JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY **60**, 393–427 (1995)

# The Effects of Task Comprehension on Preschoolers' and Adults' Categorization Choices

GEDEON DEÁK AND PATRICIA J. BAUER

*University of Minnesota*

In experimental tasks in which subjects sort sets of objects with conflicting appearances and taxonomic relations, preschoolers often have been found to categorize according to appearance. The procedures used in past studies, however, may have biased preschoolers to attend to appearance instead of taxonomic relations. This possibility was examined in two experiments. In Experiment 1, children's preference for taxonomic- or appearance-based sorting was affected by both the training and the instructions they received. Adults in Experiment 1 were affected by instructions, but not by training. In Experiment 2 preschoolers sorted above chance according to the criterion for which they received training and instruction (taxonomic relations or appearance). Consistency data, children's justifications, and spontaneous labeling support the conclusions that training and instructions have a significant effect on children's preference to sort according to taxonomic relations or appearance, and that both criteria can be used systematically by children as young as four. Implications for task comprehension, flexibility, methodology, and education are discussed. 

In the past decade there has been a dramatic change in our understanding of young children's ability to categorize objects. Prior to this time, it was widely agreed that preschoolers are "perceptually bound," or limited to categorizing objects according to overt physical similarities or overall appearance (e.g., Bruner, Olver & Greenfield, 1966; Flavell, 1985; Inhelder & Piaget, 1964; Kendler & Kendler, 1975; Vygotsky, 1962; Wern-

This research was supported by a University of Minnesota graduate fellowship to the first author; NICHD Grant (HD-07151) to the Center for Research in Learning, Perception, and Cognition at the University of Minnesota; and NICHD FIRST award (HD28425) to the second author. Experiment 1 was submitted by the first author as a predoctoral thesis to the University of Minnesota. Part of this research was presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA, 1993. The authors thank Michael Maratsos, Anne Pick, Herbert Pick, Leah Welch, and several anonymous reviewers for comments on previous versions of this paper. Kathy Kohl assisted in testing subjects, and Donald MacEachern provided statistical advice. The authors also thank the children and parents who took part in this research. Address correspondence to Gedeon Deák, Department of Psychology and Human Development, Box 512, Peabody College, Vanderbilt University, Nashville, TN 37203.



0022-0965/95 $12.00

er, 1957). This view has been dispelled by evidence that preschoolers make inferences on the basis of shared membership in a taxonomically-defined category, rather than similarity of appearance, when these relations are placed in conflict (e.g., Carey, 1985; Gelman & Markman, 1986, 1987).[1] Other research has demonstrated that preschoolers can group items according to either taxonomic relations that are not dissociated from appearances (e.g., dog and cat) or thematic relations (e.g., dog and bone) depending on contingencies such as training (Smiley & Brown, 1979), social reinforcement and instructions (Bauer & Mandler, 1989), and labeling (Markman & Hutchinson, 1984; Waxman & Hall, 1993). Taken together, these and other findings strongly suggest that by an early age (3 years) children can group objects according to a variety of relations (e.g., appearance, taxonomic, and thematic) between objects.

This conclusion has led researchers to investigate factors affecting children's choice of a categorization criterion when more than one criterion is available. For example, preschoolers are more likely to categorize according to taxonomic relations or appearances, rather than thematic relations, when items are labeled (e.g., Markman & Hutchinson, 1984; Waxman & Hall, 1993; Waxman & Kosowski, 1990). Likewise, preschoolers extend inductive inferences to a dissimilar-looking item from the same category rather than a similar-looking item from a different category when the same-category items are labeled (e.g., Gelman & Markman, 1987; Gelman & Coley; 1990; however, see Davidson & Gelman, 1990). Some attention has also been paid to non-linguistic factors that affect preschoolers' tendency to categorize by either taxonomic relations or similarity of appearance. Such factors include naturalness (Gelman, 1988), inclusiveness (Gelman & O'Reilly, 1990), and the semantic content of a to-be-inferred fact (Deák, 1995; Gelman, 1988; Gelman & Markman, 1986; Kalish & Gelman, 1992; see also Massey & Gelman, 1988).

These findings suggest that preschoolers take into account several factors when making categorization decisions. Aside from labeling, however, most of these factors have been explored using the inductive inference task, and inductive inference (i.e., generalizing or predicting properties on the basis of shared similarities) has distinct differences from more traditional measures of categorization such as sorting. For example, inductive infer-

[1] Following commonly used operational definitions, natural kinds are grouped into classes according to the current wisdom of the relevant scientific fields (Putnam, 1975; but see Dupré, 1981; and Lakoff, 1987, for objections). Artifactual objects are taxonomized according to their intended and/or typical functions or functional properties. In general, both natural kind and artifactual taxa have correlated physical properties that determine overt appearance (e.g., shape and color). Nevertheless, there are situations in which taxonomic relatedness and appearance are dissociated, and it is these situations with which we are concerned.

ence tasks may provide information on category membership, whereas sorting requires assigning category membership to items. More importantly, typical experimental inductive inference tasks are fairly specific, allowing subjects to focus on relatively constrained similarities, whereas typical experimental sorting questions are fairly general, for example, "Put the ones together that go together." Thus, although the two tasks are related, findings from inference tasks might not generalize to other categorization tasks. In fact, findings from one inference task do not generalize to another: Gelman, Collman, and Maccoby (1986) found differences between preschoolers' category-to-property inferences and property-to-category inferences. Even within a particular inductive inference task, different novel to-be-induced facts support different inferences. For example, some of Gelman and Markman's (1986, Study 3) subjects drew inferences to similar-looking items when the novel fact implied a physical trait such as weight (see also Kalish & Gelman, 1992).

In light of the differences between inductive inference and sorting/classification tasks it is not surprising that there is at least one report of systematic differences between them: Deák and Bauer (in press) found that preschoolers made relatively more taxonomic-based categorization choices when answering inductive inference questions than when answering less-specific sorting questions (e.g., "Which is the same kind of thing as this one?"). This may have been because different kinds of questions convey different information: inductive inference questions often provide clues to what kind of similarities are most relevant, whereas sorting questions usually do not. Thus, in the inductive inference task, facts about, for example, origins, diet or internal anatomy compel children to attend to taxonomic relations, whereas facts about physical traits such as weight compel them to attend to appearance.

The findings from inductive inference studies suggest that children flexibly choose from among several similarities when drawing inferences (Gelman & Markman, 1986; Kalish & Gelman, 1992). In fact, the inductive inference task inherently supports flexibility, since the experimenter can change the task by changing the semantic content of the novel, to-be-inferred fact. Traditional sorting tasks typically do not support this degree of flexibility; nor is it guaranteed that flexibility extends to sorting tasks. On the contrary, existing evidence from sorting tasks suggests that when appearances are placed in conflict with taxonomic relations, preschoolers sort according to appearances unless labels highlight taxonomic relations. It is noteworthy that evidence of the predominance of appearance- over taxonomic-based responses mostly comes from sorting tasks in which items were unlabeled (e.g., Fenson, Cameron & Kennedy, 1988; Inhelder & Piaget, 1964; Melkman, Tversky, & Baratz, 1981; Tversky, 1985). The modal interpretation of this pattern is that preschoolers are predisposed to attend to appearance over taxonomic relations in the absence of information (such as labels) to the contrary. However, this pattern may be attributable to the conditions of testing rather than a trait of the organism: just as some inductive inference questions compel preschoolers to attend to physical traits, sorting tasks using unlabeled objects might incorporate task factors that compel preschoolers to attend to overt appearances. If so, then a variety of factors that alter the testing or task conditions might also affect preschoolers' preference for appearance- or taxonomic-based sorting. Because little research has explored procedural factors affecting preschoolers' sorting performance (but, see e.g., Markman, Cox, & Machida, 1981), it is as yet difficult to evaluate this possibility.

The goal of the present research is to test variables that might affect young children's decisions to sort unlabeled objects according to either taxonomic relations or overt appearances. There are three reasons for doing so, as implied above. First, there is little research about non-linguistic factors that affect children's sorting behavior (but see Daehler, Lonardo, & Bukatko, 1979; Deák & Bauer, in press; Deák & Pick, 1994; Markman et al., 1981; Waxman & Gelman, 1986; Waxman, Shipley, & Shepperson, 1991). Clearly, we do not yet possess a well-articulated understanding of the factors (particularly non-linguistic factors) that constrain children's categorization choices. Why is this important? As Ward, Becker, Hass, and Vela (1991) point out, if only linguistic factors are investigated, any preferences or predispositions could be specifically linguistic in nature. It is necessary, then, to determine whether such biases are language-specific or whether they are related to more general cognitive tendencies. This is particularly true regarding children's tendency to categorize according to appearances versus taxonomic relations, because (with the exception of factors investigated by Deák and Bauer, in press) only labeling has been found to affect this tendency. Thus, preschoolers' ability to categorize according to taxonomic relations when appearances conflict could be a specifically linguistic phenomenon. Alternatively, it could be, as Gelman and Markman (1986, 1987) suggest, that labels constitute conceptual information linking taxonomically related items, and children prefer this conceptual information to appearances (note that it is unclear why labels should be considered conceptual information). By eliminating labels, one can evaluate whether labels are necessary for taxonomic classification. Research on non-linguistic factors affecting categorization choices might resolve this issue and provide a more complete picture of the development of categorization.

Second, this research may address the question of whether flexibility in children's categorization is a function of particular tasks that support flexibility (i.e., inductive inference tasks), or whether flexibility is a general characteristic of preschoolers' categorization choices. That is, if non-linguistic factors compel children to adopt either a taxonomic- or appearance-based sorting pattern, it will indicate flexibility in a task that does not overtly, inherently support flexibility. The issue of flexibility is extremely important (Mandler, 1983) because it addresses a fundamental question in cognitive science: how well defined and static are people's concepts (Barsalou, 1982, 1983, 1991; Deák, 1994a)? If people have very well defined and static concepts, we would expect little flexibility—that is, we would expect people to experience difficulty in recategorizing the same set of items according to different criteria. On the other hand, because flexibility entails recategorizing in response to changing tasks, it is fundamentally adaptive. In spite of its significance, there is little research on adults' flexibility in various situations (but see, e.g., Broadbent, Cooper, & Broadbent, 1978; Rabinowitz & Mandler, 1983). Moreover, there are few investigations of young children's ability to categorize the same set of objects in different ways as tasks change (Deák & Pick, 1994, and Kalish & Gelman, 1992, are exceptions). The current study will provide a test of flexibility in a sorting (non-inference) task; if children shift from appearance to taxonomic-based responses as the task changes, we may conclude that both bases for categorization are available to preschoolers, and their choice of one or the other depends on task constraints. In addition, if preschoolers demonstrate flexibility in a non-inductive inference task, it will imply that flexible categorization is not a byproduct of a particular task, but an attribute of the developing organism.

