randomized with the constraint that for any given trio, half the children in each condition heard the known question first and the other half heard the inductive inference question first. Furthermore, each child heard approximately half known and half inference questions first, and subsequently heard the remaining half of the known and inference questions. Consequently, question type order was counterbalanced across both subjects and items. Finally, question type order (known first or inference first) was randomized across items separately for each subject, so that more than one child did not hear the same question order for the same items.

Results and Discussion

Coding.—Children's responses to trios in each trio type (conflict and no-conflict) were converted to percentages based on the number of Target choices out of eight. Scores for children in the inference condition were further broken down so that each child received a percentage (out of eight) for each question type (known and inference) in both trio sets (conflict and no-conflict). Thus, each child received two (in the taxonomic question conditions) or four (in the inference conditions) percentage scores, with higher scores indicating more Target choices.

Question order effects.—Before discussing the responses of children who made inductive inferences, a methodological issue needs to be addressed: children in the inference task groups answered two questions about each trio (an inference question and a known-fact question). It could be that choosing an item in response to one question affected children's responses to subsequent questions. If children tended to choose the Target item in response to the inductive inference question *after* they had answered the known question, our results would be difficult to interpret. Because half of the subjects within each group heard any given known question first, and every child heard half known- and half inference-questions first, we can test this possibility.

For each child in the inference task conditions we calculated two new percentages: one for responses to inductive inference questions heard *before* the known question and one for responses to those heard *after* the known question. These scores were entered into a 2 × 2 × 2 mixed-measure ANOVA. The within-subjects factor was question order (inductive inference first vs. inductive inference second); the between-subjects factors were stimulus type (objects vs. drawings) and labeling (labels vs. no labels). This analysis allows us to determine whether question order had an effect on overall scores, and whether question order differentially affected different groups' scores. The ANOVA revealed no effect of question order, $F(1, 44) = 0.04$, N.S. Further, all interactions involving order were nonsignificant, $Fs(1, 44) = 0.31$, 0.36, and 0.16 for the two- and three-way interactions, respectively. Thus, there is no evidence that children's inductive inferences were affected by answering a previous question about the same trio. To further assure that our data were not biased by the repeated-measures design, a within-subjects *t* test was carried out to determine whether children's second round of responses included more Target choices than their first round of responses. The proportion of Target choices for all 16 trios was calculated separately for the first and second round of responses from each subject. This split-half analysis revealed no difference between subjects' first and second round of responses, $t(47) = 1.47$, $p > .15$ (two-tailed). In summary, there are no data or trends suggesting artifacts of repeated-measures design within the inference task groups.

No-conflict trios.—Mean scores for the no-conflict trios across all eight groups ranged from 96% to 100%, with a mean of 98% (SD = .6%). This near-unanimous choice of Target items from no-conflict trios in every condition indicates that children neither chose randomly, nor used an indiscriminate response strategy such as choosing the item that looked least like the Standard. Since there essentially was no variance in these scores, they were excluded from all subsequent analyses. This makes all analyses more conservative by increasing the overall variance (not surprisingly, all effects reported below also obtain if data are analyzed with trio type as a within-subjects variable).

Conflict trios: Analyses with inductive inferences.—Descriptive statistics for responses to conflict trios are shown in Table 2. Each child's percentage of conflict trio Target choices was entered into a 2 × 2 × 2 ANOVA with three between-subjects factors: stimulus type (object or line drawing), labeling (labels or no labels), and experimental task (taxonomic question or inductive inference). The first analysis included answers only to taxonomic questions and inductive inferences (see the top and middle sections of Table 2). Patterns of responses to known-fact questions are treated in a sepa-

TABLE 2

Mean Numbers, Standard Deviations, and Percentages of Target-Item Choices on Conflict Trios, by Stimulus Type (Objects or Drawings), Labeling (Labels or No Labels), and Task (Taxonomic, Inductive Inference, or Known Question)

	Stimulus Type					
	Drawing			Object		
Experimental Task	Mean	SD	%	Mean	SD	%
Categorization question:						
No labels	1.58	(.99)	19.8	3.66	(1.97)	45.8
Labels	6.0	(2.63)	75.0	7.34	(.78)	91.7
Inductive inference:						
No labels	2.17	(.94)	27.1	5.92	(1.09)	74.0
Labels	6.50	(1.50)	81.2	6.50	(1.45)	81.2
Known question:						
No labels	3.46	(1.80)	43.3	7.25	(.86)	90.6
Labels	7.08	(1.50)	88.5	7.58	(.66)	94.8

Note.—$N = 12$ per cell; known vs. inference questions is a within-subjects variable. The maximum possible score in each cell is 8.

rate analysis (see below). Consistent with the overall patterns predicted, there was a main effect of stimulus type, $F(1, 88) = 32.78$, $p < .001$, indicating that more conflict trio Target items were chosen in the object condition than in the line drawing condition. Children are more likely to categorize according to taxonomic relations when more physical information denoting those relations is available. A main effect of labeling, $F(1, 88) = 107.86$, $p < .001$, indicates that labels led children to choose more Target items. As in other studies (e.g., Gelman & Markman, 1986, 1987), labels assisted children in focusing on taxonomically related items. Finally, a main effect of task, $F(1, 88) = 3.99$, $p < .05$, reflects an increased tendency to choose Target items in the inductive inference task: a more specific question effectively directs children to attend to subtle features denoting taxonomic relations.

Several interactions also reached significance. A stimulus type and labeling interaction, $F(1, 88) = 12.92$, $p < .001$, indicates that labeling had a greater effect on performance in the line drawing groups than in the object groups: labeling increased the mean proportion of Target choices of the former by 0.54, versus 0.27 for the latter (both differences are significant, Tukey HSD $p < .01$). This might indicate that labeling is more effective when there is little other information on which to base categorization decisions, although it also could reflect a ceiling effect in the object groups. A second two-way interaction between labeling and experimental task reflects a task difference only in the no-labeling groups, $F(1, 88) = 6.40$, $p < .01$ (task difference without labels $p < .01$ by Tukey HSD; N.S. with labels). This suggests that the constraint imposed by the inference task was not necessary when labels were provided (alternately, because the mean proportion of conflict trio Target choices in the taxonomic question/label groups was 0.83, a ceiling effect might have obscured the task effect). When labels are available, children may match them rather than actively compare items to find features that suggest taxonomic relations. This is reasonable, because matching Standards to Targets when items are labeled need not require attention to subtle information underlying taxonomic relations. What *does* indicate that preschoolers attend to these subtle information relations is the fact that children in the object/no labels/inductive inference condition chose a mean of 74% Target objects. This is significantly higher than expected by chance (50%), $t(11) = 6.14$, $p < .001$ (two-tailed). These children focused on information denoting taxonomic relations three-fourths of the time, without the benefit of labels, despite the presence of a conflicting appearance-based answer.

