

Running Head: SPEAKER INTENT IN CATEGORY INDUCTION

Don't Believe Everything You Hear:

Preschoolers' Sensitivity to Speaker Intent in Category Induction

Vikram K. Jaswal

University of Virginia

This is an electronic version of an Article published in *Child Development* © 2004.

Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development*, 75, 1871-1885.

Don't Believe Everything You Hear:

Preschoolers' Sensitivity to Speaker Intent in Category Induction

Abstract

A label can convey non-obvious information about category membership. Three studies show that preschoolers ($N = 144$) sometimes ignore or reject labels that conflict with appearance, particularly when they are uncertain that the speaker meant to use those labels. In Study 1, 4-year-olds were more reluctant than 3-year-olds to accept that, for example, a cat-like animal was a dog just on the basis of hearing it called a "dog." In Studies 2 and 3, this reluctance was overcome when the speaker explicitly or implicitly indicated that use of the unexpected labels was intentional. These studies demonstrate that preschoolers do not treat labels as atheoretical features of objects; rather, they interpret them in light of their understanding of the labeler's communicative intent.

Don't Believe Everything You Hear:

Preschoolers' Sensitivity to Speaker Intent in Category Induction

In every domain of human cognition that involves the acquisition of knowledge, we learn from others, often through language. This is especially evident when we acquire non-obvious knowledge—for example, knowledge of subtle distinctions or non-intuitive commonalities. Consider how children learn that penguins are birds. Penguins don't look much like typical birds, so their perceptual features are unlikely to lead children to categorize them with other birds. Most children never actually interact with penguins, so it is unlikely that they would discover, on their own, non-obvious similarities between penguins and other birds. And yet, between 12 and 24 months of age, children begin to recognize many atypical exemplars as members of familiar categories (Meints, Plunkett, & Harris, 1999). The most likely explanation for this shift is that older children are more likely to have heard a speaker refer to atypical exemplars with their basic-level category labels—that is, they are more likely to have heard a speaker explicitly refer to a penguin as a *bird* (see Adams & Bullock, 1986).

When children learn that a familiar label unexpectedly applies to an object, they gain much more than just a conventional term for that object. They also gain access to a rich knowledge-base from which new (and perhaps more accurate) inferences about that object can be drawn (Gelman, 2003; Markman, 1989). Knowing that penguins are birds, for example, children might infer that they have feathers and hatch from eggs—inferences they might not even have considered prior to hearing a penguin referred to as a *bird*. For a label to have this effect, however, children have to be willing to learn from another speaker in the first place. That is, they must be willing to accept the rather surprising news that a penguin is a bird simply on the basis

of hearing it called one. This paper reports three studies investigating how preschoolers' willingness to accept and use a speaker's unexpected label as the basis for inference depends on their interpretation of the speaker's communicative intent.

Gelman and her colleagues (e.g., Gelman, 1988; Gelman & Coley, 1990; Gelman & Markman, 1986, 1987) have conducted a number of studies demonstrating that preschoolers are often willing to use an unexpected label applied to an object to make an inference different from the one they would otherwise have made. In Gelman and Markman (1986), for example, 4-year-olds learned new, non-obvious properties for each of two familiar objects. They were then shown a third, target object, which looked like one of the training items, but was given the same category label as the other, and they were asked which property it possessed. For example, they were shown a picture of a squirrel and heard "This squirrel eats bugs." Next, they were shown a picture of a rabbit and heard "This rabbit eats grass." Finally, they were shown the target, a picture of an animal that looked very much like a rabbit (e.g., it had very long ears), and heard "See this squirrel? Does it eat bugs, like this squirrel, or does it eat grass, like this rabbit?"

The child's only task was to decide whether to extend the property on the basis of appearance or on the basis of the adult-provided category label. If children ignored the category label and responded on the basis of perceptual similarity, they would most often select the property that went with the perceptual match. In the current example, this would mean responding that the target, which looked like a rabbit but had been called a *squirrel* ate grass like other rabbits. In fact, however, Gelman and Markman (1986) found that preschoolers tended to make their inferences on the basis of the category label, responding that the rabbit-like squirrel would eat bugs like other squirrels. Comparable results have been obtained with children as young as 24 months using a similar procedure (Gelman & Coley, 1990; Jaswal & Markman,

2002), and with infants as young as 13 months using novel labels and categories (Graham, Kilbreath, & Welder, 2004).

These results are interesting because an object's appearance is normally a reliable predictor of its category membership (e.g., Jones & Smith, 1993; Medin & Ortony, 1989). Children as young as 3.5 months (e.g., Behl-Chadha, 1996; Eimas & Quinn, 1994), adults (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), and non-human animals (e.g., Bhatt, Wasserman, Reynolds, & Knauss, 1988; Freedman, Risenhuber, Poggio, & Miller, 2001) alike readily use perceptual similarity to determine the category to which something belongs. Children have extensive experience with perceptually based categories: In a typical 2-year-old's productive vocabulary, most category labels represent categories whose members cohere perceptually (Samuelson & Smith, 1999). In fact, Smith and her colleagues have proposed that children's sensitivity to regularities between the appearance of objects and their labels results in a shape bias, a generalization children make that categories are organized by shape (e.g., Smith, 1999; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002).

In Gelman and Markman (1986), however, when the appearance of an object and the label applied to it conflicted, children tended to make inferences on the basis of the label. Gelman (2003) has argued that children's willingness to make inferences on the basis of unexpected labels indicates that they recognize that what something looks like is only a good, not perfect, cue as to what it is (see also Bloom, 2000). On her view, children's categories are "theory based:" They are constructed not merely on the basis of perceptual characteristics and regularities, but on the basis of children's beliefs and assumptions about the world and the way language works. According to Gelman, children assume that a label provides direct access to an

object's kind, and that an object's kind determines what nonobvious properties it is likely to have.

This theory-based approach has been challenged by a number of researchers, who argue that it is not necessary to appeal to assumptions children make about categories or language in order to explain why they resolve appearance-label conflicts in favor of the label on category induction tasks (e.g., Jones & Smith, 1993; Mareschal, 2003; Sloutsky, 2003; Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999; Sloutsky, Lo, & Fisher, 2001). For example, Sloutsky and his colleagues have proposed that children make inferences on the basis of a computation of similarity: The more features two objects share, the greater their similarity, and the more likely an inference will be drawn between them. A label is weighted more heavily than perceptual features for hypothesized maturational reasons, but the label itself is treated explicitly as a feature of an object, much like a visual feature.

The similarity-based model seems to explain one important aspect of young children's category induction that the theory-based account has not considered. In Gelman and Markman's (1986) original study, about one-third of the preschoolers' responses were made on the basis of the appearance of the items rather than the labels the researcher used for them. On hearing a rabbit-like animal referred to as a *squirrel*, for example, children sometimes inferred that it ate grass (like other rabbits) rather than bugs (like other squirrels). Recall that according to the theory-based account, a label provides direct access to an object's kind. If this were true, then it is not clear why children sometimes would fail to use that label when making an inference.