Third, this research is intended to address an important issue in cognitive development, namely, children's understanding of experimental tasks. The importance of task understanding is underscored by the often-obtained finding that preschoolers can perform complex cognitive tasks in a more or less adult-like manner, but only under fairly specific circumstances relative to older children (Gelman & Baillargeon, 1983). These circumstances often involve simplifications of the task and explicit training or modeling. This suggests that young children often either construe tasks differently than the experimenter, or fail to understand them. Similarly, data showing that preschoolers' task understanding has a substantial effect on performance has led to the more general hypothesis (Siegal, 1991) that early conceptual knowledge is difficult to access because its expression is constrained by children's understanding of pragmatics in experimental settings.

Why are findings that task factors affect task comprehension and performance important? Demonstrations of the effect of task comprehension have a direct bearing on assertions about competence: unless the child clearly understands and is motivated to perform the task requested by the experimenter, one cannot draw any valid conclusions about the child's competence. The importance of task comprehension has been shown empirically in studies of preschoolers' performance in, for instance, conservation tasks (Light & Perret-Clerment, 1989), numerical reasoning (Gelman, 1978), and class-inclusion tasks (e.g., McCabe, Siegel, Spence, & Wilkinson, 1982; Trabasso, Isen, Dolecki, McLanahan, Riley, & Tucker, 1978).

To underscore the importance of task comprehension, we draw an analogy to research on children's conservation abilities. Light and Perret-Clerment (1989) review studies showing that preschoolers' answers to post-transformation conservation questions are significantly affected by the social context, including factors that affect their beliefs about what the experimenter wants them to do and say (e.g., McGarrigle & Donaldson, 1975; Rose & Blank, 1974). These findings have increased our understanding of conservation development, and similar findings might shed light on categorization development. Critically, though, here is where the limitations of the analogy become obvious. In conservation tasks there are objectively correct and incorrect answers. Thus, if the child answers incorrectly because s/he did not understand the question or was swayed by demand characteristics, we can say that such factors are influential, but we must not lose sight of the fact that those factors swayed the child into providing an *incorrect* answer (of course, adults can be similarly swayed; see Asch, 1956). In contrast, in the typical categorization task in which children are given a choice between different ways to sort, group, extend a label, or make an inference, there is no objectively correct or incorrect answer. Instead, typically there are several acceptable answers and the question of interest is which one of these the child chooses. Because such tasks are inherently ambiguous, we might expect children to be more easily swayed by minor procedural variations which change their understanding of the task. Therefore it is exceedingly important to investigate the potential effects of different tasks and task variables on children's categorization responses, prior to asserting that children in certain developmental periods generally prefer certain categorization criteria to others. Appearance-based responding in sorting tasks might belie a particular task interpretation, rather than a limitation in children's ability to understand conceptually important relations among dissimilar-looking entities.

Two procedural factors may influence children's understanding of experimental tasks: the instructions they hear, and the training or practice they receive to familiarize them with a specific task. We focus on these variables for several reasons. First, it is possible that the instructions and training used in past studies of preschoolers' taxonomic-based categorization significantly influenced the results of those studies (this argument is detailed below). Second, instructions and training are the two manipulations commonly found in experimental investigations of children's cognitive abilities that most directly and profoundly impact (and are designed to impact) children's task comprehension. Of course, there are probably other variables that affect children's task comprehension, but training and instructions are the most obvious and overt manipulations of comprehension. As such, we might expect the quality (i.e., clarity, specificity, and age-appropriateness) of training and instructions to significantly affect task comprehension. Third, the effects of instructions and training have particular relevance for educational applications, as explained under *General Discussion.* Finally, these variables are relatively amenable to experimental manipulation.

Sorting studies have differed considerably in their methods, but certain trends can be discerned (see below). Past sorting studies (e.g., Fenson et al., 1988; Melkman, Tversky, & Baratz, 1981; Olver & Hornsby, 1964; Tversky, 1985) used instructions and/or training that might have predisposed preschoolers to interpret the tasks as requests for appearance matches. If so, the same factors might be manipulated so that children interpret the task as a request for taxonomic matches. We consider each of these factors, instructions and training, in turn.

Task instructions have been found to influence performance in a variety of cognitive domains (e.g., Bullock & Gelman, 1977; Cheng & Holyoak, 1985; Perry, 1991), including appearance-reality tasks with preschoolers (e.g., Pillow & Flavell, 1985) and sorting tasks with adults (e.g., Kemler, 1984; Medin, Wattenmaker & Hampson, 1987; Medin & Smith, 1981; Smith & Kemler-Nelson, 1984). A variety of instructions have been used in sorting tasks with preschoolers, for example, "tell me what you think goes better with a ____" (Tversky, 1985), choose the ones that "are alike in some way" (Olver & Hornsby, 1966), "point to the ones like this one" (Daehler et al., 1979), and "put the things that are alike or . . . go together into groups" (Denney, 1972). Although these instructions are all somewhat different, they have one common characteristic: they do not precisely specify whether the child should choose an item that looks like the standard or an item that is the same kind of thing as the standard. This does not always present a problem: for example, in a task contrasting a thematic relation (e.g., dog and bone) with taxonomic similarity (e.g., dog and cat), asking for the "ones that are alike" is fairly unambiguous (whereas asking for " . . . the ones that go together" is not; see Bauer & Mandler, 1989). However, in a task placing appearance (e.g., dolphin and shark) in conflict with taxonomic relatedness (e.g., dolphin and seal), such an instruction *is* ambiguous (note that not all of the aforementioned studies were designed to pit appearances against taxonomic membership, so not all of these instructions were ambiguous in their original context).

Of course it might be argued that we *should* present an ambiguous instruction in order to observe the child's interpretation, and draw a conclusion by assuming that children's predilections are reflected in their interpretation. The problem with such a strategy is that in the everyday environment, appearance and taxonomic relatedness are highly correlated. The "conflict situations" presented in these experiments are therefore fairly unusual, and it is even more unusual to present several of them sequentially. If children are not attending closely to the test items, or have not been led to expect conflict situations, they may simply select the most "obvious" answer without investing much effort or consideration into the decision. That is, they may answer impulsively, and make a decision that is not a *preferred* decision. This tendency might be further exacerbated if some children interpret the word "like" in the instruction to mean "look like." On the other hand, instructions to choose "the one that is the same kind of thing" might compel children to look for taxonomic relations in the set of items (Markman, Horton, & McLanahan, 1980). Instructions, then, may help children disambiguate an ambiguous task (that is, one in which at least two responses are possible; see Bauer & Mandler, 1989). Of course, instructions may be only one part of the formula, since (as we have just speculated) children should be more likely to sort according to appearances if they do not expect situations in which appearance and taxonomic relations conflict. This brings us to the second variable of interest.

As with instructions, training procedures have differed across sorting experiments. For example, some studies have provided no training (e.g., Denney & Moulton, 1976; Tomikawa & Dodd, 1980), whereas others used up to eight training sets (Fenson et al., 1988). Yet once again, a loose pattern may be discerned: many of the studies placing appearance and taxonomic relatedness in conflict did not incorporate training that represented the conflict between physical and taxonomic similarity found in the test trios. For example, neither Fenson et al. (1988) nor Tversky (1985) used practice trios that represented a conflict between taxonomic relatedness and similarity of appearance (Fenson et al., 1988, acknowledge that this may have contributed to their subjects' disinclination to choose dissimilar-looking taxonomically related items). Such procedures do not inform children that test items will incorporate conflicts. More seriously, they might actually train children to make appearance-based matches. Since many preschoolers are not accustomed to following instructions in school-like tasks, training might provide important information in ambiguous choice situations.

It is therefore possible that the instructions and training used in previous sorting studies biased subjects to categorize on the basis of appearance rather than taxonomic relations. More broadly, it is possible that training and instructions can bias subjects towards a particular kind of sorting choice. This could explain the discrepancy between studies investigating sorting and inductive inferences, point out a serious qualification to the interpretation of past results, and speak to the construction of future methods. In order to evaluate this possibility, Experiment 1 covaried instructions and training procedures. One instruction and training procedure was intended to support an appearance-based interpretation of the sorting task and was loosely derived from procedures used in previous studies. The other instruction and training procedure was intended to support a taxonomic-based interpretation of sorting task.

Subjects participated in a forced-choice sorting procedure in which they judged sets of three objects. The critical "conflict trios" each consist of a Standard (e.g., a triceratops dinosaur), a Target that is taxonomically related to the Standard but different in appearance (e.g., a stegosaurus dinosaur), and a Distracter that looks like the Standard but is from a more distantly related taxonomic class (e.g., a rhinoceros). Although some Standards and Distracters share membership in a superordinate category, in these cases the Target and Standard are more closely related. (The term "Distracter" is used merely for simplicity, not to imply that it is incorrect to choose that item.) Subjects were asked to sort the Target or Distracter with the Standard.

In Experiment 1 three factors were covaried: age, instructions, training. Subjects were either preschoolers or adults. Inclusion of both children and adults is important for two reasons. First, it is unknown whether task instructions will have a similar effect on both age groups: whereas adults' greater linguistic sophistication could allow for finer discrimination of meaning, if taxonomic sorting is less firmly established in preschoolers, subtle task variations might have a greater effect on their performance. Second, if preschoolers have been hypothesized to have the distinctive characteristic of sorting according to overt appearances, it should be established that older subjects do not exhibit the same behavior.

Two instructions were compared. The instruction "Which one [Target or Distracter] is most like this one [Standard]?" was hypothesized to support appearance-based sorting (hereafter referred to as the *Appearance* instruction). The instruction "Which one [Target or Distracter] is the same kind of thing as this one [Standard]?" was hypothesized to support taxonomic sorting (hereafter referred to as the *Taxonomic* instruction).