Finally, a three-way interaction involving stimulus type, labeling, and experimental task, $F(1, 88) = 5.74$, $p < .02$, reveals that labeling significantly increased Target choices in all stimulus/task groups (Tukey HSD $p < .01$), with the exception of the object/inference task group (Tukey HSD $p > .05$). A possible reason for this effect has already been discussed: labels point out taxonomic relations, but they are not necessary if informative stimuli and task-specifying information are available (alternately, a ceiling effect may obscure the effect of labeling in this condition). This finding indicates that availability of rich information and a constrained categorization task (characteristic of everyday problems) are sufficient to allow preschoolers to categorize on the basis of similarities consistent with taxonomic relations.

Another important contributor to the three-way interaction is a stimulus type × task effect in the no-label conditions. Recall that the task effect was predicted for objects but not for line drawings, because line drawings do not contain the subtle information specified by the inductive inference questions. This hypothesis was supported: in the no-labeling conditions, there was a task effect only for objects (Tukey HSD $p < .01$). There was no task effect for drawings, either labeled or unlabeled (N.S. by Tukey HSD). Apparently, the task effect is largely driven by the objects, as predicted.

Conflict trios: Analyses with known-fact questions.—Because there is no evidence of a within-subjects confound involving order in the inference conditions, children's responses to the known questions allow us to assess the effect of specific knowledge about the Target object on inference-like questions. A second 2 × 2 × 2 ANOVA (stimulus type × labeling × task) was conducted using inference task subjects' responses to the known-fact questions. The children's scores on these questions (shown in the bottom section of Table 2) were calculated and entered as in the previous analysis. There were main effects of stimulus type, $F(1, 88) = 37.29$, $p < .001$, labeling, $F(1, 88) = 91.07$, $p < .001$, and task, $F(1, 88) = 29.08$, $p < .001$, all of which were in the same direction as in the previous analysis. A stimulus type and labeling interaction, $F(1, 88) = 10.23$, $p < .002$, reflects the pattern found in the previous analysis (labeling is more effective with line drawings than with objects). An interaction of labeling and task, $F(1, 88) = 10.74$, $p <. 001$, indicates an effect of task (known vs. taxonomic question) in the no-labeling condition but not in the labeling condition. Ceiling effects also might contribute to these interactions. A three-way (stimulus type × labeling × experimental task) interaction, $F(1, 88) = 4.04$, $p < .05$, revealed a notable pattern: children chose Target items at or near ceiling levels when they were given either labels *or* a known question. Follow-up analyses revealed that performance did not differ among any group that received either of these two manipulations (Tukey HSD $p > .05$ for all possible pairs of groups who answered either known questions or labels). There was one exception to this pattern: children in the line drawing/no label/known question condition chose only 43% Target items, despite answering known questions. This result may not be as surprising as it seems: children pretested for their knowledge of potential inference questions were shown objects, not drawings. To the extent that drawings lack information about the depicted items, factual knowledge about the depicted items may be more difficult to elicit or verify. For example, if subjects do not have enough information to verify that a bar of soap and a shell-shaped piece of soap are both made of soap, it may be difficult to conclude that the bar of soap can be used in the same way as the shellshaped piece. This possibility is explored in Experiment 3.

Experiment 2

Experiment 1 demonstrated that the type of stimuli used in categorization experiments affects preschoolers' performance, presumably because different types of stimuli provide different amounts of information. If this is correct, stimulus type also should affect adults' categorization choices. Although labeling and task also are relevant with respect to adults, we did not examine their effects because adults in Deák and Bauer (1995) chose conflict-trio target objects at near-ceiling levels, so boosting performance with labels and inference tasks would not be possible. In addition, finding appropriate inference and known questions would be very difficult given individual differences in adults' knowledge.

Since drawings lack information relevant to categorization choices, adults should make more Target choices when presented with objects than when presented with drawings. This is especially true if, as suggested above, drawings lack important diagnostic information about the objects they represent. We assessed this possibility by testing two groups of adults using the taxonomic-question task. One group was tested using objects, and the other group was tested using drawings of these objects.

Method

Subjects.—Twenty-four adults (12 female, 12 male) were recruited by means of advertisements posted on the campus of a large university. Subjects were students and staff members with a mean age of 25 years. Subjects were paid $3 for their participation.

Materials.—The same set of 19 object trios and corresponding drawing trios (three training, eight conflict, and eight no-conflict) used in Experiment 1 was used in this study.

Procedure.—During the training procedure, adults were shown the three training trios, one at a time. They were asked to look carefully at each of the items in a trio and then were asked, "Which one of these [Target or Distracter] is the same kind of thing as this one [Standard]?" Subjects were not given specific feedback. Following training, adults completed a testing procedure identical to that for the taxonomic question/no labels subjects in Experiment 1.

Results and Discussion

Two scores were generated for each subject: the percentage of Target choices from the conflict trios, and the percentage of Target choices from the no-conflict trios. Since the percentages from the no-conflict trios were at ceiling with little variance ($M = 98\%$, $SD = 10.8\%$), only the percentages from the conflict trios subsequently will be considered.

Adults presented with objects chose an average of 94% Target items ($SD = 15.5\%$), whereas adults shown line drawings chose an average of 49% Target items ($SD = 6.5\%$). This difference is statistically significant, $F(1, 22) = 85.1$, $p < .001$. Thus, adults as well as children are affected by the amount of available information in to-be-categorized stimuli. This suggests that the tendency to categorize drawings according to overall physical similarity (e.g., shape and color) is not a characteristic of preschool thought per se. Relative to objects, drawings lack information that is important for some categorization decisions. When such information is lacking and subjects cannot rely on relevant declarative knowledge, both preschoolers and adults may use a default strategy of categorizing on the basis of overall appearance (i.e., shape and color). This is a reasonable strategy, because overt features such as shape and color are highly diagnostic of category relations and are available in drawings.

The results from this experiment address the third factor (listed in the Introduction) that is necessary to fully describe the dynamics of categorization decisions: subject characteristics. There are many differences between preschoolers and adults: memory and metacognitive abilities, selective attention, breadth and depth of knowledge, and motivation to perform well in an experimental task, to name a few. Yet the results of this experiment indicate that stimulus informativeness affects subjects who differ considerably in these diverse characteristics. Consequently, subject characteristics do not mediate the stimulus informativeness effect (although, strictly speaking, we cannot be sure that preschoolers and adults make fewer taxonomic-based decisions about drawings for the same reason).