The similarity-based model predicts that a shared label will lead to more label-based inferences for objects that are also perceptually similar. Indeed, Sloutsky and Fisher (2004, Experiment 4) had one group of preschoolers provide similarity judgments for the stimulus sets

used in Gelman and Markman (1986), and another group participate in an induction task using those stimulus sets. Results were in line with the prediction: Analysis of individual stimulus sets showed that the more perceptually similar two objects had been judged to be, the more likely children were to use a shared label to draw inferences between them.

The similarity-based account is appealing because it uses well known learning mechanisms to explain and predict children's behavior. All that is needed is a mechanism capable of computing similarity across two objects, and a shared label is just one attribute on which two objects might be similar (Sloutsky & Fisher, 2004). However, by treating a label merely as an attribute of an object akin to its perceptual attributes, it overlooks an important difference between labels and perceptual attributes. In particular, unlike perceptual attributes, labels are provided by people in order to refer to objects. As a result, how children interpret the intention of the person who provides a label is crucial to what effect (if any) that label will have.

A large body of work in language development has shown that children are extremely sensitive to pragmatic cues provided by a labeler (for reviews, see Baldwin & Tomasello, 1998; Tomasello, 2001, 2003). In many situations, children will learn a label for an object only when they have sufficient evidence that the speaker meant to refer to that object. For example, in Baldwin (1991), infants heard a speaker label an object in one of two conditions: In the follow-in condition, the speaker provided the label when both the speaker's and the infant's attention were focused on the same object. In the discrepant condition, the speaker provided the label when the infant was focused on one object, but the speaker was looking into a bucket containing a different object. Comprehension tests showed that 18-month-old infants mapped the label onto the visible object in the follow-in condition, but not in the discrepant condition. In the discrepant condition, they actually mapped the label onto the object the speaker was looking at, even though

it was concealed in the bucket. The fact that infants did not simply map the label onto whatever object happened to be available at the time they heard the label indicates that they were sensitive to cues about the speaker's referential intent (e.g., eye gaze, voice direction, body posture, etc.).

Older children are even more sophisticated users of information about a speaker's referential intent. In Akhtar and Tomasello (1996), for example, 2-year-olds played a game with an experimenter that involved removing objects from several containers. After several rounds of this game, the experimenter expressed an intention to "find the toma," and proceeded to attempt to open a container which was now locked. Even though these children never saw the object at the time the label was provided, comprehension tests showed that they had learned it as well as children in a condition where the label and object were presented together. Finally, Sabbagh and Baldwin (2001) showed that 3- and 4-year-olds were more likely to remember a new label if the speaker expressed confidence when labeling than if the speaker expressed uncertainty. In short, how children interpret a speaker's intent can affect how likely they are to learn or use the linguistic information she or he provides.

Returning to the induction studies described earlier, one reasonable pragmatic inference children might make on hearing a speaker use a label that conflicts with an object's appearance is that the speaker did not mean to use that label. Indeed, in this situation, a listener *should* be especially vigilant about confirming the speaker's intent. An appearance-label mismatch could easily be unintentional if the speaker misspoke or failed to see the object accurately, or if the listener misheard. Some degree of skepticism about unexpected information may be adaptive: It allows certain kinds of communicative errors to be filtered out before they have a chance to affect the child's own categorization scheme (see Baldwin & Moses, 2001).

If children are reluctant to accept and use an unexpected label to make an inference because they are not sure that the speaker meant to use it, then eliminating that doubt should make them more likely to accept and use it. For this prediction to be correct, children must have an understanding of the many ways that people can indicate an intention to refer—precisely the kind of “theory-based” knowledge that Gelman (2003) attributes to children, and that the similarity-based model eschews (e.g., Sloutsky, 2003). That is, the theory-based approach emphasizes the role that young children’s theories about objects and the world (including other people) play in children’s categories; the similarity-based approach argues that these are not necessary.

The three studies reported here investigate whether children are sensitive to cues about a speaker’s communicative intent in deciding whether to accept and use a speaker’s unexpected labels as the basis for inference. In all three studies, a speaker asked preschoolers to make inferences about objects that were sometimes referred to with unexpected labels. Study 1 was designed to provide baseline data for the later studies, and to document what children say spontaneously when they hear an unexpected label. In Study 2, the speaker explicitly indicated that use of each unexpected label was intentional by saying, “You’re not going to believe this, but this is actually an X.” Finally, in Study 3, the speaker provided a more subtle cue that he meant to use the unexpected labels by modifying each one pre-nominally.

Study 1

In Study 1, preschoolers saw color pictures of objects from two familiar categories, heard them labeled, and watched the researcher demonstrate a different activity with each (e.g., a dog ate bones, and a cat drank from a bowl of milk). After this training event, they saw pictures of hybrid objects, designed to look more like one object than the other. Sometimes these hybrids

were referred to twice with an unexpected category label (e.g., a dog-like animal was referred to as a *cat*), and sometimes they were referred to neutrally (e.g., as “this one”). Children were asked to use the pictures to act out the activity in which each engaged.

We expected that, as in previous induction studies, a shared label would often, but not always, influence children's inferences (e.g., Gelman & Markman, 1986, 1987; Sloutsky et al., 2001; Sloutsky & Fisher, 2004). On its own, Study 1 does not address whether children fail to accept unexpected labels because they do not think the speaker intended to use those labels. However, it provides baseline data for the pragmatic manipulations in Studies 2 and 3. Additionally, it allows us to document what children say when a speaker provides an unexpected label, which may provide insight into how children interpret that label. Previous induction studies have not reported on children's spontaneous utterances (e.g., Sloutsky et al., 2001), or have done so only anecdotally (e.g., Gelman & Markman, 1986; Gelman & Coley, 1991).

Method

Participants. Twenty-four 3-year-olds (mean age = 3 years, 8 months; range = 3;2 to 3;11) and twenty-four 4-year-olds (mean age = 4 years, 5 months; range = 4;0 to 5;0) participated in a single 10- to 15-minute session at their preschool or in the lab. Approximately equal numbers of boys and girls participated in each age group. Children in this and subsequent studies were recruited from preschools or from a database of interested families, and most were White and from middle-class or upper middle-class backgrounds.

Design. Each child participated in eight trial blocks, four involving sets of animal stimuli, and four involving sets of artifact stimuli. The labeling manipulation was also within-subject: the speaker provided labels for two animal and two artifact sets, and not for the others.