Subjects also received one of two training procedures. Training hypothesized to support appearance-based sorting (*Appearance* training) incorporated training trios that could be matched only according to overt physical attributes (e.g., shape and color). Training hypothesized to support taxonomic sorting (*Taxonomic* training) incorporated conflict trios that could be sorted by either appearance or taxonomic relations, and a questioning procedure in which the Experimenter provided justifications for taxonomic-based matches.

Experiment 2 provided additional evidence of the effects of instructions and training on preschoolers' categorization choices. The training procedures were revised (hereafter referred to "Revised Taxonomic" and "Revised Appearance" training) in order to more closely equate extraneous factors and to strengthen and disambiguate the effect observed in Experiment 1. We also collected consistency data and justifications for sorting choices. This was done in order to assess the reliability and validity of children's initial sorting choices, and to determine the reasons for their categorization choices. Finally, we transcribed and analyzed children's spontaneous labels in order to determine whether labeling patterns matched sorting choices.

In both Experiments, all subjects judged "conflict trios" of the kind described above, as well as "no-conflict trios" in which the Standard and Target were both perceptually similar and conceptually related; for example, a brown bull (Standard), a spotted cow (Target), and a coyote (Distracter). There were several reasons for including both kinds of trios. First, it allowed us to determine whether subjects responded randomly, because subjects should choose Target items very consistently in the no-conflict trios. More importantly, the no-conflict trios allowed us to test whether subjects followed an indiscriminate response strategy. For example, if subjects always choose the item that looks like the Standard, in every condition they should systematically choose the Distracter items from the conflict trios and the Target items from the no-conflict trios. Alternately, if subjects simply select the item that looks *least* like the Standard, they will choose the Target items from the conflict trio and Distracter items from the no-conflict trios. Finally, no-conflict trios reduce the artificiality of the task, since perceptual information (e.g., shape and color) usually correlates with taxonomic relations among objects in our environment, and consequently sorting several successive conflict trios may be quite unusual.

## EXPERIMENT 1

### *Stimulus Selection Procedure*

Twelve adults (undergraduates and staff members from a large metropolitan-area University) rated the similarity of appearance of items from within 24 trios of objects. Three-dimensional objects, rather than line drawings, were used for two reasons. First, we wished to maximize ecological validity. Second, categorization may depend on the active, knowledge-driven *selection* of sorting criteria (Murphy & Medin, 1985). For example, categorizing a whale as a mammal rather than a fish depends on ignoring overt cues such as shape and focusing on subtle cues such as skin texture, tail fin orientation, etc. Subtle cues are often lacking in experimental stimuli, and particularly in line drawings. This lack of information impairs taxonomically based categorization (Deák & Bauer, in press), presumably by limiting the information available for active selection. Thus, presenting line drawings to subjects would have artificially decreased subjects' ability to sort taxonomically. All trios consisted of real, functional objects (e.g., quartz crystal), or detailed representations of real objects (e.g., plastic animals). Half of the trios were *conflict trios* in which the Standard and Target objects were conceptually related but perceptually dissimilar, and the Standard and Distracter objects were more conceptually distant but perceptually similar. The other half were *no-conflict trios* in which the Standard and Target objects were conceptually related and perceptually similar, and the Standard and Distracter were conceptually distant and perceptually dissimilar.[2]

[2] There have been few discussions of what constitutes conceptual versus perceptual information and processing. There are reasons to believe that the two are not dissociable, but are interrelated and overlapping (Deák, 1994b; Pick & Heinrichs, 1989). For our purposes we will point out that physical features and conceptual relations are not independent, although examples can be found in which they are to some extent dissociated (e.g., our conflict trios).

Adults were asked to rate the perceptual similarity of the Standard-Target and Standard-Distracter pairs from each trio, presented in random order. Subjects were told to ignore their knowledge of the objects and base their ratings only on appearance.

Adults rated 8 of the 12 conflict trios as having significantly greater Standard-Distracter perceptual similarity (overall $M = 6.9$ out of 10 on a Likert scale) than Standard-Target perceptual similarity (overall $M = 3.0$ out of 10); $p < .05$ for each trio ($t$-scores for the remaining trios were $p > .05$). Adults rated the Standard-Target pair from each of the 12 no-conflict trios as significantly more perceptually similar (overall $M = 8.6$ out of 10) than the Standard-Distracter pair (overall $M = 2.5$ out of 10); $p < .05$ for each trio. Eight of these 12 trios were chosen to reflect the general range of stimuli among the conflict trios (e.g., approximately equal numbers of no-conflict animal trios and conflict animal trios). Thus, 8 conflict and 8 no-conflict trios were selected, and all subjects in Experiment 1 were tested on these 16 sets. Because 3 additional conflict trios were necessary for the Taxonomic training procedure, 12 more adults made perceptual similarity ratings for an additional 7 candidate conflict trios. Of these, 3 trios had significantly greater perceptual similarity between the Standard-Distracter pair (overall $M = 6.0$ out of 10) than the Standard-Target pair (overall $M = 3.9$ out of 10), $p < 0.5$ for each trio, and an additional trio was marginally significant. Three of these trios (including the marginally-significant one) were selected as training trios (sets are listed in Table 1).

*Method*

*Subjects.* Preschool subjects were 48 children (23 F, 25 M), with a mean age of 4.4 (range 4.0 to 4.11) recruited through a volunteer pool drawn from a mid-sized metropolitan area. Adult subjects were 48 undergraduate students and staff members from a large urban university (28 F, 20 M, mean age 25 years) recruited via flyers. Preschool subjects received a small gift; adults were paid \$3.

*Materials.* Sixteen trios of objects constituted the test stimuli. Trios were displayed in $40 \times 32$ cm trays with the Target and Distracter closer to and on either side of the subject, and the Standard in the middle of the tray and closer to the experimenter. Two trios of colored paper shapes were used to train subjects in the *Appearance training* condition; three trios of objects were used to train subjects in the *Taxonomic training* condition (see Table 1).

*Procedure.* Subjects were tested in a room decorated with child-appropriate pictures and furnished with children's furniture. Equal numbers of subjects were randomly assigned to one of two instruction conditions in which they were either asked to "choose the one [Target or Distracter] that is *most like* this one [Standard]" (*Appearance instruction* condition), or to "choose the one [Target or Distracter] that is *the same kind of thing as* this one [Standard]" (*Taxonomic instruction* condition). Within each instruction condition,

TABLE 1
Stimuli in Experiments 1 and 2, Including Training Trio Justifications

| Standard | Target | Distracter |
|---|---|---|
| | *Training trios—Experiment 1* | |
| 1. orange rectangle | orange rectangle | purple oval |
| 2. green circle | green circle | yellow rectangle |
| | *Training trios—Experiments 1 and 2 (Revised Taxonomic justifications) (Revised Appearance justifications)* | |
| 1. white collie | brown dachshund | white Persian cat |
| ("These bark."/"These have baby puppies.")("These are shaggier"/"These are taller") | | |
| 2. red convertible | blue motorcycle | red speedboat |
| ("These have wheels."/"These go on the ground.")("These are longer"/"These are wider") | | |
| 3. polar bear | brown bear | white sheep |
| ("These eat meat."/"These live in the wild.")("These are white"/"These are on all fours") | | |
| [4. raw wool ball | dyed wool yarn | spool of twine] |
| ("These are made of wool."/"These are soft and fluffy.")("These are round"/"These are tan") | | |
| | *Conflict trios* | |
| [1. light bulb | ornamental bulb | juice bottle] |
| 2. brown sugar | sugar cubes | brown dirt |
| 3. flat, white shell | spiral, brown shell | flat, white rock |
| 4. white scallop soap | green soap bar | white pottery scallop |
| 5. black panther | tabby house cat | black stallion |
| 6. purple cedar ball | unpainted cedar egg | purple bath bead |
| 7. round, white balloon | long, blue balloon | round, white candle |
| 8. triceratops | stegosaurus | rhinoceros |
| | *No-conflict trios* | |
| 1. steel fork | steel & resin fork | plastic spoon |
| 2. quartz crystal | quartz crystal | galena crystal |
| 3. French horn | hunting horn | guitar |
| 4. pliers | pliers | hex wrench |
| 5. brown bull | spotted cow | coyote |
| 6. moth | butterfly | goose |
| 7. orange | tangerine | grapes |
| 8. paper clip | paper clip | binder clip |

*Note.* Items in brackets were used in Experiment 2 only. Justifications in parentheses were provided in Experiment 2.

half of the subjects received Appearance training and half received Taxonomic training. Subjects in the Appearance training condition were shown the two construction paper training trios and questioned using the assigned instruction. Subjects invariably chose the Target items during training. Preschoolers in the Taxonomic training condition were presented each of the three training trios, and asked to identify each object prior to answering one of the two categorization questions ("Which one of these [Target or Distracter] is *most like/the same kind of thing as* this one [Standard]?"). Subjects invariably identified each object. They were then asked the categorization question, and those who chose the Target were asked for a justification to ensure that they understood the task. When a child chose a Distracter the experimenter asked her or him to recall the objects' names, clarified the question (e.g., asked the child, "Is a cat the same kind of thing as a dog, or is a dog the same kind of thing as another dog?") and repeated the question. This was meant to point out the disparity between the child's identifications and the conceptual relationships between objects. Although some children were confused by the first training trio, most learned what was being asked of them by the third training trio: children had to be asked the question a mean of 2.25 times for the first trio (because they first made non-conceptual choices, non-conceptual justifications, or both) and a mean of 1.65 times for the third trio before choosing the Target object *and* providing a conceptual justification (matched $t$ test, $p <$ .05). Thus, taxonomic training was effective (although not perfectly effective) in orienting children to the task.

Adults were shown each training trio and given their assigned instruction. If the subject chose the Distracter the experimenter asked, "Do you see more than one possible way to answer the question?" This only occurred three times (out of 144 trials), and in those cases the subject subsequently provided a taxonomic justification for the Standard-Target match.