We should emphasize that the drawings used in these experiments were not somehow worse, or less informative, than those used in most studies. Our drawings were made by an artist instructed to draw the objects accurately, and, as indicated by the accuracy of adults' drawing-to-object matching (see Experiment 1), they were not systematically misleading. However, eliminating depth, texture, contour, and other subtle physical information has a profound effect on the categorization decisions of both preschool and adult subjects. We have suggested that the effect of stimulus type is attributable to the difficulty or impossibility of selectively attending to subtle physical features in line drawings, most importantly (for the current question) those features that denote taxonomic relations. This hypothesis is explored in more detail in Experiment 3.

Experiment 3

The first purpose of Experiment 3 was to replicate the effect obtained in the inductive inference condition of Experiment 1. Recall that children in that experiment were asked two questions about each trio. Although several analyses indicated no effects of repeated exposure to the materials, it seemed worthwhile to gain further confirmation, and so two of the conditions (objects/no labels and line drawings/no labels) were replicated using only the inductive inference task. The second purpose of Experiment 3 was to further investigate the effect of stimulus type on categorization decisions. The use of line drawings might reduce the rate of Target choices in categorization and inductive inference tasks because drawings do not depict subtle physical similarities that underlie taxonomic relations. If this is true, we would also expect subjects to be able to name or identify drawings only at more general levels of inclusion.

Because the Standard-Target matches in the stimulus trios are based on natural language taxonomic groupings (i.e., hierarchical class taxonomies that are encoded in linguistic terms), the lack of relevant physical information in drawings should affect performance in tasks other than sorting and inference tasks. It should also affect naming or identification, because the intensions of natural language terms (i.e., names) presumably rest on at least some of the same features that underlie taxonomic-based categorization judgments. However, because drawings portray gross physical features, subjects should be able to inexactly name or identify depicted items. For example, subjects should be able to differentiate drawings of animals from drawings of machines, although drawings might not allow subjects to differentiate, for example, a wolf from a coyote. Therefore, when asked to name or identify drawings, subjects should produce more superordinate-level names and more names that are correct at the global or superordinate level but incorrect at the basic level (e.g., identifying a drawing of a coyote as an animal but incorrectly labeling it "dog" or "wolf"). In contrast, subjects should label objects more precisely, accurately identifying these more-informative items at the basic- and subordinate-levels, as well as the superordinate level. If this pattern of findings, called the "identifiability effect," is obtained, it will support our interpretation of the stimulus-type effect in children's categorization and inference choices.

The identifiability effect was investigated by asking subjects to name stimulus items individually (i.e., out of the context of a trio) as well as make an inductive inference about the trio. If our hypothesis is correct, children should provide more general labels for drawings and more specific labels for objects. Furthermore, children should provide more general labels for items in trios about which children made appearance-based inferences, and more specific labels for items in trios about which children made taxonomic-based inferences.

In order to evaluate the identifiability effect, children's responses were coded in several ways (see "Method"). First, each label was coded according to its level of precision—from names that are correct at the subordinate- or basic-level (e.g., calling a diamond "diamond"), to those that are correct at the superordinate-level or within superordinate-level boundaries (e.g., calling a quartz crystal "rock" or "diamond"), to those that violate superordinate-level boundaries (e.g., calling a rhinoceros "dinosaur"). This coding scheme is intended to directly examine whether line drawings are more difficult to name (and, by extension, to identify) at more specific levels. Second, the relations between the labels produced for items in a trio and the inference choice about that trio were compared. If a subject makes an inference linking items X and Y (rather than X and Z), it is more likely that X and Y will receive the same label, and in fact this is often the case (see Deák & Bauer, 1995). However, when X and Y are *not* given the same (or nearly synonymous) label, we have no evidence that identifiability is playing a role in children's judgments. Such "mismatches" between naming and inferences are inconsistent with the identifiability effect and should be relatively infrequent.

This unique combination of measures—child-produced labels and inductive inference choice—also allows us to examine the effects of child-produced labels on the child's own inference choices. It is possible that covert labeling typically plays a mediating role in children's subsequent inference responses or vice versa. Asking children to complete both tasks might affect their responses. For example, it is unusual for children to systematically produce overt labels prior to making inferences, and their labels might affect subsequent inferences. In order to examine this possibility, half of the children labeled the items first, and half made inductive inferences first. If either task affects the other one, an effect of task order should be evident.

Method

Subjects.—Thirty-two 4-year-olds (16 female, 16 male; mean age 4-6; range 4-4 to 4-8) were recruited from the same population and in the same manner as in Experiment 1. Two additional subjects were excluded, one for failure to complete the task and one due to experimenter error. None of the children had participated in any of the prior pretesting or experimental groups. Children were given a small gift for their participation.

Design.—Half of the children (n = 16 per group) were presented with line drawings, the other half were presented with objects. Within each of these stimulus type groups half of the children (n = 8) made inductive inferences about the trios first and labeled individual items second; the order of tasks was reversed for the other half of the children.

Materials and procedure.—The 16 object trios and 16 line drawing trios used in Experiments 1 and 2 were used. The procedure for the inductive inference task (training and testing) was identical to that used in Experiment 1, with one exception: children were asked only one question (inductive inference) about each trio. Trio order and object placement were randomized as in Experiment 1.

In the naming task, children were told that they would see some things and were asked to name the things for the experimenter. The experimenter then presented every item (objects or line drawings) one at a time until the child named the item. If the child did not generate a label for an item, the experimenter used several prompts to encourage the child: "What kind of thing do you think it is?" and "It's okay if you're not sure; what do you *think* it is?" Item presentation order was determined as follows: each child in a stimulus type condition was given the 16 trios in a different random order, which was repeated three times, each time involving a different item from each trio. The first item type presented (Standard, Target, or Distracter) was counterbalanced across children, and item-type order then was rotated, so that if the first item presented was a Target, the next item (from the next trio in the order) was a Distracter, and subsequently a Standard, etc. In this way no two items from the same trio or object type were ever presented sequentially. In addition, items were presented one at a time, out of the context of the other items in the trio. This was desirable because we wished to assess the identifiability of individual items, and presenting the entire trio simultaneously might have led children to label difficult-to-identify items by extending the label given to a more easily identified, similar item from the same trio.