Stimuli. Eight animals from familiar categories were grouped into four sets based roughly on shape similarity, and eight artifacts from familiar categories were grouped into four sets in a similar manner. The result was the following eight stimulus sets: cat-dog, horse-cow, bear-pig, fish-bird, key-spoon, shoe-car, cup-hat, and button-ball. Realistic color illustrations of a prototypical exemplar of each animal category were obtained from commercially available picturebooks, and were digitized for computer manipulation. Color photographs of a prototypical exemplar of each artifact category were obtained from a digital library of photo-objects (Hemera Technologies). These images will be referred to as the “standard objects.”

One additional typical exemplar of each category was created (“typical test objects”), primarily by manipulating the color of each standard. Additionally, two hybrid objects (“hybrid test objects”) were created by using one of the standards (or an additional typical image) as a base, and adding features of the other standard image of that pair. For each set, one hybrid was designed to look more like a member of one category; the other hybrid was designed to look more like a member of the other. For example, in the cat-dog set, the hybrids had perceptual features of both cats and dogs, but one was designed to look more like a cat and the other was designed to look more like a dog. Baseline trials (described below) confirmed that the hybrid objects looked like members of the intended categories, but were less good exemplars than the typical test objects. The complete cat-dog continuum is shown as an example in Figure 1. A set of warm-up stimuli consisted of typical exemplars of dolls and shovels.

Insert Figure 1 about here

All images were sized to approximately 2 to 4 inches wide, and 2 to 5 inches tall. Each image (and its left-right reverse) was printed in color and mounted into a small stand that allowed it to remain upright.

Each category was associated with a particular activity or function, as shown in Table 1. Color photographs of the props associated with each activity were mounted onto two easels. For example, a photo of a bowl of milk was mounted onto one easel, and was used to demonstrate the cat drinking milk; a photo of some dog bones was mounted onto the other easel, and was used to demonstrate the dog eating bones. It was expected that the activities and functions might be familiar to children (though this was not necessary).

Insert Table 1 about here

Procedure

Children were tested individually in a small room at their school or in the laboratory. They sat at a small table, with the researcher across from them. Each session began with a warm-up trial that served to familiarize them with the task: On one easel, the experimenter displayed a photo of a bed and explained that a doll slept in the bed. As the explanation was being given, the experimenter demonstrated putting the doll to sleep in the bed. On the other easel (displayed simultaneously), the experimenter showed a photo of a bucket and demonstrated and explained aloud that a shovel could be used to scoop sand into or out of the bucket. Children were then shown additional dolls and shovels in alternating order and were asked where each went, until they succeeded in putting a doll with the bed and a shovel with the bucket consecutively. Correct

selections were praised, and incorrect selections were corrected. Fewer than one-third of the children in this or any of the following studies required correction on the warm-up trial.

Test trials were similar to the warm-up trial: For example, the experimenter showed a photo of a bowl of milk, and explained that a cat drank the milk, and then showed a photo of dog bones, and explained that a dog ate bones. In order to reduce memory load, following the demonstration, each standard exemplar remained standing next to its appropriate background photograph. Children were then shown three test objects, one at a time and in a pseudo-random order (described below), and they were asked to show the activity or function associated with each: Two of the test objects were the additional typical exemplars from that set, and the third was one of the two hybrids from that set. One hybrid per set was shown to half of the children, and the other hybrid from that set was shown to the other half (e.g., half of the children saw the cat-like hybrid animal and half saw the dog-like one).

Children were presented with eight such trial blocks, corresponding to the eight pairs of objects in Table 1. On the four *label* trial blocks, the researcher used a category label to introduce each of the test objects. Labels were provided twice: “This is a/Look at this X. Can you show me what this X does?” [Due to experimenter oversight, some children heard the label introduced with the phrase “This is an X,” and others heard it introduced with the phrase “Look at this X.” Analyses showed no difference between the two carrier phrases, and so data are reported together.] The typical test exemplars were always called by labels that matched their appearance. By contrast, the hybrids were referred to by labels that were the *opposite* of their appearance (i.e., the cat-like hybrid was referred to as “this dog;” see Figure 1). On the four *no-label* trial blocks, the researcher used the phrase “this one” to introduce each test object, including the hybrid (for example, “Look at this! Look at this one! Can you show me what this

one does?"). Regardless of their selection, children were given neutral feedback in a positive tone ("Okay!"), and the researcher then proceeded to the next test object or trial block.

The label and no-label trial blocks were made up of two animal sets and two artifact sets each. Children were yoked into pairs so that the stimulus sets making up one child's label trial blocks corresponded to another child's no-label trial blocks and vice versa. For each yoked pair, the presentation order of the two typical objects and the hybrid in each set was random with the constraint that the hybrid object was presented first, second, and third at least once each in the label trial blocks and at least once each in the no-label trial blocks. Four random orders of trial blocks were used, with the constraint that no more than two label or no-label trial blocks could occur successively. Additionally, the left-right positions of the background photographs matching the appearance of the hybrid stimuli were counter-balanced within and across children.

Coding in this and subsequent studies was conducted off-line, via videotape, and involved noting which of the two possible activities or functions was selected for each test object. For reliability, a second coder coded a randomly selected 25% of all the sessions from this and subsequent studies with the sound off. Agreement between the two coders on this and subsequent studies was at least 98%.

Results and Discussion

Preliminary analyses revealed no effects or interactions involving gender or whether the stimulus sets involved animals or artifacts; subsequent analyses collapsed across these factors. Inferences about the typical items served as a check to insure that children understood the task: When dealing with familiar, prototypical items either labeled appropriately or not labeled at all, children should perform at near-ceiling levels. Indeed, 3- and 4-year-olds made appropriate perceptually based inferences for the typical items 99% and 100% of the time, respectively. In

other words, when a typical cat was shown and referred to as “this cat” or “this one,” preschoolers at both ages readily inferred that it would drink milk rather than eat bones.

The remainder of the analyses focus on responses to the hybrid stimuli, which were the stimuli that had features of two categories, but were designed to look more like one category than the other. Table 2 shows the percentage of perceptually based inferences children made as a function of whether the hybrids were referred to neutrally or with a counter-intuitive label. As the table shows, children made more perceptually based inferences about the unlabeled hybrids than the labeled ones. On no-label trials, 3- and 4-year-olds made perceptually based inferences 88% and 91% of the time, respectively, more frequently than would be expected by chance of 50% [t 's > 12.46, p 's < .001]. In contrast, on the label trials, 3-year-olds made perceptual inferences just 33% of the time, less often than would be expected by chance, $t(23) = 2.64$, $p < .05$, and 4-year-olds made perceptually based inferences 59% of the time, which is not different from chance, $t(23) = 1.46$, $p > .10$.

Two conclusions can be drawn at this point: First, the hybrid objects were perceptually compelling in the manner intended: When they did not hear it labeled, both 3- and 4-year-olds treated the cat-like object as a cat, for example. Second, hearing an object referred to with an unexpected label did influence the inferences that children made, as in previous studies (e.g., Gelman & Markman, 1986). However, it seemed to have more of an influence on the 3-year-olds than the 4-year-olds.