Following training the experimenter presented the 16 test trios in 1 of 24 quasirandom orders, in which no more than three trios of either type (conflict or no-conflict) occurred sequentially. Left-right placement of Target and Distracter objects was quasi-randomly determined. Subjects were encouraged to examine each of the objects prior to making a judgment. However, subjects were not encouraged to label the objects nor did the experimenter label the objects for the subject. The experimenter repeated the designated instruction (" . . . like . . . " or " . . . kind of thing . . . ") on every trial, and during the trial the experimenter paid equal attention to the Target and Distracter objects (e.g., the experimenter pointed to each object an equal number of times while asking the sorting question, and did not look at either object while subjects made sorting decisions). Subjects indicated their choices either verbally or non-verbally (by pointing or picking up). The experimenter did not request verbal justifications (since providing justifications might have influenced subjects' sorting choices on subsequent trials), nor did the experimenter provide any reinforcement or feedback for the subjects' responses. Sorting responses were recorded during the session by the experimenter. A random sample 25% of the subjects' sorting responses were recoded from videotapes by an independent coder who was unaware of the hypotheses. Intercoder reliability was 99%.

*Results and Discussion*

A percentage score (derived from the number of Target choices out of the maximum possible) was generated for each trio type (conflict and no-conflict) for each subject. Thus, every subject received two percentages between zero and one, with higher scores indicating a preponderance of Target choices. The mean Target choices for the *no-conflict* trios was 97% for children and 100% for adults, indicating almost unanimous choice of the Target object by both age groups, with no difference due to instruction or training condition. This confirms our expectation that subjects did not follow a strategy of choosing items that looked least like the Standard (even in the Taxonomic training + instruction condition), since this would have led to far more Distracter choices from the no-conflict trios. Further, the result strongly suggests that subjects did not answer randomly or indiscriminately. This pattern does not vary between subjects: every subject chose at least 6 out of 8 no-conflict trio Target items. Since there was essentially no variance among the no-conflict trio proportion scores, they were excluded from all analyses. This makes our analysis of the conflict-trio choices more conservative (i.e., decreases the likelihood of finding effects), since the overall variance is increased.

Unexpectedly, adults did not agree upon a taxonomic classification of one of the conflict trios (trio #1): Only 42% of adults sorted the Target item with the Standard in the Taxonomic training + instruction condition. This percentage was far lower than in any other trio in this condition, suggesting that there is no clear taxonomic classification for the Standard item. Consequently, this trio was dropped from all subsequent analyses, and all proportions of conflict trio Target choices are out of the remaining seven trios, rather than eight. An analysis of variance with all eight conflict trios revealed the same main effects and interactions obtained in the 7-item analysis described below, so exclusion of this trio does not affect our results.

Each subject's percentage of conflict trio Target choices was entered into a 2 (age: child, adult) $\times$ 2 (instruction: Appearance, Taxonomic) $\times$ 2 (training: Appearance, Taxonomic) analysis of variance. There was a main effect of age, $F(1,88) = 139.39, p < .001$, reflecting the difference between 73% and 27% overall Target choices by adults and children, respectively. There was a main effect of instruction type, $F(1,88) = 49.01, p < .001$, reflecting the difference between 64% and 36% overall Target choices in the Appearance and Taxonomic instruction conditions, respectively. In addition, there was a main effect of training, $F(1,88) = 6.17, p < .02$, due to more Target choices following Taxonomic training (55% Target trio choices) than Appearance training (45% Target trio choices; see Table 2). These data indicate that instructions and training do have separate and significant effects on subjects' construal of the task and consequently their decision to categorize according to either taxonomic relations and appearances.

Two interactions also reached significance: age $\times$ training, $F(1,88) = 5.45, p < .05$; and age $\times$ instruction, $F(1,88) = 4.79, p < .05$. The former reflects more Target choices following Taxonomic training for children ($p < .05$, Tukey HSD) but not adults (ns. by Tukey HSD). The latter in-

TABLE 2
Mean Percentages of Conflict Trio Target Choices by Age, Instruction, and Training Procedure, Experiment 1

| Condition | Age: Children | Age: Adults |
|---|---|---|
| Appearance training | | |
| Like | 13.1 | 55.9 |
| | (18.7) | (20.6) |
| Kind of thing | 21.4 | 90.4 |
| | (23.2) | (9.3) |
| Taxonomic training | | |
| Like | 21.4 | 54.7 |
| | (18.8) | (33.8) |
| Kind of thing | 51.2 | 92.8 |
| | (7.3) | (7.5) |

*Note.* Proportions are based on seven conflict trios. Numbers in parentheses are Standard deviations. See text for significance values and discussion.

teraction reflects a greater effect of the Taxonomic instruction on adults than children, although children also were affected by instruction ($p < .05$, Tukey HSD). The Taxonomic training procedure had no effect on adults' performance, but it allowed preschoolers to take advantage of the Taxonomic instruction and increased children's Target choices. Apparently, the training variable had little affect on adults' understanding of the experimental task, but instructions did have such an affect. Closer examination reveals that 7 out of 24 (29%) adults who heard the Appearance instruction chose less than half of the Target items from the conflict trios, whereas all of the adults in the Taxonomic conditions chose more than half of the conflict trio Target items. The Appearance instruction (" . . . which one is most like this one," which does not clearly specify either taxonomic- or appearance-based sorting) apparently leads some adults to interpret the task one way and others to interpret the task another way.

It is difficult to interpret the fact that children who received Taxonomic training + instructions chose a mean of 51% conflict trio Target items. One possibility that children's responses were bimodal: some children in this group made mostly taxonomic-based choices, and others made mostly appearance-based responses. Inspection of individual children's responses to the conflict trios, however, belies this possibility: all 12 children chose between 2 and 6 (out of 8) Target items. This leaves at least two possible interpretations. On one hand, this percentage of Target choices is significantly greater than the percentage observed in the Appearance training + instruction condition. It could indicate that children adopted a taxonomic-based response strategy but were limited by the difficulty of the trios, lack of knowledge about the items, or other factors. Thus Taxonomic training and instructions compelled subjects to sort taxonomically, but for some reason was not completely effective. On the other hand, a 51% rate of responding is not different from the rate expected by chance (50%). Children in the Taxonomic training + instruction group did not respond completely randomly, as demonstrated by the near-unanimous choice of Target objects from the no-conflict trios and the fact that the overall (conflict plus no-conflict) percentage of Target choices (73.5%) was significantly greater than chance, $t(95) = 3.15$, $p < .005$. Nevertheless, children may have responded systematically to the easier (no-conflict) items but were confused by the more ambiguous (conflict) items. The issue of preschoolers' construal of the experimental task hinges upon the resolution of this ambiguity: if children in the Taxonomic training + instruction group chose more Target items because of taxonomic relations among those items, then procedural factors (Taxonomic training and instructions) induced a correspondence between the experimenters' and the subjects' understanding of the task. On the other hand, if they chose conflict-trio items randomly, either there was a discrepancy between the task set up by the experimenter and that understood by the subjects, or preschoolers are unable to sort conflict trios according to taxonomic relations without the benefit of labels. More importantly, the former possibility suggests that preschoolers can sort taxonomically when appearances conflict, even without the benefit of labels or other information that specifically points out the taxonomic relations, whereas the latter suggests the necessity of those mediating factors. These alternatives were explored in Experiment 2.

## EXPERIMENT 2

In order to determine whether preschoolers can understand and perform an experimental task requiring taxonomic-based sorting of unlabeled items, we designed a revised version of the Taxonomic training procedure. If subjects in the Taxonomic training + instruction group in Experiment 1 chose more Target items because they understood the task differently, then enhancing taxonomic training should further increase this understanding and lead to a higher rate of Target choices. If, on the other hand, the Taxonomic training procedure confused subjects and led to random conflict-trio sorting, then enhancing the training procedure probably will not reduce this confusion and subjects will still choose at chance levels.

All subjects in the *Revised Taxonomic* training condition heard the Taxonomic instruction from Experiment 1, "Which one of these is the same kind of thing as this one?" In order to clarify that instruction, they were shown two initial stimulus pairs. Moreover, a fourth training trio was added to facilitate abstraction of a common sorting principle, namely, taxonomic relations.

In order to confirm that the Revised Taxonomic training procedure (including the Taxonomic instruction) itself facilitated taxonomic-based sorting, a second group of subjects participated in a *Revised Appearance* training condition. Children in the condition heard the Appearance instruction from Experiment 1, "Which one of these is most like this one?" and the same two initial stimulus pairs were used to introduce the instruction. Children saw the same four training trios as did the Revised Taxonomic training group, but were told different justifications that supported Appearance-based sorting choices (see Table 1).

The Revised Appearance group is critical for interpreting the performance of the Revised Taxonomic group because the two groups are matched except for those factors hypothesized to affect children's interpretation of the task, namely, instructions and the feedback and justifications provided by the experimenter during training. Any differences between the sorting performances of the Revised Taxonomic and Revised Appearance groups therefore are due to those procedural differences.

Several additional responses were elicited and/or collected to provide reliability and validity measures, as well as to provide a more detailed understanding of the effects of the training manipulation. First, data on the consistency of children's sorting responses were collected: after subjects answered the sorting question for all 16 test trios, they saw the test trios a second time and were asked the sorting question again. This second response was used to assess the consistency of their judgments. In the event that subjects in the Revised Taxonomic condition also choose close to 50% conflict trio Target items, consistency data will indicate whether subjects chose randomly, because if they were doing so we would expect low consistency between first and second sorting choices. Similarly, the consistency response provides converging evidence, along with the no-conflict trio responses, of subjects' overall attentiveness and systematic responding. In addition, consistency data are a general indication of the robustness of the training manipulation because they indicate whether the effect lasts beyond the initial presentation of the test trios.

In addition to the consistency response, when shown test trios a second time subjects were asked to provide justifications for their responses. There were two reasons for this. First, justifications provide converging evidence (in conjunction with a comparison of responses to conflict and no-conflict trios) that subjects are not using an indiscriminate response strategy such as choosing the item that looks least like the Standard. Similarly, justifications provide additional evidence that subjects' sorting choices are not random, because we expect justifications to be commensurate with sorting choices. For example, if a subject chooses a Distracter from a conflict trio, we expect that subject to mention overt physical similarities between the Distracter and Standard. Second, and more importantly, justifications should be affected by training. That is, during training the experimenter provides certain kinds of justifications, and it is believed that these will generalize to subsequent test trios, thus affecting children's sorting responses. If the experimenter-provided justifications for the training trios actually mediate subsequent test-trio responses, we expect children to provide relatively more of the *same kinds* of justifications for subsequent test trios. If the training manipulation affects sorting choices, Revised Taxonomic subjects should produce relatively more justifications citing taxonomic relations, whereas Revised Appearance subjects should produce relatively more justifications citing overt physical similarities.