Scoring.—Inductive inference responses were scored as in Experiments 1 and 2 (i.e., Target or Distracter). Labels were assigned to one of several categories according to the degree of precision or accuracy with which the label identifies the item. *Within-Basic* labels accurately identified the item at a subordinate or basic level, for example calling the orange "orange" or the panther "panther." *Within-Superordinate* labels accurately identified the item at the superordinate level, either by providing the superordinate category name (e.g., calling the pliers "tool") or by providing the name of another basic-level kind from within a more inclusive superordinate category (e.g., calling the pliers "screwdriver" or the coyote "dog"). Because there are usually multiple levels within a taxonomic hierarchy, and because the levels considered "basic," "superordinate," etc. are not clearly defined and depend upon subjects' knowledge and experience (e.g., Palmer, Jones, Hennessy, Unze, & Pick, 1989), we defined "superordinate" fairly conservatively.[4] The superordinate level was defined as moderately more inclusive than the adult basic level (e.g., all mammals, all crystalline minerals). Global categories such as "all animals" (including birds, fish, etc.) or "all rocks or minerals" were considered more general than the superordinate and therefore were the basis for *Between-Superordinate* labels. *Between-Superordinate* labels either identified the item at a global level (e.g., calling the French horn "music thing") or identified another basic-level kind from outside of the item's superordinate class (e.g., calling the rhinoceros "dinosaur" or the sugar cubes "ice cubes"). Included in this category were instances in which the child appeared to incorrectly identify an item, for example calling the candle "bomb." Finally, *Other* labels included attribute names (e.g., calling the orange "circle") and ambiguous names (e.g., calling sugar cubes "coffee stuff"). All labels were independently coded by two coders. Intercoder reliability was 91%; disagreements were resolved by discussion.

There were three kinds of label/inductive inference responses, each of which reflected different relations between the labels and inductive inference choice for a trio (see also Deák & Bauer, 1995). *Matches* occurred when children produced the same name for the two objects across which an inference was made. For example, if a child made an inference from the Standard (light bulb) to the Distracter from trio 1 (bottle) and called them both "bottle," it would be considered a *Match* response (however, if the child also labeled the Target item, the ornamental light bulb, "bottle," it would be considered a *Nonmatch*). *Nonmatches* occurred when items united by inference were not distinguished from the nonchoice item by their labels. This happened when all three items were given different labels, all three items were given the same label, or the inference-connected items were given different labels and the third item was not labeled. For example, if a child called the Standard (light bulb) "light," the Distracter (bottle) "bottle," and the Target (ornamental bulb) "candle," that would be coded a *Nonmatch*. Finally, *Mismatches* occurred when subjects gave the same label to the Standard and the item to which an inference was *not* drawn. For example, if a child drew an inference from the Standard (light bulb) to the Distracter (bottle), but called both the Standard and *Target* "light bulb," and called the Distracter a different name (e.g., "bottle"), it would be a *Mismatch*.

Results and Discussion

Each child was assigned two percentage scores based on the number of Target choices for the inductive inference questions about the conflict trios and no-conflict trios, respectively. Children who saw objects chose an average of 93.7% Target items from the no-conflict trios (SD = 11.2%), and children who saw drawings chose an average of 96.9% Target items from the no-conflict trios (SD = 7.2%). Because the overall rate (95.3%) is at ceiling, the no-conflict trios are excluded from further analyses, making all subsequent analyses slightly more conservative. Furthermore, these data reinforce our previous conclusion that children are neither answering randomly nor using an indiscriminate response strategy such as choosing the different-looking item.

The average percentages of conflict-trio Target choices of children in the four condi-

[4] Note that an item can fall into several distinct taxonomic hierarchies and, therefore, can have several basic-level and superordinate-level names. This is a correlate of the point (n. 1 above) that any entity falls into a number of taxonomic categories. Again, however, we qualify this generality by limiting our consideration to specific hierarchical relations.

TABLE 3

MEAN NUMBERS, STANDARD DEVIATIONS, AND PERCENTAGES OF TARGET-ITEM CHOICES ON CONFLICT TRIOS, BY STIMULUS TYPE (Object or Drawing) AND ORDER (Inference Task First or Labeling First), EXPERIMENT 3

	STIMULUS TYPE					
	Drawing			Object		
ORDER	Mean	SD	%	Mean	SD	%
Inference task first	2.13	(1.55)	26.6	5.50	(1.07)	68.7
Labeling task first	2.75	(1.17)	34.4	5.41	(1.41)	67.6

NOTE.—N = 16 per cell; task is a within-subjects variable. The maximum possible score in each cell is 8.

tions (objects/inference task first, objects/labeling first, drawings/inference task first, drawings/labeling first) are shown in Table 3. These means were entered into a 2 × 2 ANOVA with two between-subjects variables: stimulus type (objects or line drawings) and order (inference task first or labeling task first). Children shown objects chose an average of 68% Target items (SD = 15.1%), compared to 30.5% among children shown drawings (SD = 17.0%). This difference is significant, $F(1, 28) = 41.78$, $p < .0001$, and closely resembles the results for the analogous groups from Experiment 1 (74% and 27%, respectively). The groups that made inferences prior to labeling most closely replicate the inductive inference/no label/objects and inference/no label/line drawings groups from Experiment 1. These groups chose an average of 68.7% and 26.6% Target items overall, respectively, figures which are very similar to the means from Experiment 1. Furthermore, the average of 68% in the object group is significantly greater than chance, $t(7) = 4.86$, $p < .01$ (two-tailed). When we pool the subjects from the inductive inference/objects groups in Experiments 1 (no label group) and 3 (inference task first group) ($N = 20$), there is an overall mean of 5.75 conflict trio Target choices out of 8 (71.9%), with a standard deviation of 1.07 (13.4%) and a range of 4 to 8 Target choices. Nineteen out of these 20 children chose five or more Target items, 11 chose six or more Target objects, and five children chose seven or eight Target items. The probability (by binomial theorem) of one child choosing seven or more Target items out of eight is .035, and the binomial probability of five or more out of 20 children choosing seven or eight out of eight Target items is less than .001. Thus, at least a substantial minority of the children chose Target items consistently, and no child showed the opposite pattern.

The effect of task order was not significant, $F(1, 28) = .29$, nor was the interaction between stimulus type and task order, $F(1, 28) = .65$. This indicates that overt labeling prior to categorizing does not increase the rate of taxonomic-based choices. This might seem surprising, given the robust effect of adult-provided labels on preschoolers' taxonomic-based categorization (e.g., Experiment 1; Gelman & Coley; 1990; Gelman & Markman, 1987; Markman & Hutchinson, 1984; Waxman & Hall, 1993; Waxman & Kosowski, 1990). There are at least three possible reasons for the lack of an effect (aside from the possibility that we failed to detect an effect, which is unlikely because the respective means, 47.5% and 50.7%, are so similar). One is that children spontaneously (i.e., without the experimenter's intervention) covertly label or identify all items when making a categorization judgment. As a result, *overt* labeling in response to an experimenter's request does not further affect the child's reasoning—it simply entails changing a covert response into an overt one. In support of this hypothesis, Deák and Bauer (1995) found that children spontaneously labeled 78.1% of the test items, suggesting that labeling frequently occurs without the experimenter's prompting. Another possibility is that children's self-produced labels did not encode taxonomic relations as effectively as did experimenter-provided labels. This explanation is quite plausible: experimenter-provided labels are carefully selected in order to induce taxonomic-based choices, typically by providing the same label to the Standard and Target and a different label to the Distracter. Child-generated labels are, of course, not so systematic. Note that one could experimentally provide labels that would not increase taxonomic-based responding, for instance by providing a name shared by all three items (e.g., "animal"). A third possible explanation is that children did not refer to their self-produced labels when subsequently making an inference. The current data do not support or rule out any possible explanation or combination of explanations for the lack of an effect.