 Insert Table 2 about here

A 2-way mixed ANOVA on the responses to the hybrid stimuli (label/no-label x 3's/4's) showed that both main effects were significant: Children at both ages made fewer perceptually-based inferences when they heard the hybrids labeled than when they did not, $F(1,46) = 95.00$, $p < .001$, and 3-year-olds made fewer perceptually-based inferences than 4-year-olds, $F(1,46) = 6.88$, $p < .05$. A significant interaction confirmed that, compared to 4-year-olds, 3-year-olds made fewer perceptually based inferences about the hybrids that were called by counter-intuitive labels, $F(1,46) = 6.84$, $p < .05$. That is, 3-year-olds who heard the cat-like hybrid referred to as “this dog” were more likely than 4-year-olds to have it eat bones like the other dog.

This age difference was also evident in the individual data: On the four label trials, only 10 of the 24 three-year-olds (42%) made two or more perceptually based inferences despite hearing the labels, whereas 19 of the 24 four-year-olds (79%) did so, distributions that are significantly different from each other, $\chi^2(1, N = 48) = 7.06$, $p < .01$. In contrast, on the four no-label trials, all children at both age groups made two or more perceptually based inferences.

To our knowledge, this is the first time that younger children have been found to be more influenced by unexpected labels than older ones on a category induction task. In fact, these results conflict with predictions and findings from the similarity-based model (Sloutsky et al., 2001). According to that model, there is more variability in what members of the same category look like than what they are called. As children grow older, they gain experience with different-looking objects being referred to with the same label, and so learn that a label can be a more reliable predictor of category membership than appearance. In one study supporting this developmental progression, Sloutsky et al. found that when 4- to 5-year-olds heard two objects called by the same label, they were influenced both by the shared label and by the number of features the objects had in common. In contrast, 11- to 12-year-olds (and adults) *always* drew an

inference between objects sharing the same label, regardless of their perceptual characteristics. That is, older children were more influenced by the labels than younger ones—exactly the opposite of our finding.

Sloutsky et al. (2001) used unfamiliar (rather than familiar) labels paired with familiar and unfamiliar objects, so the two studies may not be directly comparable. Additionally, Sloutsky et al. used a larger spread of age groups. Perhaps if they had analyzed the preschool data by year, they would have found 4-year-olds to be more influenced than 5-year-olds. Or, perhaps if we had run an elementary school age group, we would have found that children this age always relied on the shared labels. Whatever the reason for the difference between studies, it is clear that 3-year-olds in our study were more open to using labels as a basis of inference than 4-year-olds. Why?

Naigles, Gleitman, and Gleitman (1983) found an analogous age difference in how children interpret anomalous information about the subcategorization of familiar verbs. In their study, younger preschoolers who heard a sentence with a familiar verb used in an inappropriate frame were more likely to treat that sentence as if it were grammatical than older preschoolers. For example, when asked to use a small Noah's Ark set of animals to act out an ungrammatical sentence like "The zebra brings," 2-year-olds tended to move the zebra on its own toward a location—as if *brings* could be used intransitively (e.g., as in *comes*). Three- and 4-year-olds, on the other hand, tended to have the zebra bring another animal to a location—as if *brings* could only be interpreted transitively, which necessitated a causal action. Naigles et al. argue that the younger subjects were more open to the possibility that they had not yet heard every possible frame in which a familiar verb could appear. For the older preschoolers, however, an anomalous usage on one occasion was more likely to be treated as an error. Indeed, some older preschoolers

actually misheard the anomalous sentences so that a sentence like “The lion *comes* the giraffe” resulted in an acting out of “The lion *combs* the giraffe.”

The analogy with the present study is that the 3-year-olds might have assumed that they had not seen every breed of, for example, cat or dog, and so they were more willing than 4-year-olds to accept on faith that a cat-like animal could actually be a dog. Because 4-year-olds may have had more experience with the various categories represented by the stimuli used in our study, they may have been more conservative. Unfortunately, we do not have any information about the amount of experience children at the different ages had with the various categories we used, so this possible explanation will require further study.

Children's spontaneous reactions to the unexpected labeling events are useful to consider for two reasons. First, if children spontaneously comment on the label that the speaker uses, then this provides evidence that the children heard and processed the label. Second, the content of such comments may provide insight into why children sometimes made a perceptually based inference about a hybrid despite hearing it labeled. Working from transcripts, two coders independently coded into one of three categories the spontaneous comments that immediately followed hearing a hybrid stimulus labeled: explicit denials of the speaker's label (including rejections and corrections); other comments about the speaker's label and/or the appearance of the item (including “looks like” statements, questions, expressions of confusion, etc.); and irrelevant comments. Table 3 shows examples of each. The two coders agreed on 94% of the 104 comments (Cohen's kappa = .88); the few disagreements were resolved through discussion.

 Insert Table 3 about here

Figure 2 provides a detailed breakdown of the percentage of each type of comment children made (if any) when they heard the hybrids labeled, as a function of whether they subsequently used the label to make an inference, or whether they instead chose to make a perceptually based inference. What is striking about this figure is that when children made a perceptually based inference despite hearing a label, they almost always made a relevant comment—3-year-olds did so on 88% of such trials, and 4-year-olds did so on 86% of them. Moreover, as the hatched areas of the figure show, these comments most frequently involved an explicit denial of the label that the speaker used. For example, on hearing a spoon-like object called a “key,” one child (age 4;1) replied “That’s not a key; (it’s) a spoon,” and used it to eat cereal from a bowl rather than start a car.

Insert Figure 2 about here

The frequency of these comments rules out the possibility that children failed to make label-based inferences because they did not hear or process the unexpected labels. Rather, it was most often because they did not *accept* that the label applied to the object being labeled in the first place. This is a different explanation for why children fail to make label-based inferences than the one provided by the similarity-based model of induction (Sloutsky et al., 2001; Sloutsky & Fisher, 2004). On that account, children do not actively try to interpret the information provided by another speaker. Rather, labels simply enter into a computation of similarity whose output decides whether a label-based or perceptually based inference is warranted. The present results suggest that a label may not even enter into the computation of similarity.

In summary, when preschoolers—particularly 4-year-olds—were faced with a label-appearance disparity, they did not always make a label-based inference (see also Gelman & Markman, 1986; Sloutsky et al., 2001; Sloutsky & Fisher, 2004). Their spontaneous utterances suggested that they often refused to accept that the label applied to the object being labeled in the first place. As suggested earlier, a listener *should* be skeptical about unexpected information, as it may have been provided unintentionally (e.g., the result of a speech production error, or a failure to see the object appropriately). If the speaker makes it clear that use of those labels was intentional, then children should be more likely to accept and use them as the basis for inference. Study 2 addressed this possibility.