The final measure reported here is spontaneous label production. In Experiment 2, as in Experiment 1, the experimenter did not encourage subjects to label the test trio items. Nevertheless, subjects produced spontaneous labels fairly frequently. They are reported here because labels provide another measure of perceived similarity. That is, when a subject gives the same name to two items from a trio, and gives the third item a different name (or no name), we may conclude that the subject apprehended greater similarity (of some sort) between the items given the same label. Such spontaneous labels are important because they can either match or mismatch the sorting response to a given trio. For example, if a subject categorizes the black panther with the black stallion (Distracter item, conflict trio #5), yet calls both the panther and tabby house cat (Target item) "cat," we conclude that in spite of choosing the Distracter, the subject was nevertheless aware of the taxonomic relation between the Standard and Target. This is important because, as noted in the Introduction, a principal purpose of this research is to see whether children are strongly predisposed to focus on overt appearances, or instead if they tend to sort according to appearances only in response to certain experimental conditions. If subjects produce taxonomic-based labeling patterns on trios from which they made appearance-based sorting choices, this strongly suggests that the subjects were aware of the taxonomic relation but were biased by the sorting procedure to make an appearance-based sorting response. Thus, if there is a larger number of mismatches between conflict trio sorting and labeling responses in the Revised Appearance condition, we can conclude that on at least some trials subjects were aware of the taxonomic relation in the trio but were biased by training to sort according to appearance.

### *Method*

*Subjects.* Twenty-four preschoolers (12 F, 12 M; mean age 4.6; range 4.4 to 4.8) participated. Subjects were from the same population and were recruited as in Experiment 1.

*Materials.* All materials in Experiment 1 were also used in Experiment 2, with two exceptions. First, one conflict test trio was added to replace the test trio excluded from analyses of Experiment 1 (see *Results and Discussion,* above, and *Procedure,* below). Second, several training stimuli not included in Experiment 1 were used. All stimuli not used in Experiment 1 are described below and listed in Table 1.

*Revised taxonomic training procedure.* Several changes were made to the taxonomic training procedure used in Experiment 1. First, two pretraining object pairs were presented prior to the training trios. One pair was a toy shark and dolphin (similar appearance; different taxonomic category), and the other was a toy white wicker chair and a toy red rocking chair (dissimilar appearance; similar taxonomic category). For each pair, subjects first labeled the items. They were then asked, "Are these the same kind of thing or different kinds?" If the child answered "yes" for the chairs and "no" for the dolphin and shark, the experimental responded, "That's right—they're the same kind of thing because they're both chairs," or "That's right—they're different kinds of thing because this one is a shark and this one is a dolphin." If the child gave the opposite response, the experimenter pointed out the taxonomic relation, such as "They are different kinds of things because this one is a shark and this one is a dolphin." Pretraining was intended to ensure that all subjects had some understanding of the phrase, "same kind of thing." Second, an additional training trio was added (see Table 1). This trio had marginally greater Standard-Distracter than Standard-Target perceptual similarity, as determined by stimulus selection in Experiment 1. Each child in Experiment 2 therefore saw a total of four training trios. Third, when children made Target choices during training the experimenter provided two justifications. These justifications are listed in the Table 1, and were provided unless either or both were spontaneously produced by a child. Finally, after providing justifications the experimenter indicated the Standard and Target and said, "Sometimes things look different even though they are the same kind of thing."

*Revised appearance training procedure.* The same two initial pretraining pairs were presented in order to sensitize children to the Appearance instruction. Children labeled the objects and were asked, "Are these two alike or different?" If a child answered "yes" to the dolphin and shark and "no" to the chairs s/he was told, "Yes—these are alike because they're both long and gray and smooth and they have fins;" or "Yes—these are different because this one is big and white, but this one is smaller and red, and it rocks." If a child gave the "incorrect" answer, the experimenter provided a justification for the opposite response, such as "They are alike because they are both long and gray and smooth, and they have fins."

Children saw the same training trios as the Revised Taxonomic training group. However, they heard justifications focusing on the physical similarities between the Standard and Distracter items. These justifications are listed in Table 1. After providing justifications for each training trio, the experimenter said, "You see, these [indicating Standard and Distracter] are alike in a lot of ways."

*Testing procedure.* The testing procedure was similar to that used in Experiment 1 (as in Experiment 1, subjects were given no reinforcement or feedback for sorting choices), with two modifications. First, an additional test trio was added to replace the one excluded from Experiment 1. The replacement trio (trio #1; see Table 1) had significantly greater Standard-Distracter than Standard-Target perceptual similarity ($p < .05$), as rated by adults in the second stimulus selection pretest for Experiment 1. Second, after subjects completed sorting judgments for all of the test trios, they were shown the test trios again. They were then re-asked the categorization question. Note that children were not asked to repeat their previous answer but were simply asked the question a second time. After re-questioning, subjects were asked to justify their choice. The experimenter elicited justifications by asking, "Why are these [indicating Standard and subject's choice] *the same kind of thing* (*alike*)? What makes them *the same* (*alike*)?" Subjects were asked repeatedly for justifications until they indicated that they could not think of any more reasons.

*Scoring*

Initial sorting responses were scored as in Experiment 1 (i.e., "Target" or "Distracter"). Consistency (second) sorting responses were scored as *same* if the subject chose the same item the first and second time, or *switch* if the subject chose different objects the first and second time.

Justifications were assigned to one of three categories by two independent coders. *Taxonomic* justifications referred to basic-level nominal essence (e.g., "These are both light bulbs") or shared superordinate class ("These are both insects"). *Relational* justifications referred to origins ("They came from a river"), context ("I saw them both at the zoo"), or dynamic or causal properties ("These pop if you stick them with a pin," "You can eat with them"). *Physical* properties referred to static, perceivable attributes of objects ("They're both white," "They're both heavy"). More precise definitions and additional examples of subcategories within each of the three major justification type (Taxonomic, Relational, and Physical) are shown in Table 3. An additional *Ambiguous* category encompassed responses that did not fit into a well-defined category (e.g., "One's a bottle and one's a lamp, and lamps and bottles are kinda alike"). Only justifications specific to the Standard and the sorting choice item were coded and analyzed. For example, with regard to the no-conflict trio butterfly (Standard), butterfly (Target), and goose (Distracter), the justification "These are insects" is specific to the Standard and Target because it differentiates them from the Distracter. In contrast, the justification "These have wings" is not specific to the Standard and Target because it is also true of the Distracter. Two independent coders assigned a subcategory type to every such justification. Intercoder agreement was 92.7%; disagreements were resolved by discussion.

Consistency responses and justifications were recorded by the experimenter during the session. A second, independent coder who was unaware of the hypotheses watched randomly selected videotapes of 25% of the subjects and recorded consistency responses and justifications. Reliability for consistency responses was 100%, and reliability for justifications was

TABLE 3
Justification-Type Subcategories: Examples and Total Numbers Produced in the Revised Taxonomic and Revised Appearance Training Conditions, Experiment 2

| Property type & subcategory | Example | Total no. rev. tax. | Total no. rev. app. |
|---|---|---|---|
| Taxonomic | | | |
| Basic-level name | "These are both cows." | 81 | 42 |
| Superordinate name | "These are kinds of tools." | 2 | 5 |
| Relational | | | |
| Origins | "They came from the sea." | 2 | 1 |
| Location/context | "These two go in a parade." | 18 | 0 |
| Dynamic/casual properties | "My dad uses them to fix things." | 75 | 29 |
| Physical | | | |
| Overall | "They look the same." | 5 | 2 |
| Substance/parts | "They're both made of glass." | 21 | 37 |
| Shape | "Both are round." | 4 | 32 |
| Color | "They're both orange" | 27 | 103 |
| Size/weight/density | "Both are hard." | 5 | 26 |
| Texture/reflectance | "These are shiny." | 3 | 7 |
| Other | "They make a squeaky noise." | 12 | 2 |
| Ambiguous | "They're both the same." | 7 | 10 |

89.3% (because disagreements about justifications seemed to result from uneven audibility of children's utterances on the videotapes, the Experimenter's notes were used in cases of disagreements).

Two independent coders unaware of the hypotheses of the study each transcribed half of the children's spontaneous labels. A third coder then coded the relation between the labels and the sorting choices. A codable labeling response occurred when a child produced nouns for all three items in a trio (or two of the three items, in situations described below).

There were three kinds of labeling responses, each of which reflected different relations between the labels and sorting choice for a trio. *Matches* occurred when children produced the same name for the two objects that they sorted together. For example, if a child sorted the Distracter from trio #1 (bottle) with the Standard (light bulb) and called them both "bottle," it would be considered a *Match* response (unless the child also labeled the Target item "bottle"). *Nonmatches* occurred when items sorted together were not distinguished from the non-chosen item by their labels. This happened when either all three items were given different labels, all three items were given the same label, or the two items sorted together 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," it would be a *Nonmatch*. Finally, *Mismatches* occurred when subjects gave the same label to the Standard and to the item that was *not* chosen as a sorting match. For example, if a child sorted the Standard (light bulb) with the Distracter (bottle), but called both the Standard and *Target* "light bulb" (and either called the Distracter something different (e.g., "bottle") or did not label the Distracter), it would be a *Mismatch.*

*Results*

Initial sorting responses were coded as in Experiment 1, yielding two scores for each subject: percentage of conflict trio Target choices, and percentage of no-conflict trio Target choices. Overall, children chose 99% no-conflict trio Target items in the Revised Taxonomic training group and 99% in the Revised Appearance training group. This indicates that children were generally attentive and systematic in the sorting task, and that children in the Revised Taxonomic group did not use an indiscriminate strategy such as choosing the item that looked least like the Standard.

Children in the Revised Taxonomic group chose 65.0% conflict trio Target items overall ($SD$ = 18.6), which is significantly greater than chance (50%), $t(11) = 2.47$, $p < .025$ (one-tailed). This suggests that the Revised Taxonomic training procedure induced children to adopt a sorting strategy based on taxonomic relations rather than (conflicting) similarity of appearance. Children in the Revised Appearance group chose 12.5% conflict trio Target items ($SD$ = 13.4), which is significantly less than chance, $t(11) = 9.21$, $p < .01$ (one-tailed). This indicates that the Revised Appearance training procedure induced children to sort according to similarity of appearance rather than (conflicting) taxonomic relations. Because these data include responses to replacement trio #1, which was not used in Experiment 1, separate analyses were performed on each subject's responses to the seven conflict trios excluding replacement trio #1. Subjects in the Revised Taxonomic condition made an average of 66.7% Target choices, which is significantly greater than chance, $t(11) = 3.1$, $p < .01$ (one-tailed). Subjects in the Revised Appearance condition made an average of 11.9% Target choices, which is significantly less than chance, $t(11) = 9.85$, $p < .001$. The results of Experiment 2, then, clearly are not attributable to the replacement conflict trio.