Children produced labels for all but 14 out of a total of 768 objects (1.8%) and all but 26 out of 768 line drawings (3.4%). No difference was found in the overall number of labels for the two stimulus types, $F(1, 30) = 1.04$. The numbers of labels in each of three categories of precision (Within-Basic, Within-Superordinate, Between-Superordinate) were tallied for each subject in the two stimulus-type groups (see Table 4). These totals were entered into a 2 × 3 mixed-measure ANOVA, with stimulus type (objects or drawings) as the between-subjects factor and label precision category as the within-subjects factor. As expected, there was a significant interaction between label type and stimulus type group, $F(2, 90) = 18.77$, $p < .001$. Tests for simple effects revealed a greater mean number of Within-Basic labels in the objects than the line drawings group, $F(1, 96) = 23.57$, $p < .01$, and more Between-Superordinate labels in the drawings group than the objects group, $F(1, 96) = 36.64$, $p < .01$. No difference between stimulus type groups was observed in the mean number of Within-Superordinate labels, $F(1, 96) = 2.1$. Thus, whereas children in the two groups produced an equivalent number of labels, these were distributed differently among levels of precision. Children who saw objects produced more within-basic and fewer between-superordinate labels than did children who saw drawings. These data support the hypothesis that the lack of information in line drawings obscures taxonomic relations among items, affecting categorization, induction, and labeling.

Converging evidence for this conclusion is obtained by examining the patterns of label precision and inductive inference choices within a trio: the conditional probability that a subject identified one or more of the items within a trio at the Between-Superordinate level, given that the subject chose the Distracter, was .72. Thus it was likely that subjects inaccurately identified at least one item in a trio when they made an appearance-based inference about that trio. Clearly, both identification and induction are related in systematic ways to stimulus type. It is highly plausible that this relation is mediated by the amount of physical information about taxonomic relations that is available in the stimuli. An alternative to this possibility is that there is a causal relation between identification and induction; for example, failure to identify an item leads to appearance-based inferences. Because no order effect was observed, there is no evidence that the relation between the two tasks is causal (in either direction). Furthermore, there is some evidence that appears to contraindicate a causal effect.

If the relation between identification and inductive inference choices is causal, we might expect that inaccurate identifications cause Distracter choices. Although we cannot rule out this possibility completely, we can examine whether identifiability is *not* playing a role in some of children's inductive inferences, thus suggesting that the relation between identification and induction is noncausal (i.e., mediated by an underlying variable such as amount of available information).

TABLE 4

MEAN NUMBER OF LABELS PRODUCED AT EACH LEVEL OF PRECISION (Within-Basic, Within-Superordinate, Between-Superordinate, and Other) PER SUBJECT, BY STIMULUS TYPE CONDITION (Objects or Drawings), EXPERIMENT 3

	LABEL SPECIFICITY CATEGORY			
STIMULUS TYPE	Within-Basic	Within-Superordinate	Between-Superordinate	Other
Objects	28.7	11.4	4.7	2.2
Line drawings	23.9	10.0	10.7	1.7

NOTE.—$N = 12$ per stimulus type condition. Each subject was asked to name 48 items (three items per trio; eight conflict and eight no-conflict trios). Rows do not total 48 because subjects occasionally did not name an item.

There are four situations in which it is ikely that identification could be having ecisive effect on induction. These situa-1s involve both label precision and the ation between label patterns and categori-ion choices. Two situations involve object iflict trios: (1) trials in which subjects)se the Target and there was a Mismatch Nonmatch and (2) trials in which subjects)se the Target, and either the Standard or rget (or both) were identified at the Be-een-Superordinate level. In these situa-ns, subjects made a taxonomic-based in-ence but did not appropriately name the ms; we therefore infer that precise identi-ation played little role in the inference. e other two situations involve conflict os in the line drawing condition: (3) trials which subjects chose the Distracter and ere was a Mismatch (or Nonmatch in iich all items were given different labels) d (4) trials in which subjects chose the stracter and all items were identified at rels more specific than the Between-perordinate. In these cases, subjects led to make a taxonomic-based inference it accurately identified the drawings; we erefore infer that the subject chose the istracter despite accurate identification of ms in the trio. Subjects exhibited one of ese four response patterns in 37 out of the 6 trials. This indicates that in *at least* 21% the conflict trio trials, identification or ming of the items probably did not play a usal role in the inductive inference.

Additional data, on the relation between m labels and induction within a conflict o, support this point. On some trials iden-ication of items in a trio did not decisively fect inference choices. In the object group, ildren made a total of 87 conflict-trio Tar-t choices. On 17, or 19%, of those trials, at ast one object was labeled at the Between-iperordinate level. On at least a substantial inority of the trials, then, children chose irget objects but did not precisely identify 'ery object within a trio. In the drawing oup, children made a total of 39 Target ioices from conflict trios. On 17, or 44%, of ose trials at least one drawing was identi-ed at the Between-Superordinate level. gain, Target choices were not completely ntingent on precise identification of every em. A similar conclusion is based on con-ct trio Distracter choices. In the object oup, 19 out of 41 (46%) Distracter choices ere accompanied by Within-Basic or 'ithin-Superordinate labels for every item the trio. In the drawing group, seven out of 89 (8%) Distracter choices were accompanied by Within-Basic or Within-Superordinate labels for every item. This suggests that Distracter choices are not strictly contingent upon inaccurate identification of items. Further, it suggests that children do not make appearance-based inferences about drawings *strictly* because drawings are more difficult to identify than objects.

Although these data are not conclusive, they certainly are difficult to reconcile with the possibility that identification causally affects inference choices. Instead, they warrant an alternative interpretation, such as the possibility that both identification and inference choices are mediated by the amount of available physical information.

The results of the labeling data suggest two conclusions. First, among trios for which subjects chose the Distracter, subjects also found those objects difficult to identify precisely. Second, this pattern was not invariant—other factors or processes apparently are at work. The stimulus type effect is not strictly mediated by identification of objects and drawings. Further studies will be needed to more conclusively explore the relation between identifiability and the ability to categorize according to subtle attributes underlying taxonomic relations. Experiments 1–3 demonstrate that a stimulus' physical informativeness affects both of these variables. It is not known what other variables (e.g., breadth of declarative knowledge) also affect the relation between identifiability and taxonomic categorization. Similarly, it is not known what other variables impede taxonomic-based categorizing of line drawings.