Study 2

The goal of Study 2 was to investigate whether an explicit pragmatic cue indicating that the speaker meant to provide the unexpected labels would make 3- and 4-year-olds more willing to accept them. The procedure was exactly the same as that used in Study 1; however, when introducing each of the unexpected labels, the speaker said, “You’re not going to believe this, but this is actually an X.”

The similarity-based model of category induction (Sloutsky, 2003) does not consider the intention of the speaker in calculating similarity; labels are merely features of objects. Thus, manipulating cues to speaker intent would not be expected to have an effect on whether children accept and use the labels the speaker provides. The theory-based account (Gelman, 2003), in contrast, holds that how children interpret the intentions of another person is crucial to how they interpret that person’s behavior (e.g., Bloom & Markson, 1998; Diesendruck et al., 2004; Gelman & Ebeling, 1998; Gergeley, Bekkering, Kiraly, 2002). On the theory-based account but

not the similarity-based one, then, the pragmatic manipulation here should cause children to make more inferences on the basis of the speaker's labels.

Method

Participants. Twenty-four 3-year-olds (mean age = 3 years, 4 months; range = 3;0 to 3;11) and twenty-four 4-year-olds (mean age = 4 years, 7 months; range = 4;0 to 5;6) participated in their preschools or the lab. None had participated in Study 1. Within each age, there were approximately equal numbers of boys and girls. [Due to the limited availability of participants, four children who had already turned 5 years old were included in the group of 4-year-olds. Results and analyses are the same regardless of whether they are included.]

Procedure

The procedure was exactly the same as that used in Study 1, but the researcher introduced each of the four labeled hybrid items per child by saying, "You're not going to believe this, but this is actually a [label]. Can you show me what this [label] does?" The labels were provided twice, as in Study 1.

Results and Discussion

As in the previous study, there were no effects or interactions involving gender or ontological category, and so subsequent analyses collapsed across these factors. Children at both ages made perceptually based inferences for the typical items at ceiling levels (between 98% and 99% of the time), regardless of whether the stimuli were labeled or not. As before, the children clearly understood the requirements of the task.

The middle section of Table 2 shows the results from the hybrid trials as a function of the percentage of perceptually-based inferences children made. What is immediately obvious is that, unlike Study 1, where 3-year-olds were more influenced by the unexpected labels than 4-year-

olds, the performance of the two age groups in this study was nearly identical: Children at both ages were influenced considerably by the unexpected labels. Both 3- and 4-year-olds made perceptually based inferences about the hybrid stimuli when the stimuli were referred to neutrally, and they made the opposite (label-based) inferences about them when they were referred to with counter-intuitive labels—in all cases, at levels significantly different from what would be expected by chance [t 's > 6.33, p 's < .001].

A 2-way mixed ANOVA on these data (3's/4's x label/no-label) yielded a significant effect of whether the items were labeled, $F(1,46) = 237.22$, $p < .001$, but no effect of age and no interaction. In terms of the number of children who made at least two perceptually based inferences despite hearing the items labeled, only 4 of the 24 (17%) children at each age did so.

As in Study 1, two independent coders categorized children's spontaneous comments about labeled hybrid stimuli into one of three categories: explicit denials, other relevant comments, and irrelevant comments. Coders agreed on 98% of the 89 utterances (Cohen's kappa = .95); the single disagreement was resolved through discussion. The middle panel of Figure 2 shows the details of the data. As the figure shows, on the few occasions when they made a perceptually based inference despite hearing the label, children at both ages still tended to explicitly deny the label or to make some other relevant comment about the label or the appearance of the hybrid. Thus, as in Study 1, when children failed to make label-based inferences, it was most often because they rejected the appropriateness of the speaker's label—even though the speaker had alerted them to the fact that it was going to seem inappropriate.

Before discussing the implications of these results for the similarity- and theory-based accounts of category induction, we present a third study. Whereas Study 2 asked whether

children could use an explicit cue that the speaker meant to use the unexpected labels, Study 3 considered whether children might also be able to make use of a more subtle pragmatic cue.

Study 3

When a speaker provides extra information about an object, it is usually a means of providing the listener with additional content. For example, pre-nominal modification can convey subtype information: “Siamese cat” indicates a kind of cat, and “oak tree” indicates a kind of tree. Children as young as 2.5 years seem to understand that the head noun indicates the category, and the modifier specifies the sub-category (Clark, Gelman, & Lane, 1985; Gelman, Wilcox, & Clark, 1989; Hall & Moore, 1997). If the speaker modified an unexpected label to suggest that it referred to an unfamiliar subtype (e.g., “Pashtun dog”), children might assume that they had never encountered that kind of thing, and so might be more willing to use the category label to make an inference.

Interestingly, the very modification of a label increases the probability that the speaker meant to use that label in the first place. For example, the likelihood that the speaker meant to use the label “dog” to refer to a cat-like animal may be higher when the label is introduced as a “Pashtun dog” than merely a “dog.” This pragmatic argument follows from one of Grice’s (1975) proposed maxims underlying successful conversational exchanges. According to the maxim of quantity, speakers should make their contributions neither less nor more detailed than required. When a speaker purposely provides extra information, Grice suggests that it could be a means of indicating that the speaker recognizes that there may be controversy about what she or he is saying. Thus, modification of an unexpected label may make children more willing to use that label as the basis for inference for at least two reasons: It may signal to them that the referent

is an unfamiliar subtype, and/or it may signal to them that the speaker intended to use the unexpected label.

To investigate these possibilities, children in Study 3 participated in one of two conditions. In the *subtype* condition, the unexpected labels were introduced with unfamiliar modifiers. For example, the cat-like hybrid was sometimes introduced as a “Pashtun dog.” “Pashtun” does not in and of itself explain why the dog looks unusual. But, as mentioned earlier, given children’s understanding of compound nouns (e.g., Clark et al., 1985), children may assume that it represents an unfamiliar kind of dog.

In the *familiar adjective* condition, the unexpected labels were introduced with known, non-explanatory pre-nominal adjectives. For example, the cat-like animal was sometimes introduced as a “furry dog.” “Furry” has a meaning, known to preschoolers, and cannot explain why the cat-like animal is being referred to as a “dog.” It may indicate a kind of dog (e.g., furry as opposed to bald), but, unlike “Pashtun dog,” “furry dog” does not imply that the dog is a kind of dog children have never seen before. If the provision of additional information alone (rather than the content of that information) can indicate that a label was produced intentionally, then children in the familiar adjective condition should make as many label-based inferences as those in the subtype condition.

Because the interest in Study 3 was whether subtle modifications of the unexpected labels would enhance their influence, we chose to focus only on 4-year-olds. Recall from Study 1 that unmodified labels were more likely to influence the responses of 3-year-olds than 4-year-olds; thus, 4-year-olds had more to gain from the manipulations planned here.