Consistency data provide converging evidence that preschoolers' interpretation of the task was affected by training. Children's second sorting responses were almost always identical to first responses: only 12 of the 192 total sorting responses (3.2%) switched when children were asked a second time. Thus, children's choices were systematic. Notably, 11 of the 12 changes occurred in conflict trios, suggesting that children were sometimes aware that both choices were appropriate.

Subjects produced 541 choice-specific justifications: 258 for conflict trios and 283 for no-conflict trios. 255 of the justifications were produced by subjects in the Revised Taxonomic condition, and 286 by subjects in the Revised Appearance condition. Thus, justifications were produced with approximately equal frequently in both conditions and both kinds of trials. Children's justifications were generally clear and meaningful: only 17 of the 541 responses (3.1%) were *Ambiguous.*

Numbers of justifications of the three major types (*Taxonomic, Relation, Physical*) produced in the Revised Taxonomic and Revised Appearance conditions are shown in Table 4. In order to test the apparent difference between the distribution of justifications across types in the two conditions, the number of Physical justifications produced by a subject was subtracted from the total number of Taxonomic and Relational justifications produced by that subject. This yielded a difference score for each subject, which were on average positive for subjects in the Revised Taxonomic condition ($M = 8.7$, $SD = 5.8$) and negative for subjects in the Revised Appearance condition ($M = -11.1$, $SD = 11.4$). The average difference between the number of Physical justifications and the total number of Taxonomic and Relational justifications differed significantly between the two groups, $t(22) = 5.4$, $p < .01$. In addition, most children produced justifications consistent with their condition: all children in the Revised Taxonomic condition produced more Taxonomic and Relational justifications than Physical justifications, whereas 10 out of 12 children in the Revised Appearance condition produced more Physical than Taxonomic and Relational justifications. This difference is significant, $\chi^2\,(1) = 18.7, p < .001$.

Not only did justifications differ for conflict and no-conflict trios, but within conflict trios they differed for Target and Distracter choices. Children across the two conditions justified Target choices by reference to 43 Taxonomic, 52 Relational, and 28 Physical properties. These same children justified their Distracter choices by reference to 8 Taxonomic, 11 Relational, and 116 Physical properties. Clearly, children justified Targets choices in terms of Taxonomic and Relational properties, whereas they justified Distracter choices in terms of Physical properties.[3]

The relations between spontaneous labels and sorting choices in Experiment 2 were as follows: in the Revised Taxonomic condition there were 97 *Matches*, 23 *Nonmatches*, and 6 *Mismatches;* in the Revised Appearance condition there were 29 *Matches*, 8 *Nonmatches*, and 14 *Mismatches*. This distribution is striking because relatively more mismatches occurred in the Revised Appearance condition than the Revised Taxonomic condition: 27% of the labels in the Revised Appearance condition were Mismatches, compared to only 5% in the Revised Taxonomic condition. This indicates that subjects in the Revised Appearance condition occasionally chose Distracters but were aware of the taxonomic relation between the Standard and Target, as reflected by their labeling patterns.

[3] Due to the nature of these data (e.g., the fact that the same child produced multiple justifications), they are not amenable to nonparametric statistical tests (see Siegel & Castellan, 1988). Similar constraints prevent statistical analysis of the data on spontaneous labeling.

TABLE 4
Numbers of Major Justification Types (Taxonomic, Relational, and Physical) by Condition (Revised Taxonomic or Revised Appearance), Experiment 2

| | Justification type | | |
|---|---|---|---|
| Condition | Taxonomic | Relational | Physical |
| Revised taxonomic | 83 | 95 | 77 |
| Revised appearance | 47 | 30 | 209 |

Most labeling responses matched sorting responses (71%). It is not surprising, then, that on *Match* trials subjects usually chose the Target item (because overall 68% of sorting choices were Target choices). Specifically, there were 118 Matches and 3 Mismatches among Target choices, compared to 8 Matches and 17 Mismatches among Distracter choices. The fact that 17 out of the 20 Mismatches (85%) occurred in trios from which subjects chose the Distracter supports our interpretation of the relationship between labeling and sorting; namely, that children who chose Distracter objects were nonetheless (at least occasionally) aware of the Standard–Target relation.

*Discussion*

The results of Experiment 2 provide converging evidence that training (including instructions) can powerfully affect preschoolers' sorting choices when appearances and taxonomic relations conflict. Children in the Revised Taxonomic training condition made a mean of 65% conflict trios Target choices, compared to 12% in the Revised Appearance group, suggesting that the effect of training found in Experiment 1 is reliable and replicable. The high rate of Target choices in the Revised Taxonomic condition cannot be attributed to a strategy of choosing the item that looked less like the Standard, since such a pattern is contraindicated by children's responses to no-conflict trios. These data strongly suggest that training, which provided children with "model" justifications for particular kinds of relations and similarities, strongly influenced children's sensitivity and preference for similar relations in subsequent test trials.

Children demonstrated uniform and high consistency when re-asked the categorization question, indicating that children's initial sorting choices were not random or indiscriminate, and that the training effect was robust, because children retained the tendency to make training-appropriate sorting choices during re-sorting. As previously noted, the few inconsistent responses suggest that preschoolers may sometimes have been aware of the different bases for sorting the conflict trios.

Children's spontaneous labels provided both converging evidence for the sorting data (because most labels matched sorting choices) and additional evidence that children who made appearance-based sorting choices nevertheless (at least occasionally) were aware of the taxonomic relation between

Target and Standard, because most mismatches occurred when subjects chose the Distracter object.

Children's justifications were consistent with their sorting choices. The fact that children produced more Taxonomic and Relational justifications when they chose Targets and more Physical justifications when they chose Distracters indicates that children's sorting choices and justifications are systematic and predictable. It is noteworthy that approximately half of the total justifications (255 out of 541, or 47%), were based on non-Physical attributes—that is, Taxonomic or Relational connections. This provides further evidence that preschoolers are not strongly predisposed to sort by appearances, because apparently they can justify categorization choices by referring to non-obvious attributes.

Justifications speak to a long-standing issue in the literature on children's categorization—namely, the salience of various categorization criteria. A great many theorists have, at one time or another, suggested that some bases for categorization are more salient than others. For example, theorists have argued whether children use shape (e.g., Gentner, 1989; Tomikawa & Dodd, 1980) or function (Nelson, 1973) to generalize word meaning and sort objects. Other theorists have questioned whether shape or color is more important for certain categorization functions (e.g., Baldwin, 1989). Others have debated whether preschoolers prefer taxonomic or thematic relations (e.g., Denney & Moulton, 1976; Greenfield & Scott, 1986). Still others, as noted in the Introduction, have asked whether overall appearance or taxonomic relations determine categorization decisions. All such hypotheses implicitly suggest that preschoolers possess a "response hierarchy" specifying which categorization criterion is most salient, which is next most salient, etc. Presumably, children prefer to use the most-salient sorting criterion available in a given set of stimuli. An alternative to these "response hierarchy" hypotheses is that children's preferred categorization criteria depend on task variables, contextual variables, or both (Deák & Bauer, in press; Deák & Pick, 1994; Jones, Smith & Landau, 1991; Smith & Heise, 1992; Smith & Jones, 1992). If our subjects were constrained by rigid response hierarchies, the distribution of justifications would have been invariant across the two training conditions. For example, if the preference for color-, shape-, and size-based justifications found in the Revised Appearance condition had also been found in the Revised Taxonomic condition, this would have evinced a preference for those factors. Because the relative distribution of different kinds of justifications differed in the two conditions, the data support the view that categorization criteria are task and context dependent.

It might be argued that the relative frequencies of different kinds of justifications indicates that some attributes are more salient that others. For example, in both conditions there are many more color-based justifications than size- or texture-based justifications. This finding is somewhat equivocal with regard to the question of response hierarchies. On the one hand, children know many more color words than texture words, so it may not be surprising that fewer texture-based justifications were produced (the fact that English-speaking children learn more color words than texture words is itself interesting, but discussion of this fact is beyond the scope of this article). On the other hand, the fact that children also produced fewer size-based justifications does lend some credence to the possibility of an attribute hierarchy, because preschoolers know several comparative size terms, and size terms were modeled three times in the experimenter-produced training justifications. In sum, then, the numbers of various kinds of justifications do not conclusively address the question.

The fact that the two training groups made significantly different kinds of sorting choices supports our hypothesis that procedural factors such as instructions and training bias children's interpretation of sorting tasks. This finding clearly has implications for past research on preschoolers' sorting: results indicating that children tend to sort by appearances (e.g., Fenson et at., 1988; Melkman et al., 1981; Tversky, 1985) may be attributed to test procedures rather than a general age-related preference for appearances. These data also speak to the construction of future tests of preschoolers' sorting abilities: such tests should take into account the possibility that some seemingly innocuous methods bias children's preference for one kind of relation or another. Finally, there is a broader implication for preschoolers' categorization in general. The fact that children from the same population sorted either by appearance or by taxonomic relations suggests that both strategies are available to children, and their preference is affected by the task at hand. Thus, children are not restricted to sort according to a static "response hierarchy" which leads them to invariably prefer some categorization criteria to others. Instead, a range of categorization criteria are available to preschoolers, and their choice depends on task factors.

## GENERAL DISCUSSION

A preponderance of evidence now suggests that young children can categorize objects according to a number of different kinds of category relations, such as taxonomic, thematic, and appearance. A growing body of evidence outlines factors that affect preschoolers' preference for one or another of these relations. This work has focused on preferences for thematic- or taxonomic-based sorting (e.g., Bauer & Mandler, 1989; Markman & Hutchinson, 1984; Smiley & Brown, 1979), the role of labeling in taxonomic-based sorting and inference-making (e.g., Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Markman & Hutchinson, 1984; Waxman & Gelman, 1988; Waxman & Kosowski, 1990; Waxman & Hall, 1993; although in most of these studies subjects could have been sorting on the basis of appearances), and inference-making on the basis of taxonomic relations versus appearances (e.g., Deák & Bauer, in press; Gelman & Markman, 1986, 1987; Gelman & Coley, 1990). A few studies have even explored non-linguistic factors affecting preschoolers' categorization choices in a sorting task (e.g., Bauer & Mandler, 1989; Deák & Pick, 1994; Markman et al., 1981). Yet no studies have examined factors affecting children's preference for taxonomic- versus appearance-based sorting in a unlabeled sorting task.