General Discussion

The traditional view that preschoolers are "perceptually bound" (e.g., Inhelder & Piaget, 1964; Vygotsky, 1934/1986) has been supplanted by the view that young children can categorize according to conceptually important and nonobvious similarities (e.g., Gelman & Markman, 1986, 1987). It is now imperative to outline the factors affecting preschoolers' categorization choices. In doing so, rather than focusing on dichotomous "either/or" questions such as "Do preschoolers primarily use perceptual or conceptual information?" (a question that is problematic because perception and cognition cannot easily be dichotomized), it is useful to describe the interactions and dynamics among three kinds of factors—available information, experimental task, and subject characteristics—affecting categorization choices. The data presented here, as well as other recent data (described below) exemplify the usefulness of this descriptive taxonomy.

Previous work has explored the manner in which these factors affect children's tendency to categorize according to either appearances or taxonomic relations. With regard to the task and information factors, Deák and Bauer (1995) showed that two task variables, instructions and training, have a significant effect on children's decision to categorize according to either appearance or taxonomic relations. Gelman and her colleagues (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987) have shown that one kind of information, labels, has a substantial effect on preschoolers' categorization choices when appearances and taxonomic relations are in conflict (see also Markman & Hutchinson, 1984; Waxman & Hall, 1993; Waxman & Kosowski, 1990). This conclusion was supported in Experiment 1. Experiment 1 revealed additional effects of information and task factors. In the aforementioned descriptive taxonomy, labeling and stimulus type are information variables, and task is a task variable. Experiment 2 revealed effects of stimulus type on adults' categorization choices, demonstrating that certain variables (e.g., stimulus informativeness) have similar effects across diverse age groups. Also, as in Deák and Bauer (1995), adults chose more Target items than did children. This suggests that age, a subject variable, affects performance. However, we cannot determine whether the age difference is due to differences in factual knowledge, understanding of pragmatics, selective attention, or other variables.

The results of Experiment 1 make it clear that physical information and experimental task are critical determinants of categorization performance. Compared to unlabeled drawings, unlabeled objects induce substantially more categorization choices based on taxonomic relations or subtle features underlying taxonomic relations. Our drawings are not, however, systematically misleading, as shown by adults' accurate performance on a picture-to-object matching task. Experiment 3 showed that this effect is strongly, but not strictly, related to the difficulty of precisely identifying or naming items depicted in line drawings. It may seem paradoxical that adults accurately matched pictures to objects but might have had difficulty precisely identifying pictures. This is not difficult to explain, because the two tasks are quite different. A pertinent difference is that in the matching pretest, drawings and objects were simultaneously present. Subjects in the pretest did not need to label or identify items; they merely had to find the closest match between an object and a corresponding drawing. The pretest strictly demonstrates that the drawings were not *misleading* with regard to the particular object they were designed to represent. In other words, each drawing looked less like the objects it *did not* represent than the object it *did* represent. This is meaningful, because in every trio two objects were chosen to look similar, and it could have been the case that pictures of these objects were not discriminably different. In fact, data from the matching pretest argue that pictures of the similar-looking items were accurately discriminated.

There are other ways in which the physically impoverished nature of line drawings might discourage taxonomic-based categorization. For example, line drawings lack subtle information that is diagnostic of taxonomic relations. Notably, Melendez, Bales, and Pick (1993) found that preschoolers consistently sorted objects by function after examining the objects but did not do so with pictures of the objects (see also Daehler, Lonardo, & Bukatko, 1979).

The stimulus-type effect, which was robust in all three experiments, has serious implications for the interpretation of past research and the construction of future research. First, it suggests that several past studies (e.g., Fenson et. al., 1988; Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Melkman et al., 1981; Olver & Hornsby, 1966; Tversky, 1985), which reported a tendency for children to categorize by appearances in the absence of labels, might have biased children's responses by using impoverished stimuli. Second, it underscores the importance of using rich, informative stimuli, rather than simple line drawings, when devising tests from which the results are intended to generalize to everyday behavior. More generally, these data raise intriguing questions about the possible effects of using impoverished stimulus materials in tasks for young children that involve complex decision making and reasoning. Of course, objects and line drawings probably do not invariably lead to qualitatively different performance. The critical issue likely is not one of objects versus line drawings (or two vs. three dimensions) per se. Rather, the issue is whether a particular stimulus contains information pertinent to a particular categorization decision. Certainly some three-dimensional representations are distorted and/or impoverished compared to the objects they represent, just as some two-dimensional representations (e.g., large, color photographs) depict subtle features such as texture, reflectance, etc. However, objects usually contain more information than drawings, and it is this fact that underlies our findings.

Compared to a more open-ended categorization question, a question that implies which information should be attended can induce taxonomic-based categorization, if the information is diagnostic of taxonomic relations. For instance, if a child is told that the Standard object is made from an unknown substance, she must look for another object with physical properties (e.g., color, texture, reflectance, density, etc.) that specify the same substance. Such properties are difficult or impossible to portray in drawings, so the task effect is limited to stimuli that contain the relevant information. Apparently, then, a constrained task and informative stimuli are sufficient to induce taxonomic categorization in preschoolers. Of course, a constrained task may also induce appearance-based categorization if the questions imply attributes that determine overall appearance (see Gelman & Markman, 1986, Study 3).

Another kind of information, labels, also had a profound effect. Labels might draw children's attention to similarities among items, or children might simply match items that are given identical labels. The possibility that easily matched labels reduce or eliminate further processing of stimuli has not been demonstrated. In fact, it seems unlikely that labels typically have this effect. Rather, in some particularly ambiguous situations, such as reasoning about conflict trios, children might look for additional information to verify or "double check" the connection drawn by adult-provided labels. Tasks requiring decisions about known characteristics also facilitate choosing taxonomically related items. Notably, in this circumstance it cannot be determined whether the child is actively comparing and drawing a connection between two items (i.e., categorizing) or simply choosing an item with a particular characteristic (e.g., something that lights up). The latter possibility seems unlikely, however, because it is unclear why children would ignore overall appearances to categorize on the basis of subtle, nonsalient information, unless that information gained importance by virtue of its association with a taxonomic relation.

