

# The Effect of Vocabulary Size on Toddlers' Receptiveness to Unexpected Testimony About Category Membership

Vikram K. Jaswal  
*Department of Psychology*  
*University of Virginia*

Children must be willing to accept some of what they hear “on faith,” even when that testimony conflicts with their own expectations. The study reported here investigated the relation among vocabulary size, object recognition, and 24-month-olds' ( $N = 40$ ) willingness to accept potentially surprising testimony about the category to which an object belongs. Results showed that children with larger vocabularies were better able to recognize atypical exemplars of familiar categories than children with smaller vocabularies. However, they were also most likely to accept unexpected testimony that an object that looked like a member of one familiar category was actually a member of another. These results indicate that 24-month-olds trust classifications provided by adult labeling patterns even when they conflict with the classifications children generate on their own.

The appearance of an object is normally a very good cue to its category membership (e.g., Jones & Smith, 1993; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Even 3.5-month-old infants can form perceptually based categories: They will habituate to a series of pictures of cats, for example, and will recover interest when shown a picture of a dog (Quinn, Eimas, & Rosenkrantz, 1993). As adults, however, we know that looks are not everything. An eel may look a lot like a snake, but it is, in fact, a fish. A mechanical pencil may look like a pen, but it is, in fact, a pencil. In short, things are not always what they seem. How do children learn about category members whose appearance is uninformative or even deceiving?

One way they can do so is through personal experience. For example, given sufficient experience with a set of objects, children may categorize them by non-obvious causal or functional similarities rather than obvious perceptual ones (e.g., Gopnik & Sobel, 2000; Kemler Nelson, Russel, Duke, & Jones, 2000). Of course, this requires not only that children have the opportunity to explore, observe, and interact with the objects in question, but also that they have the expertise necessary to detect the relevant nonobvious similarities. If children were limited to learning about categories and category members on the basis of their own experience, most would never learn, for example, that a whale is a mammal rather than a fish.

Another potentially more efficient way children can learn about nonobvious categories and category members is by taking advantage of another person's knowledge or expertise—for example, by being receptive to labels another person uses when referring to objects. A child who accepts an adult's assertion that a penguin is a bird, for instance, can go on to make inferences about penguins that may be different from those she or he would otherwise have made (e.g., that they hatch from eggs, that they have feathers, etc.). A willingness to accept another person's word, or their "testimony"—even when its basis is unclear or counterintuitive—is essential to the transmission of culturally acquired knowledge (Coady, 1992; Harris, 2002; Harris & Koenig, 2006; Markman & Jaswal, 2003; Robinson, 1994). The study reported here investigates the relation among vocabulary size, object recognition, and 24-month-olds' willingness to accept potentially surprising testimony about the category to which an object belongs.

There is a good deal of evidence showing that the names that children hear applied to objects can have a profound influence on the categories they form. Although many of their earliest categories are made up of perceptually similar objects (Samuelson & Smith, 1999), infants as young as 12 to 13 months old can form categories of perceptually diverse objects if they hear those objects referred to by the same name (Graham, Kilbreath, & Welder, 2004; Nazzi & Gopnik, 2001; Waxman & Markow, 1995). Moreover, hearing a label applied to an object can actually lead children to reclassify an object from one known category to another (e.g., Gelman & Coley, 1990; Gelman, Collman, & Maccoby, 1986; Gelman & Markman, 1986, 1987; Jaswal, 2004, 2006; Jaswal & Malone, *in press*; Jaswal & Markman, 2007; Sloutsky & Fisher, 2004). In Gelman and Markman (1986), for example, 4-year-olds were shown a picture of a squirrel and told, "This squirrel eats bugs." They were then shown a picture of a rabbit and told, "This rabbit eats grass." Finally, they were shown a picture of a squirrel that looked very much like a rabbit (e.g., it had very long ears), and were asked, "See this squirrel? Does it eat bugs, like this squirrel, or does it eat grass, like this rabbit?" Children tended to respond on the basis of the label: Even though the animal looked like a rabbit, they inferred that it would eat bugs like a squirrel.

Jaswal and Markman (2007) extended this finding to 24-month-olds. Toddlers watched as the experimenter used a set of small props to show that, for example, a

key was used to start a car and a spoon was used to eat cereal from a bowl. They were then presented with a specially designed hybrid artifact, which looked like a key, but which had some features of a spoon (e.g., a silver bowl-like end). Children were encouraged to use the props to act out whether the hybrid was used to start a car or to eat cereal. Children who heard this key-like object referred to as “this one” used it to start the car. In contrast, those who heard it referred to as “this spoon” were more likely to use it to eat cereal. From these findings, Jaswal and Markman argued that toddlers are willing to set aside a spontaneously generated classification (based on appearance) in favor of one provided via an unexplained and potentially quite baffling category label. They accepted “on faith” that there was a reason why something that looked very much like a key, for example, was being called a spoon.

It is possible, however, that some children were more likely to recognize the hybrid objects than others, which may have had an effect on their willingness to accept the unexpected labels. In a recent study, Smith (2003) showed that vocabulary size plays an important role in children’s ability to recognize atypical exemplars. In that study, the larger a child’s vocabulary, the better able she or he was to recognize highly stylized three-dimensional caricatures. For example, a caricature of a camera consisted of two simple geometric shapes, painted gray: a Styrofoam circle placed on top of a Styrofoam rectangle. Children with more than 100 object labels in their productive vocabularies were more likely than those with fewer than 100 object labels to select the appropriate caricature from an array when asked, for example, “Where is the camera?” Further, children with larger vocabularies were more likely to play with the caricatures in category-appropriate ways. For example, they were more likely to use the caricature of the camera to pretend to take pictures. Interestingly, however, vocabulary size did not influence children’s performance when the stimuli were richly detailed replicas or toys rather than Styrofoam caricatures. Smith argued that as children learn object labels, they also learn to attend to those features of objects that are most relevant to object recognition and to deemphasize less relevant features.