Method

Participants. Forty-eight 4-year-olds participated in their preschools or in the lab. Twenty-four children participated in a *subtype* condition (mean age = 4 years, 7 months; range = 4;2 to 5;0), and 24 in a *familiar adjective* condition (mean age = 4 years, 4 months; range = 4;0 to 5;0). None had participated in Studies 1 or 2. Within each condition, there were equal numbers of boys and girls. In the subtype condition, one girl's data were collected over two sessions separated by a week, and another girl's responses were coded during her session because she did not have permission to be videotaped. In the familiar adjective condition, two additional children participated, but their data were excluded because they became bored during the first half of the study and asked to return to their classroom.

Procedure. The procedure was exactly the same as that used in Studies 1 and 2, but the hybrid items were always introduced with a pre-nominal modifier. In the *subtype* condition, modifiers were blank—they had no known meaning for preschoolers. The modifiers were randomly selected (without replacement) from the following list: Tamil, Moroccan, Lebanese, Celtic, Flemish, Kashmiri, Pashtun, and Berkshire. Real words were used in order to insure that they sounded realistic, but they were assumed to be novel to 4-year-olds. Additionally, these modifiers were selected because their endings varied, so children would be unable to use one ending in particular as a cue to subtype. All hybrid items were introduced with a compound phrase, regardless of whether they included a label (e.g., “this Pashtun dog”) or not (e.g., “this Tamil one”).

In the *familiar adjective* condition, modifiers were known and described perceptually obvious features of the objects. The eight stimulus sets were assigned the following adjectives: *furry* dog/cat, *big* horse/cow, *fat* bear/pig, *brown* fish/bird, *shiny* key/spoon, *red* shoe/car, *smooth*

cup/hat, and *round* button/ball. As in the subtype condition, the hybrid items were introduced with a modifier, regardless of whether they were labeled (e.g., “this furry dog”) or not (e.g., “this fat one”).

Results and Discussion

As in the previous studies, preliminary analyses did not reveal any main effects or interactions involving gender or ontological category, and so subsequent analyses collapsed across these factors. Also like the previous studies, children made perceptually based inferences about typical items at near ceiling levels (between 98% and 99% of the time), regardless of whether the stimuli were labeled.

The bottom section of Table 2 shows the results from the hybrid trials. As the table shows, children in both conditions were equally affected by the anomalous labels. Indeed, to preview the results from the following analyses, performance was statistically the same in the subtype and the familiar adjective conditions. In both conditions, children made primarily perceptually based inferences about the unlabeled hybrid stimuli, more frequently than would be expected by chance (t 's > 7.88, p 's < .001). By contrast, on hearing the hybrid stimuli labeled with perceptually counter-intuitive labels, they made fewer perceptually based inferences than expected by chance (t 's > 2.48, p 's < .05).

A 2-way mixed ANOVA on these data (subtype/familiar adjective x label/no-label) showed a main effect of labeling, indicating that children in both conditions were more likely to make perceptually based inferences about the unlabeled hybrids than the labeled ones, $F(1,46) = 134.89$, $p < .001$. There was no effect of condition and no interaction. Analysis by individual confirmed that both types of pre-nominal modification resulted in the same pattern of responses: In the subtype condition, 8 of the 24 participants (33%) made two or more perceptually based

inferences despite hearing the items labeled, and in the familiar adjective condition, 9 of the 24 (38%) did so.

The spontaneous comments children made on hearing the stimuli labeled were categorized by the same two coders as in the previous studies into one of three categories: Explicit denials of the speaker's label; other comments relevant to the appearance of the stimulus and/or its label; and irrelevant comments. As before, reliability was high, with the two coders agreeing on 95% of the 80 comments (Cohen's kappa = .91); disagreements were resolved through discussion. The right panel of Figure 2 shows the detailed distribution of children's comments as a function of whether they made label- or perceptually based inferences. As the figure shows and as in the previous studies, when children made a perceptually based inference despite hearing the hybrid labeled, they frequently commented in some relevant way on the disparity between the label and the object's appearance.

That children in the familiar adjective condition responded in the same way as those in the subtype condition is intriguing. Why did hearing a ball-like object referred to as a "round button" have the same effect as hearing it referred to as a "Flemish button?" After all, whereas "Flemish button" at least holds out the promise that an explanation exists (e.g., it's a kind of button I have never seen), "round button" does not. The proposal here is that the content of the modification was less important than the fact that the category label was being modified in the first place. When the speaker provided additional information, the provision of that information alone made it more obvious that the speaker intended to use that particular label (Grice, 1975).

Comparisons of 4-year-old Results

Given that all three studies used the same task and stimuli, we compared how likely 4-year-olds were to use unexpected labels as the basis for inference as a function of how the labels

were introduced: on their own (Study 1), preceded by the “actually” phrase (Study 2), or modified pre-nominally (Study 3). These data were analyzed (and results are reported) as if children participated in a single experiment in one of four conditions: bare label, actually, subtype, or familiar adjective.

Four-year-olds made perceptually-based inferences about the labeled hybrid stimuli most often in the bare condition (59% of the time), less in the subtype and familiar adjective conditions (24% and 32%, respectively), and least often in the actually condition (13%). A one-way ANOVA on these data yielded a significant effect of condition, $F(3,92) = 10.37, p < .001$. Post-hoc Tukey analyses revealed that children who heard a cat-like animal referred to as a “dog,” for example, were more likely to ignore or reject that label than those who heard the label preceded by the “actually” phrase, or than those who heard it referred to as a “Tamil dog” or a “furry dog.” The actually, subtype and familiar adjective conditions did not differ from each other.

Two important conclusions can be drawn from this analysis. First, as the individual study analyses suggested, children were more influenced by unexpected labels when they were preceded by an explicit statement indicating that the speaker meant to use them (e.g., “You’re not going to believe this, but . . .”) or by a blank or familiar modifier (e.g., “Tamil” or “furry”) than when they were simply presented on their own. Furthermore, pre-nominal modification was as effective in influencing children’s responses as was explicit pragmatic information that the speaker intended to use that label.

General Discussion

These three studies investigated whether preschoolers’ willingness to accept a speaker’s unexpected label and to use it as a basis for inference depended on their interpretation of the

speaker's communicative intent. Study 1 provided baseline data by showing that preschoolers—particularly 4-year-olds—did not always draw inferences on the basis of a speaker's unexpected labels. Study 2 showed that when the speaker made it clear that use of the unexpected labels was intentional, children were much more likely to use them. And Study 3 showed that simple modification of the unexpected labels also made children more likely to use them. Together, these studies serve as an important reminder that labels derive their power not from their mere presence, but from their intentional use (see also Baldwin & Tomasello, 1998; Tomasello, 2001).