This constellation of effects is fleshed out in more detail by the interactions observed in Experiment 1. Labels had a greater effect with line drawings than with objects: perhaps the scarcity of physical information in line drawings compelled children to attend more exclusively to the labels. The ease of label matching is suggested by the fact that the task difference (inductive inference > taxonomic question) occurred only in the no-label conditions. This is not surprising: Why bother figuring out what physical information is important from the content of a question, when a simpler basis for categorization (i.e., matching labels) is available? Perhaps the most notable interaction, however, was the finding of a task effect for unlabeled objects, but not drawings. This supports our contention that preschoolers can categorize according to taxonomic relations if (1) children are given a task that implies features underlying taxonomic relations and (2) those features are available in the stimuli. Whereas it might be argued that establishing these conditions essentially provides children with a particular answer, this argument is spurious. For one thing, the procedures used in these experiments by no means unambiguously highlighted a particular feature. Furthermore, the objects were sufficiently rich and varied that the relevant feature(s) differed from trial to trial. It is more plausible to suggest the following: categorization tasks (e.g., inference and naming tasks) in which criteria are dissociated (e.g., appearances and taxonomic relations), are inherently ambiguous. Children as young as 4 years are capable of choosing according to either criterion, and a number of factors affect their choices. In the present series of experiments, the richness of the stimuli and the specificity of the task profoundly affected this decision. Other research shows that procedural factors that affect subjects' interpretation or construal of the experimental task have a similar effect (Deák & Bauer, 1995).

This work, along with Deák and Bauer (1995), also demonstrates that children do not need labels in order to make categorization choices that are consistent with conventional taxonomies. First, the effects of labeling were greater in the taxonomic-question task than the inference task. It is possible that labeling is more useful in a task that provides little guidance about what information is important. Second, when children saw unlabeled objects and made inductive inferences, they chose an average of 74% Target (taxonomically related but dissimilar-looking) objects from the conflict trios. This is significantly higher than expected by chance, and it is replicated in Experiment 3, in which children who made inductive inferences about objects prior to labeling those objects chose an average of 69% Target items. Thus, when preschoolers were asked a question that required the selection of subtle or nonobvious attributes that underlie taxonomic relations, and when that information was available, they drew an inference consistent with those relations. It is clear, then, that preschoolers are not limited to using one type of decision-constraining information (e.g., labels); rather, the selection process is quite sensitive and complex.

These data demonstrate that preschoolers can use different types of information to make categorization decisions. However, they do not answer several questions about preschoolers' categorization. First, although labels are salient cues to categorization, there may be situations in which labels do not help and may even hinder the categorization process. For example, calling both a pampered Persian and a wild tiger, "cat," obscures meaningful differences between the two. Determining how the role of labels changes across situations may clarify the reason for their usefulness.[5] Second, the differences between drawings and objects are many, and further investigation is necessary to specify how the information available in objects, but not drawings, is used for categorization decisions. For example, how precisely can young children select particular kinds of physical information on the basis of a specific task or question? Another important aspect of this question is the role of knowledge. For example, an individual with extensive knowledge of biological taxonomies will categorize a picture of a dolphin with that of a whale instead of a shark (if the task calls for a taxonomic match) as long as the individual can identify all three items. In this situation it might not matter whether physical details such as blowholes and horizontal tail fins are depicted. A third question, related to the former point, concerns the fact that no attempt was made in this study to control for children's familiarity with the experimental stimuli. Although this does not qualify our conclusions (because pretesting established that children did not know the answers to the inductive inference questions), and although the issue of familiarity itself is quite problematic, experiments using novel objects will show whether experience with specific kinds of items is necessary to make inductive inferences that ignore overall appearances. Exploration of these and related problems will lead to a better understanding of the nature of inductive inferences. Finally, our interpretation of the task effect calls for more detailed study of inference making. We suggest that preschoolers can use the semantic context of an inference question to constrain their selection of relevant information. This implies that when preschoolers hear a novel fact with, for example, the frame, "This one is *made of X*?" they will attend to substance-specifying cues such as texture and reflectance, as opposed to, for instance, shape or parts. There is some evidence that children do in fact use these cues: Deák and Bauer (1995) asked children who were trained and instructed to categorize objects according to taxonomic relations to justify their conflict-trio choices. Features not available in drawings were cited to justify 19.8% of their choices (e.g., "these kinda smell good," "they're both squishy," "they have bumpy things on them"). Future research should attempt to provide further evidence that such information is used in categorization decisions.

The results of the research reported here show that when making categorization decisions, children are capable of using many kinds of information, including information in the environment and expectations and knowledge about the world. This abundance of information, however, must be reduced in order to make categorization decisions under conditions of uncertainty (i.e., where there is more than one possible way

[5] Note that in Experiment 1 we presented identical labels close together in time with clear referents. Using synonyms (e.g., "puppy" and "dog," "bunny" and "rabbit") Gelman and Markman (1986, Study 2) also found an effect similar to that reported here. In order to understand how different labeling variables affect categorization, further research is required. For example, research imposing a delay between labeling and sorting, or comparing identical to nonidentical labels (e.g., taxonomically related labels such as "plant" and "flower"), will provide a clearer picture of the role of labeling in everyday categorization.

to categorize items). It is this process of selecting relevant information that is implied by our data (see also Barsalou, 1991; Jones & Smith, 1993; Medin, Goldstone, & Gentner, 1993; Murphy & Medin, 1985; Smith & Heise, 1992).

Several other recent findings also emphasize the process of selecting relevant information on the basis of the task at hand. For example, Kalish and Gelman (1992) found that 4-year-olds made inductive inferences about unfamiliar properties ostensibly connected to material (e.g., "becomes sodden in water") on the basis of common material, whereas they made inferences about properties ostensibly connected to object function (e.g., "is used for accelerating") on the basis of a common basic-level object name. Similarly, 3-year-olds sorted items according to what room they belong in on the basis of object name, whereas they sorted according to breakability on the basis of material. These findings demonstrate that preschoolers can select relevant information according to the task at hand, dictated by the semantic content of an inference question.

Other data (Jones et al., 1991) indicate that the shape of artificial objects is a more powerful criterion than texture or size for preschoolers' word generalization, but when eyes are added texture also becomes important (see also Landau, 1993). In addition, when a novel word was presented in an adjectival syntactic frame, and a particular attribute of the object was emphasized (e.g., a spotlight shone on its glittery surface), the word was extended to objects sharing that attribute. Similarly, Landau (1993) reports that when the novel word is presented as a preposition, the object's shape matters less than its location relative to other objects. Clearly, children take into account various kinds of information about objects, as well as the syntactic frame in which a word is presented, when determining the extension of that word (see also Dickinson, 1988; Soja, Carey, & Spelke, 1991).

Finally, Ward, Becker, Hass, and Vela (1991) showed that, when preschoolers extend a novel word, shape matters more if the drawings depict a creature than if the parts are rearranged to depict a nonanimate object. Apparently, some aspects of shape are considered more diagnostic than others of certain kind of things (e.g., animals). Even this is an oversimplification, however: Becker and Ward (1991) showed that children consider the outer contours of an animal irrelevant when those contours merely suggest a postural change. Landau (1993) reports a similar finding.