The present research was designed to explore the role of procedural factors in preschoolers' preference for taxonomic- or appearance-based sorting. Procedural factors were hypothesized to have an effect because many of the empirical studies reporting a preference for appearance-based responding (e.g., Fenson et al., 1988; Melkman et al., 1981; Tversky, 1985) incorporated subtle procedural details (e.g., instructions and training) that may have biased children to sort by appearance.

The results of Experiments 1 and 2 support this hypothesis by demonstrating that preschoolers' tendency to sort according to appearance or taxonomic relations can be strongly affected by different instructions and training. This was most evident in Experiment 2, in which children trained and instructed to sort by appearances did so 88% of the time, whereas children trained and instructed to sort by taxonomic relations did so 65% of the time. The only differences between these conditions were the instructions and the feedback and justifications given during training. Although training and instructions were not independently varied in Experiment 2, Experiment 1 showed that each of these variables has an effect on children's sorting choices. Furthermore, the effect is not limited to preschoolers. Adults in Experiment 1 also showed an effect of instructions, indicating that sensitivity to procedural factors in sorting tasks is not constrained to a specific age range. The fact that adults were sensitive to instructions but not training is intriguing, but currently we cannot explain this difference.

The validity of preschoolers' sorting data is supported by consistency data, justifications, and spontaneous labeling in Experiment 2. The consistency data showed that preschoolers sorted systematically (also shown by responses to the no-conflict trios) and that the effects of training persisted into a subsequent re-sorting period. Children's justifications also demonstrate the validity of their sorting choices: justifications were almost always consistent with sorting choices, and differed in the predicted manner across conditions. Of course, justifications should be treated cautiously since we do not know the extent to which they reflect the actual processes underlying the sorting choice as opposed to ad hoc reasoning in response to the experimenter's question, nor do we know the extent to which they are constrained by the child's productive language and vocabulary. In addition, since subjects were exhaustively questioned for justifications, and therefore often produced more than one type of justification for a given trio, we cannot determine *which* justification, if any, was preferred. Finally, children's spontaneous labels generally matched sorting responses, suggesting that children persistently apprehended a more compelling relation between the Standard and sorting-choice item.

There was some hint, interestingly, that children were aware of both answers in the conflict trios, since 11 out of 12 (92%) of the sorting choice switches occurred in conflict trios. This is reinforced by the fact that 85% of the Mismatches between sorting choice and spontaneous labels occurred when children chose the Distracter object, and all but one Mismatch occurred in conflict trios. Because of their low frequency, these findings are merely suggestive. Nevertheless, they may be taken to indicate that children are sometimes aware of both taxonomic relations and appearances, and they choose one of these relations in response to task factors. Children's sensitivity to both relations is also suggested by the experimenters' and coders' subjective (but consensual) impression that children took more time to categorize conflict trios than no-conflict trios. The possibility that preschoolers can be simultaneously aware of multiple ways to categorize a given set of items is extremely interesting and important, because it indicates flexibility in categorization (i.e., the ability to re-categorize a set of items according to different criterion), a little-understood problem in the study of categorization.

Nevertheless, these data are suggestive of flexibility in preschoolers' sorting decisions, since task factors strongly influenced children's proclivity for appearance- or taxonomic-based sorting. This finding is incompatible with a common (sometimes implicit) family of hypotheses in the developmental literature which suggest that young children have a categorization criterion "response hierarchy" in which some criteria are more salient or compelling than others. By this view, children are predisposed to use the most-salient feature available in the stimuli as a categorization criterion.

In a trivial way this hypothesis is obviously correct: under no circumstances will we find preschoolers sorting animals according to the infrared-spectrum wavelengths they reflect, or sorting rocks according to caloric content per liter. But within the realm of reasonably plausible categorization criteria, it is possible that children's choice of such criteria depends strongly and significantly on their reasons for categorizing. That is, task and context factors may play a more significant role in categorizing than do children's preferences for certain features over others. This may even be true for transitory and accidental features such as "dirty" or "sleepy" or "fell on the floor" (see Gelman, 1988). Although preschoolers may never use these attributes in an artificial, decontextualized experimental task, one can imagine everyday situations in which such sortals are used consciously and effectively. For example, we can easily imagine a situation in which entities are sorted by sleepiness: Day-care workers may decide, "Those children who are sleepy may take a nap now and have snack later, whereas those who are not sleepy may have a snack now and nap later." Gelman (1988) may be correct that such features usually do not predict more enduring properties, but it is clear that particular tasks call for very different categorization criteria and groupings, and whereas there is some research on such "ad hoc" groupings in adults (e.g., Barsalou, 1983, 1991), there has been very little investigation of children's use of such groupings. Nevertheless, the current findings as well as others (e.g., Coley, 1994; Deák & Pick, 1994; Kalish & Gelman, 1992) strongly suggest that preschoolers are neither limited nor strongly predisposed to use particular categorization criteria in all situations, but that they are quite sensitive to local task constraints.

The current findings can be integrated into a descriptive taxonomy suggested by Deák and Bauer (in press) for describing the dynamics among the various factors affecting people's categorization decisions. This taxonomy differentiates three broad, non-orthogonal factors: characteristics of the *subject*, characteristics of the *task*, and the *information* available to the subject. Developmental theories such as Piaget's have typically focused on the causal effects of only one variable representing the first (subject) factor, namely age. More recent studies, such as Gelman and Markman (1986, 1987), have focused on a variable representing the latter factor (information), namely labeling. This study represents the remaining factor, namely task characteristics. Without evidence of the effects of each of these factors and interactions among the factors, any theory of children's categorization risks neglecting potentially critical variables and interactions among variables. For example, Experiment 1 indicated that both children's and adults' performance is affected by their task comprehension, and that children's task comprehension might be susceptible to certain variables (e.g., training) that less-powerfully influence adults' task comprehension.

It might be suggested that in order to compel children to categorize taxonomically (Revised Taxonomic training condition, Experiment 2), we had to impose rather extreme interventions (e.g., modeling certain kinds of justification for taxonomic-based sorting choices). Perhaps the need for stronger manipulations, combined with the fact that children in this group nevertheless made appearance-based decisions for 35% of the conflict trios, shows that taxonomic-based categorization is more difficult for children than is appearance-based categorization. There are several responses to this suggestion.

The first response is that it indeed may be difficult for some children to use taxonomic relations. For example, two of the children in the Revised taxonomic condition made 6 out of 8 Distracter choices in the conflict trios. When these subjects are excluded, the overall rate of conflict-trio Target choices rises to 73%. However, this fact does not really address the question because we do not know whether those two subjects had global difficulty representing taxonomic relations among objects, or whether they still did not understand—or chose to ignore—the task specified by training and instructions.

It may be more fruitful to address the issue conceptually, and here there are at least three relevant issues. First, it has already been stated that appearance and taxonomic relations tend to be highly correlated. For example, most goldfish look more similar to other goldfish than to, say, Dachshunds or petunias. Young children may use this as a heuristic—indeed, even adults perform simple sorting tasks according to overall appearances when distracted or working under time pressure (Smith & Kemler Nelson, 1984). Preschoolers are nevertheless capable of categorizing by Taxonomic relations without perceptual support or taxonomic labels, but they may need explicit warning that the appearance heuristic is invalid for the task at hand. We attempted to provide such warning during training, but may not have been completely successful. Second, it is not clear that our manipulations were so extreme. Our training manipulation required children to abstract a sorting strategy across only four training trios incorporating very different items and justifications, and subsequently apply it to test trios for which they received no feedback or reinforcement. In this light, our training could be seen as rather minimal and subtle, thus leading us to wonder (contrary to the prior suggestion) how such a subtle manipulation profoundly affected children's behavior. Note, here, that 3 of the 8 justifications provided during Revised Taxonomic training were *Physical,* which may have weakened the desired manipulation. Further, we imposed manipulations within a task that is artificial and unfamiliar to preschoolers. It is difficult to say exactly how extreme our training was, given that it occurred in an unusual context. Perhaps in an everyday situation that inherently calls for ignoring appearance and focusing on taxonomic relation, preschoolers need no social input whatsoever. After all, a stuffed dog might look very much like a real dog, but preschoolers neither pet nor try to feed stuffed dogs (except in pretend play). Third, categorizing according to taxonomic relations in the absence of similar appearances often depends on specific factual knowledge. For example, unless you know that dolphins bear live young and surface to breath, you might not be able to judge a dolphin a mammal (and of course, you must first know that such attributes are characteristic of mammals). Young children often do not possess the factual knowledge necessary to make such judgments about taxonomic relations, and although we selected conflict trios in which the taxonomic relations were likely to have been within many middle-class American 4-year-olds' knowledge, some subjects were no doubt ignorant of some relations. These knowledge gaps alone might account for the 35% Distracter choices in the conflict trios of the Revised Taxonomic condition (however, we have no direct evidence for this possibility).

The idea that preschoolers' categorization decisions depend on task factors—or, more accurately, factors that affect their understanding of the goals of a task—leads to a particular interpretation of experimental categorization paradigms. In the current paradigm, children are presented with sets that pose a conflict between two possible bases for categorization, taxonomic relations and appearance. In this task there are two appropriate responses, taxonomic-based and appearance-based. Procedural factors influence the probability that a child will choose one of these relations. In early work that led to the belief that children are "perceptually bound," or limited to categorize by appearance (e.g., Inhelder & Piaget, 1964; Olver & Hornsby, 1966), researchers' methods (e.g., instructions and training or lack thereof) biased subjects' interpretation of the task so that a particular response strategy—categorizing by appearances—was favored. Later research, although methodologically much improved, incorporated subtler but nevertheless potentially biasing procedures (e.g., Fenson et al., 1988; Melkman et al., 1981; Tversky, 1985). Two such factors, training and instruction, were investigated here. Other factors have been investigated elsewhere: Deák & Bauer (in press) found, for example, that children and adults are more likely to make appearance-based categorization decisions when the stimulus items are line drawings than when they are real objects. Real objects apparently contain subtle physical information upon which children can base taxonomic choices, even when overall appearances are in conflict. Line drawings, in contrast, do not portray subtle physical features and therefore preclude the use of such features as a basis for categorization. In support of this argument, 46 of the 541 justifications (8.5%) produced in Experiment 2 referred to features not depicted in line drawings (e.g., substance, texture, density, etc.).