Before continuing, an alternative explanation of these results needs to be addressed: Perhaps the addition of the extra verbal information in Studies 2 and 3 simply increased children's attention to the labels (e.g., Samuelson & Smith, 1998). On this account, this increased attention (rather than information about the speaker's intention) led them to weight the labels more heavily in these studies than in Study 1. Children's spontaneous comments in Study 1, however, make this attentional account unlikely. As discussed earlier, and as shown in Figure 2, when children failed to make a label-based inference about the labeled hybrids, they almost always made a comment indicating that they had heard and processed the unexpected label. In fact, 75% of 3-year-olds' and 68% of 4-year-olds' perceptually based responses in Study 1 were accompanied by an explicit denial of the label the speaker used (e.g., "It's not a hat; it's a cup"). This makes it clear that children in Study 1 were paying attention to the speaker, heard the label, processed it, and frequently decided it was inappropriate.

These findings cannot be explained by the similarity-based model of category induction (Sloutsky, 2003). According to this model, labels are merely attributes of objects: Just as it does not matter in a computation of similarity whether the physical features of a particular object (e.g.,

color, texture) were produced intentionally, it should not matter whether the label was produced intentionally. Studies 2 and 3, however, demonstrated that children were more likely to accept and use an unexpected label as the basis for inference when they interpreted it as having been used intentionally.

The theory-based approach (Gelman, 2003) is better suited to account for these data. On this view, children's categories reflect more than just regularities between cues in the environment; they also reflect children's beliefs and assumptions about the world. How children decide which cues to pay attention to depends on what they already know and believe about the relevance of any given cue in a particular situation. In the three studies reported here, the same label (e.g., "dog") had a different effect on children's inferences, depending on whether the children interpreted that label as having been produced intentionally.

A recent study by Koenig and Echols (2003) provides additional evidence that children respond differently to the same labels, depending on whether they interpret them as having been produced intentionally. They asked how 16-month-old infants would respond to true versus false labeling events. For example, infants were shown a picture of a ball and heard either "That's a ball" (true) or "That's a duck" (false). In some studies, a human seated next to the infant produced the utterances, and in other studies, an audio speaker did. Koenig and Echols found that infants were more likely to spontaneously respond to a false label with the veridical one—as if correcting the speaker—if the false label had been produced by a human than the audio speaker. (Control studies ruled out the possibility that it was the presence of the human alone that was responsible for this difference.) If a label were merely an attribute of an object, it should not have mattered what the source of that label was.

A number of induction studies also suggest that a label's influence depends on children's theories. For example, in Study 1 of Gelman and Markman (1986), preschoolers tended to use the label as the basis for inference when the inference concerned a kind-relevant non-obvious property (e.g., what an animal was likely to feed its young). In contrast, in their Study 2, when the inference concerned which of two colored dots should be applied to an object, most children did not consistently make label-based inferences. Likewise, in their Study 3, most children did not consistently use a label to draw an inference about, for example, whether a bird's legs get cold at night. These results suggest that children base their decision about whether to make a label-based inference at least in part on their understanding of whether kind membership is relevant to that inference.

Similarly, McCarrell and Callanan (1995) showed that preschoolers recognize that an object's appearance is sometimes a better basis for inference than its label. In their study, children were shown a novel animal with an accordion-like neck. The experimenter called it a *wug* and explained that it could "stretch tall to eat cherries." When shown another animal that had an accordion-like neck, children were quite likely to agree that it "stretched tall to eat cherries"—even when it was called by a different label. Even though children had presumably never seen an animal with an accordion-like neck and it certainly had never been associated with stretching tall to eat cherries, they were able to create, on the fly, a causal theory connecting the two: "It can stretch tall to eat cherries because it has an accordion-like neck." As Markman and Jaswal (2003) pointed out, children may have drawn upon theories of physics (an accordion-like neck would be able to stretch), psychology (it wants to reach the cherries, so it will try to), and biology (it is an animate object and so therefore eats). When confronted with another animal with

an accordion-like neck and asked whether it could stretch tall to eat cherries, they recognized that its label (and therefore kind membership) was irrelevant.

Similarity between the target and test items is clearly an important factor in children's induction (e.g., Sloutsky & Fisher, 2004; Sloutsky et al., 2001), but children's theories also play a prominent role. Their theories enable them to judge whether a speaker intended to use a particular label, and whether that label (as an indication of kind membership) is relevant to the inference at hand. Perhaps the similarity-based model could add additional terms to represent this kind of information. However, as shown in the three studies reported here, cues to intentionality can be diverse, including explicit cues (like saying "You're not going to believe this, but . . .") and implicit ones (like pre-nominal modification). Accounting for the range of pragmatic cues may require exactly the kind of theory-based knowledge that the similarity-based model was explicitly designed to avoid.

It should be noted that researchers who challenge the theory-based explanation of category induction do not deny that children may have conceptual knowledge. Indeed, Sloutsky and Fisher (2004) note that the similarity-based model requires that children be able to distinguish between ontological categories, but they have not yet provided an atheoretical mechanism for this. Likewise, Mareschal's (2003) explanation of category induction focuses on perceptual similarity and variability, but he acknowledges that "there is clear evidence of [verbal, theory-based categories] in both the developmental and adult literature" (p. 380). However, both Sloutsky and Mareschal argue that atheoretical mechanisms may be sufficient to account for young children's behavior on category induction tasks.

The studies reported here suggest otherwise: Preschoolers' willingness to accept and use a speaker's unexpected label as the basis for inference depended on whether they thought the

label had been used intentionally. What purpose would this serve? Most speakers, most of the time, do their best to provide accurate information when they communicate with children. But, for a variety of reasons (e.g., lack of knowledge, speech production error, lack of perceptual access), speakers may not always be accurate. Filtering out information that conflicts with one's own hypotheses and that does not seem intentionally conveyed is an adaptive, theory-based strategy for maintaining veridical knowledge about the world.

References

- Adams, A. K., & Bullock, D. (1986). Apprenticeship in word use: Social convergence processes in learning categorically related nouns. In S. A. Kuczaj & M. D. Barrett (Eds.), *The development of meaning* (pp. 155-197). New York: Springer-Verlag.
- Akhtar, N., & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology, 14*, 79-93.
- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875-890.
- Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309-329.
- Baldwin, D. A., & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In E. Clark (Ed.), *The Proceedings of the Twenty-ninth Annual Child Language Research Forum* (pp. 3-23). Chicago: Center for the Study of Language and Information.
- Behl-Chada, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition, 60*, 105-141.
- Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial kind stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 219-234.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science, 9*, 200-204.