These studies, as well as the current research, suggest that young children's use of task characteristics to constrain information selection when making categorization decisions may be a relatively general phenomenon. It is becoming increasingly difficult to make broad generalizations such as, "Children prefer [information type *x*] to [information type *y*] when categorizing." Instead, it appears that task, context, knowledge, and the particular kinds and combinations of information available to children all jointly constrain their categorization decisions. It remains to be seen whether a sufficiently powerful and useful theory can be constructed to accurately describe and explain these findings.

References

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, **6**, 429–451.

Baillargeon, R. (1987). Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology*, **23**, 655–664.

Baillargeon, R. (1993). The object concept revisited: New directions in the investigation of infants' physical knowledge. In C. E. Granrud (Ed.), *Visual perception and cognition in infancy*. Hillsdale, NJ: Erlbaum.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. **27**). New York: Academic.

Becker, A. H., & Ward, T. B. (1991). Children's use of shape in extending novel labels to animate objects: Identity versus postural change. *Cognitive Development*, **6**, 3–16.

Biederman, I. (1988). Aspects and extensions of a theory of human image understanding: In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective*. Norwood, NJ: Ablex.

Bruner, J., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth*. New York: Wiley.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1993). *The origin of concepts*. Paper presented at the meeting of the Society for Research in Child Development, New Orleans.

Daehler, M. W., Lonardo, R., & Bukatko, D.

(1979). Matching and equivalence judgments in very young children. *Child Development*, 50, 170–179.

Deák, G. O. (1994). *The role of Descartes' framework in explanations for developmental phenomena.* Paper presented at the Annual Symposium of the Jean Piaget Society, Chicago.

Deák, G. O., & Bauer, P. J. (1995). The effects of task comprehension on preschoolers' and adults' categorization choices. *Journal of Experimental Child Psychology*, 60, 393–427.

Deák, G. O., & Pick, A. D. (1994). *What is a penlight? Preschoolers' systematic and flexible use of abstract categorization criteria.* Paper presented at the sixth annual convention of the American Psychological Society, Washington, DC.

Dickinson, D. K. (1988). Learning names for materials: Factors constraining and limiting hypotheses about word meanings. *Cognitive Development*, 3, 15–35.

Fenson, L., Cameron, M. S., & Kennedy, M. (1988). Role of perceptual and conceptual similarity in category matching at age two years. *Child Development*, 59, 897–907.

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

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a Dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, 26, 796–804.

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

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, 58, 1532–1541.

Gentner, D. (1978). A study of early word meaning using artificial objects: What looks like a jiggy but acts like a zimbo? *Papers and Reports on Child Language Development*, 15, 1–6.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child.* New York: Harper & Row.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, 8, 113–139.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, 62, 499–516.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development*, 63, 1536–1557.

Kemler-Nelson, D. G. (1991). *Principle-based inferences in preschoolers' categorization of novel artifacts.* Poster presented at the meeting of the Society for Research in Child Development, Seattle.

Kendler, H. H., & Kendler, T. S. (1975). From discrimination learning to cognitive development: A neobehavioristic odyssey. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1). Hillsdale, NJ: Erlbaum.

Landau, B. (1993). Learning the language of space. In E. V. Clark (Ed.), *Proceedings of the 24th Annual Child Language Research Forum.* Palo Alto, CA: Stanford Linguistics Association.

Lee, C. L., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception and Psychophysics*, 15, 529–532.

Leslie, A. M. (1988). The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language.* Oxford: Clarendon.

MacKay, D. G. (1973). Aspects of the theory of comprehension, memory, and attention. *Quarterly Journal of Experimental Psychology*, 25, 22–40.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, 16, 1–27.

Massey, C. M., & Gelman, R. (1988). Preschoolers' ability to decide whether a photographed unfamiliar object can move itself. *Developmental Psychology*, 24, 307–317.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.

Melendez, P. H., Bales, D. W., & Pick, A. D. (1993). *Direct and indirect perception: Four-year-olds' grouping of toys.* Poster presented at the meeting of the Society for Research in Child Development, New Orleans.

Melkman, R., Tversky, B., & Baratz, D. (1981). Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering, and retrieval. *Journal of Experimental Child Psychology*, 31, 470–486.

Miller, R. (1973). The use of concrete and abstract concepts by children and adults. *Cognition*, 2, 49–58.

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

Olver, R. R., & Hornsby, J. R. (1966). On equivalence. In J. Bruner, R. R. Olver, & P. M. Greenfield (Eds.), *Studies in cognitive growth* (pp. 68–85). New York: Wiley.

Palmer, C. F., Jones, R. K., Hennessy, B. L., Unze, M. G., & Pick, A. D. (1989). How is a trumpet known? The "basic level object" concept and perception of musical instruments. *American Journal of Psychology*, **102**, 17–37.

Smith, J. D., & Kemler-Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, **113**, 137–159.

Smith, J. D., & Shapiro, J. H. (1989). The occurrence of holistic categorization. *Journal of Memory and Language*, **28**, 386–399.

Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In B. Burns (Ed.), *Advances in psychology—percepts, concepts, and categories: The representation and processing of information*. Amsterdam: Elsevier.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, **38**, 179–211.

Spelke, E. S. (1991). Physical knowledge in infancy: Reflections on Piaget's theories. In S. Carey & S. A. Gelman (Eds.), *The epigenesis of mind*. Hillsdale, NJ: Erlbaum.

Starkey, P., & Cooper, R. S. (1980). Perception of numbers by human infants. *Science*, **210**, 1033–1035.

Tomikawa, S. A., & Dodd, D. H. (1980). Early word meanings: Perceptually or functionally based? *Child Development*, **51**, 1103–1109.

Tversky, B. (1985). Development of taxonomic organization of named and pictured categories. *Developmental Psychology*, **21**, 1111–1119.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT. (Original work published 1934)

Ward, T. B., Becker, A. H., Hass, S. D., & Vela, E. (1991). Attribute availability and the shape bias in children's category generalization. *Cognitive Development*, **6**, 143–167.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, **167**, 392–393.

Waxman, S. R., & Hall, D. G. (1993). The development of a linkage between count nouns and object categories: Evidence from fifteen- to twenty-one-month-old infants. *Child Development*, **64**, 1224–1241.

Waxman, S. R., & Kosowski, T. D. (1990). Nouns mark category relations: Toddlers' and preschoolers' word-learning biases. *Child Development*, **61**, 1461–1473.

Wellman, H. M., & Gelman, S. A. (1988). Children's understanding of the nonobvious. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale, NJ: Erlbaum.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, **43**, 337–375.

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development*. Minneapolis: University of Minnesota Press.

Yonas, A., Arterberry, M., & Granrud, C. E. (1987). Space perception in infancy. In R. A. Vasta (Ed.), *Annals of Child Development*. Greenwich, CT: JAI Press.