Together, these two studies provide important caveats for the interpretation of past research on children's categorization, as well as principles for the design of future research. In the current study, for example, the fact that two procedural factors—instructions and training—designed to impart the nature of the task to the child significantly affected their interpretation of the task suggests that the same factors were operating in past studies (e.g., Fenson et al., 1988; Melkman et al., 1981; Tversky, 1985) of children's taxonomic-based categorization across variations in appearance. Because these studies typically used instructions and training that closely resembled our appearance-based procedures, it is reasonable to assume that their results, which typically supported the hypothesis that preschoolers cannot categorize according to taxonomic relations when appearances differ, were affected by their procedures (however, it should be noted that Fenson et al.'s subjects were considerably younger than our own). In addition, the results motivate a caveat for the design of future research: it is clear that in tasks requiring children to choose between two or more acceptable responses (e.g., most experimental categorization tasks), the instructions and training must be very carefully constructed so that they do not unintentionally bias children toward one kind of response. Of course, it is not necessary to do so if the results of studies using potentially biasing procedures are not generalized beyond those procedures.

The relevance of these findings extends beyond the straightforward (but nonetheless critical) methodological issue concerning the interpretation of past research and design of future research, and beyond the important issue of flexibility discussed above. They are important for several additional reasons. First, they add to our knowledge of non-linguistic factors affecting categorization decisions. It is clear that children's ability to categorize taxonomically without the benefit of similar appearances is not strictly tied to linguistic information, or to information denoting conceptual connections between taxonomically related items. Rather, preschoolers have the ability to focus on taxonomic relations without the benefit of any "special" information (linguistic or conceptual) provided by the experimenter: the child merely needs to understand that s/he is being asked to point to the taxonomic relations, and needs enough information to select that relation.

Second, the results point to the importance of task comprehension. As several theorists have pointed out (Gelman, 1978; Siegel, 1991), it is invalid to draw conclusions about children's competence without also demonstrating that performance was not impeded by task-specific attention and memory factors or their construal of the task. Demonstrations of the effects of instructions and training on task comprehension have further-reaching relevance. As pointed out by Siegal (1991), Piaget's assertion that preschoolers are perceptually bound has serious educational implications because it implies that preschoolers cannot benefit from instruction in abstract (e.g., mathematical and scientific) concepts, which therefore should not be taught until children are older. If task comprehension has a significant affect on children's abstract reasoning, it would suggest, first, that it *is* possible to teach children abstract concepts, and second, that the procedure for orienting children to the task must be constructed very carefully. Training and instruction are therefore ecologically important variables, the importance of which is directly proportional to the cultural prevalence of and value placed on early educational programs and materials. In general, these results, as well as many others (reviewed by, for example, Gelman, 1978; Gelman and Baillargeon, 1983; and Siegal, 1991) suggest the general principle that preschoolers are quite sensitive to a variety of procedural factors operative in most experimental testing situations. This sensitivity, which Piaget construed as a absence or fragility of cognitive operations, may instead be construed as a tendency to respond to many social and non-social factors that often are arbitrarily designated as "incidental." This, then, suggests a very different picture of the preschooler than the egocentric, perceptually bound organism of Piaget's theory. Instead, it suggests an organism that has been rapidly acquiring a great deal of information about the social and physical world, and one who utilizes much of this knowledge when placed in an ambiguous or confusing situation.

## REFERENCES

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, **70**(9, Whole No. 416).

Baldwin, D. A. (1989). Priorities in children's expectations about object label references: Form over color. *Child Development*, **60**, 1291–1306.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, **10**(3), 211–227.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition,* **11**(3), 211–227.
Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *Psychology of learning and motivation: Advances in theory and research* (Vol. 27). New York: Academic Press.
Bauer, P. J., & Mandler, J. M. (1989). Taxonomies and triads: Conceptual organization in 1-to-2-year-olds. *Cognitive Psychology,* **21,** 156–184.
Broadbent, D. E., Cooper, P. J., & Broadbent, M. H. P. (1978). A comparison of hierarchical and matrix retrieval schemes in recall. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 486–497.
Brown, A. L. (1989). Analogical learning and transfer. What develops? In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 369–412). Cambridge: Cambridge Univ. Press.
Bullock, M., & Gelman, R. (1977). Numerical reasoning in young children: The ordering principle. *Child Development,* **48,** 427–434.
Carey, S. (1985). *Conceptual change in childhood.* Cambridge: MIT Press.
Cheng, P. W. & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology,* **17,** 391–416.
Coley, J. D. (1994, April). *Emerging differentiation of folkpsychology & folkbiology.* Poster presented at the Conference of Human Development; Pittsburgh, PA.
Daehler, M. W., Lonardo, R., & Bukatko, D. (1979). Matching and equivalence judgments in very young children. *Child Development,* **50,** 170–179.
Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development,* **5,** 151–176.
Deák, G. (1994a). *Flexibility of categorization in children and adults.* Unpublished Orals paper, University of Minnesota.
Deák, G. (1994b). *The role of Descartes' framework in explanations for developmental phenomena.* Paper presented at the annual symposium of the Jean Piaget Society, Chicago, IL., June.
Deák, G. (1995). *The functional nature of induction: Flexibility in children's inferences about object properties.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.
Deák, G. O., & Bauer, P. J. (in press). The dynamics of preschoolers' and adults' categorization choices. *Child Development.*
Deák, G. O., & Pick, A. D. (1994). *What is a penlight? Preschoolers' systematic and flexible use of abstract categorization criteria.* Paper presented at the annual convention of the American Psychological Society, Washington, DC, July.
Denney, N. W. (1972). A developmental study of free classification in children. *Child Development,* **43,** 221–232.
Denney, N. W. & Moulton, P. A. (1976). Conceptual preferences among preschool children. *Developmental Psychology,* **12,** 509–513.
Dupré, J. (1981). Natural kinds and biological taxa. *The Philosophical Review,* **40**(1), 66–90.
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. (1985). *Cognitive Development* (2nd ed.). New Jersey: Prentice–Hall.
Gelman, R. (1978). Cognitive development. In M. Rosenzweig & L. Porter (Eds.), *Annual review of psychology* (Vol. 29, pp. 297–332).
Gelman, R. & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman, (Eds.), *Handbook of child psychology. Vol. 3: Cognitive development.* New York: Wiley.
Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology,* **20,** 65–95.
Gelman, S A., & Coley, J. D. (1990). The importance of knowledge a Dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology,* **26,** 796–804.

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

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

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

Gelman, S. A., & O'Reilly, A. W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*, **59**, 876–887.

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* (pp. 199–241). Cambridge: Cambridge Univ. Press.

Greenfield, D. B., & Scott, M. S. (1986). Young children's preference for complementary pairs: Evidence against a shift to taxonomic preference. *Developmental Psychology*, **22**(1), 19–21.

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

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. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior*, **23**, 734–759.

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.

Light, P., & Perret-Clerment, A. (1989). Social context effects in learning and testing. In A. Gellatly, D. Rogers, & J. A. Sloboda (Eds.), *Cognition and social worlds* (pp. 99–112). Oxford: Oxford Univ. Press.

Mandler, J. M. (1983). Representation. In J. H. Flavell, & E. M. Markman (Eds.), Mussen, P. H. (Series Ed.), *Handbook of child Psychology: Vol. 3, Cognitive development*. New York: Wiley.

Markman, E. M., Cox, B., & Machida, S. (1981). The standard object-sorting task as a measure of conceptual organization. *Developmental Psychology*, **17**(1), 115–117.

Markman, E. M., Horton, M. S., & McLanahan, A. G. (1980). Classes and collections: Principles of organization in the learning of hierarchical relations. *Cognition*, **8**, 227–241.

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

McCabe, A. E., Siegel, L. S., Spence, I., & Wilkinson, A. (1982). Class-inclusion reasoning: Patterns of performance from three to eight years. *Child Development*, **53**, 780–785.

McGarrigle, J., & Donaldson, M. (1975). Conservation accidents. *Cognition*, **3**, 341–350.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, **7**(4), 241–253.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, **19**, 242–279.

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.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**(3), 289–316.

Nelson, K. (1973). Some evidence for the cognitive primacy of function and its functional basis. *Merrill-Palmer Quarterly*, **19**, 21–39.

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.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, **6**, 449–468.

Piaget, J. (1952). *The child's conception of number.* London: Routledge & Kegan Paul.

Pick, A. D., & Heinrichs, M. (1989). *Perceptual learning in cognitive development.* Paper presented at the 5th International Conference on Event Perception, Oxford, OH.

Pillow, B. H., & Flavell, J. H. (1985). Intellectual realism: The role of children's interpretations of pictures and perceptual verbs. *Child Development*, **56**, 664–670.

Putnam, H. (1975). *Mind, language, and reality.* Cambridge: Cambridge Univ. Press.

Rabinowitz, M., & Mandler, J. M. (1983). Organization and information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**(3), 430–439.

Siegal, M. (1991). *Knowing children: Experiments in conversation and cognition.* Hillsdale, NJ: Erlbaum.

Siegel, S., & Castellan, Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences*, 2nd Ed. New York: McGraw–Hill.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, **28**, 249–257.

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, 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.

Smith, L. B., & Jones, S. S. (1993). Cognition without concepts. *Cognitive Development*, **8**, 181–188.

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

Trabasso, T., Isen, A. M., Dolecki, P., McLanahan, A. G., Riley, C. A., & Tucker, T. (1978). How do children solve class-inclusion problems? In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum.

Tversky, B. (1985). Development taxonomic organization of named and pictured categories. *Developmental Psychology*, **21**(6), 1111–1119.

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

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.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, **1**, 139–156.

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.

Waxman, S. R., Shipley, E. F., & Shepperson, B. (1991). Establishing new subcategories: The role of category labels and existing knowledge. *Child Development*, **62**, 127–138.

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.

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, MN: University of Minnesota Press.

RECEIVED: September 14, 1993; REVISED: September 16, 1994