- Clark, E. V., Gelman, S. A., & Lane, N. M. (1985). Compound nouns and category structure in young children. *Child Development, 56*, 84-94.
- Diesendruck, G., Markson, L., Akhtar, N., Reudor, A. (2004). Two-year-olds' sensitivity to speaker intent: An alternative account of Samuelson and Smith. *Developmental Science, 7*, 33-41.
- Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development, 65*, 903-917.
- Freedman, D. J., Risenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*, 312-316.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford: Oxford University Press.
- Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition, 66*, B35-B47.
- Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796-804.
- Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought* (pp. 146-196). Cambridge, England: Cambridge University Press.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183-208.

- 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., Wilcox, S. A., & Clark, E. V. (1989). Conceptual and lexical hierarchies in young children. *Cognitive Development, 4*, 309-326.
- Gergeley, G., Bekkering, H., Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755.
- Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development, 75*, 409-427.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Volume 3: Speech acts* (pp. 41-58). New York: Academic Press.
- Hall, D. G., & Moore, C. E. (1997). Red bluebirds and black greenflies: Preschoolers' understanding of the semantics of adjectives and count nouns. *Journal of Experimental Child Psychology, 67*, 236-267.
- Jaswal, V. K., & Markman, E. M. (2002). Children's acceptance and use of unexpected category labels to draw non-obvious inferences. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society* (pp. 500-505). Hillsdale, NJ: Erlbaum.
- Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8*, 113-139.
- Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition, 87*, 179-208.
- Mareschal, D. (2003). The acquisition and use of implicit categories in early development. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making*

- sense of the blooming, buzzing confusion* (pp. 360-383). Oxford: Oxford University Press.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- Markman, E. M., & Jaswal, V. K. (2003). Abilities and assumptions underlying conceptual development. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 384-402). Oxford: Oxford University Press.
- McCarrell, N. S., & Callanan, M. A. (1995). Form-function correspondences in children's inference. *Child Development, 66*, 532-546.
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge, England: Cambridge University Press.
- Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology, 35*, 1072-1078.
- Naigles, L. G., Gleitman, H., & Gleitman, L. R. (1992). Children acquire word meaning components from syntactic evidence. In E. Dromi (Ed.), *Language and cognition: A developmental perspective* (pp. 104-140). Norwood, NJ: Ablex.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.
- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054-1070.

- Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94-104.
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition, 73*, 1-33.
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Science, 7*, 246-251.
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166-188.
- Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgement. *Developmental Psychology, 35*, 1478-1492.
- Sloutsky, V. M., Lo, Y., & Fisher, A. V. (2001). How much does a shared name make things similar: Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72*, 1695-1709.
- Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp. 277-303). Mahwah, NJ: LEA.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*, 13-19.
- Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 132-158). Cambridge, England: Cambridge University Press.
- Tomasello, M. (2003). *Constructing language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Author Note

This research is based, in part, on a Ph.D. dissertation submitted to Stanford University. The research was supported by a predoctoral NIMH National Research Service Award and a Stanford Graduate Research Opportunity grant. Generous support was also provided by the University of Virginia. Portions of this work were presented at the biennial meetings of the Society for Research in Child Development, April 2003, in Tampa, FL, and the Cognitive Development Society meeting, October, 2003, in Park City, UT.

Thanks to Ellen Markman for her support and advice throughout this project; to Herb Clark, Judy DeLoache, John Flavell, Susan Johnson, Angel Lillard, and Tom Wasow for helpful discussions and comments on previous drafts; to Tauna Szymanski, Leslie Neely, Mira Cole, and Jessica Lunsford for assistance in data collection and coding; to Tania Lombrozo for assistance in creating stimuli; and to the children, parents, teachers and staff at Bing Nursery School, the Menlo-Atherton Parents' Cooperative, and Merridale Preschool.

Correspondence concerning this article should be addressed to: Vikram Jaswal, Department of Psychology, University of Virginia, P.O. Box 400400, 102 Gilmer Hall, Charlottesville, Virginia 22904-4400. E-mail: jaswal@virginia.edu.

Table 1

Stimulus Sets

	Stimulus set	Activity/background
Warm-up	Doll	Sleeps on the bed
	Shovel	Goes with the bucket
Animals	Cat	Drinks milk
	Dog	Eats bones
	Horse	Sleeps in the barn
	Cow	Sleeps outside on the grass
	Bear	Lives in the forest
	Pig	Lives in the mud
	Fish	Lives in the lake
	Bird	Lives in the nest
Artifacts	Key	Goes with the car
	Spoon	Goes in the bowl
	Shoe	Goes on the baby
	Car	Goes on the road
	Cup	Goes on the table
	Hat	Goes on the man's head
	Button	Goes on the coat
	Ball	Goes through the hoop

Table 2

Percentage of perceptually based inferences about hybrid items in Studies 1-3. (Standard deviations in parentheses)

	Trial type	
	No-label	Label
Study 1: Bare Label		
3-year-olds	88% (15) *	33% (31) *
4-year-olds	91% (14) *	59% (32)
Study 2: Actually		
3-year-olds	83% (20) *	18% (25) *
4-year-olds	88% (20) *	13% (28) *
Study 3: Pre-nominal modification		
Subtype	91% (14) *	24% (26) *
Familiar Adjective	85% (22) *	32% (35) *

* Significantly different from chance of 50%, $p < .05$.

Table 3

Categories of Spontaneous Comments Made about Hybrid Stimuli in Study 1

Category	Examples
Denials, rejections, corrections	<p>Age 3;5: On hearing a bird-like animal called a "fish:" "No, this is a bird."</p> <p>Age 4;1: On hearing a horse-like animal called a "cow:" "It's not a cow because it doesn't have nipples."</p> <p>Age 4;1: On hearing a cup-like object called a "hat:" "It's not a hat; it's a cup."</p>
Other comments relevant to the appearance and/or label	<p>Age 4;3: On hearing a cat-like object called a "dog:" "That doesn't look like a dog. Look at big ears. Like a cat because this has (???) ears. And a tail too. What should it be? Doesn't look like a dog."</p> <p>Age 4;1: On hearing a shoe-like object called a "car:" "It's a funny, funny shoe-car!"</p> <p>Age 4;4: On hearing a button-like object called a "ball:" "Only it's a brown one."</p>
Irrelevant, no comment	Comments including "Yeah!" or animal sounds.

Figure Captions

Figure 1. Cat-dog continuum. All children saw the standards and typical stimuli. Half of the children in each study saw the cat-like hybrid stimulus, and half saw the dog-like one. Children saw the stimuli in color.

Figure 2. Distribution of spontaneous comments in response to hearing hybrid stimuli labeled, as a function of whether label-based or perceptually based inferences were subsequently made.

Standard Cat Typical Cat Hybrid Cat Hybrid Dog Typical Dog Standard Dog



Label: "Cat"
No-Label: "Cat"

"Cat"
"One"

"Dog"
"One"

"Cat"
"One"

"Dog"
"One"

"Dog"
"Dog"