Like the caricatures used in Smith (2003), the hybrid objects used in Jaswal and Markman (2007) represented unusual exemplars. For example, as described earlier, the key-like object was designed to look primarily like a key, but it had some features of a spoon, and adults rated it as an atypical key exemplar (Jaswal, 2003). Jaswal and Markman (2007) did not measure the vocabulary size of their participants, but given Smith’s (2003) findings, it is possible that children with large vocabularies were more likely to identify the unusual hybrids as members of the categories they were designed to resemble. If this were true, then when the experimenter asserted that, for example, the key-like object was a spoon, this would have presented a conflict primarily to children with large vocabularies. Indeed, to children with small vocabularies, the label may have served the useful function of identifying an otherwise unrecognizable object.

The goal of this study was to investigate the relation among vocabulary size, object recognition, and toddlers' willingness to accept potentially surprising testimony about an object's category membership. Specifically, we asked whether children with larger vocabularies would indeed be more likely to recognize the hybrid objects than those with smaller vocabularies, and whether this would affect their willingness to accept the experimenter's labels—labels that, to an adult, clearly conflicted with the appearance of the hybrid objects. One possibility is that children with larger vocabularies will be more reluctant than those with smaller vocabularies to accept that a key-like object is a spoon, for example, because they may be more likely to assume spontaneously that the object is a key. An alternative possibility is that hearing an adult refer to an object with a particular label is such a powerful source of information that toddlers will defer to that label regardless of their initial expectations.

To investigate these possibilities, we followed the procedure used by Jaswal and Markman (2007). Toddlers were shown a series of objects from familiar categories. In addition to typical exemplars, children would occasionally be presented with one of the hybrid objects described earlier. They would either hear the experimenter refer to the hybrid neutrally ("this one"), or by using a label that may have conflicted with children's expectations about the object given its appearance. Children were asked to demonstrate the function of each object by acting out either an activity that was consistent with its appearance or one that was consistent with the label (if one was provided). We obtained a measure of participants' productive vocabulary size by asking parents to complete and return the MacArthur-Bates Communicative Development Inventory (Fenson et al., 1994).

## METHOD

### Participants

Forty 24-month-olds ( $M = 24$  months, 12 days; range = 23;8–26;7) participated in a single 10- to 15-min session. Nineteen were girls, and 21 were boys. Five additional toddlers were tested, but their data were not included due to extreme fussiness resulting in an inability to complete at least half of the session ( $n = 3$ ) or experimenter error ( $n = 2$ ). Children came from a variety of backgrounds, but most were White.

### Design

Children were randomly assigned to a label or no-label condition, resulting in 20 participants per condition, approximately balanced for sex. The average age was 24;9 in the label condition and 24;15 in the no-label condition.

## Materials

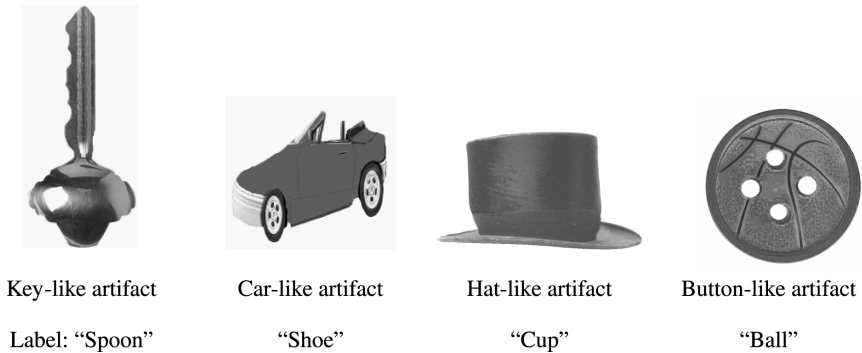
Eight artifacts from familiar categories were grouped into four pairs based roughly on shape similarity: key–spoon, shoe–car, cup–hat, and button–ball. To confirm that these categories were familiar, when parents arrived at the lab, they completed a brief vocabulary checklist that asked the level of their child’s understanding of each category label: definitely understands, maybe understands, or does not understand. Due to experimenter error, this checklist was not completed correctly for 4 children (3 in the label condition and 1 in the no-label condition). Of those children whose parents did complete the checklist correctly, seven of the eight labels were said to be definitely understood by all children in the label condition, and six of the eight were said to be definitely understood by all children in the no-label condition. Finally, all but 1 child were reported to at least “maybe” understand all eight labels. Table 1 shows the percentage of children in each condition reported to definitely understand each label. This analysis confirms that the categories and labels used here were familiar to all children.

Color photographs of a typical exemplar of each artifact were obtained from a digital library of photo objects (Hemera Technologies, Gatineau, Quebec, Canada). These images will be referred to as the demonstration exemplars. An additional typical exemplar of each category was created (typical test exemplars), primarily by manipulating the color of each standard. Finally, two hybrids were created for each set: The hybrids had features of both categories, but were computer-generated to look more like a member of one category than the other. For example, in the spoon–key set, the hybrids had perceptual features of both spoons and keys, but one was designed to look more like a spoon and the other was designed to look more like a key. Examples of the hybrids are shown in Figure 1.

TABLE 1  
Stimulus Sets

<i>Stimulus Set</i>	<i>Average Comprehension<sup>a</sup></i>		<i>Function (Background Photo in Bold)</i>
	<i>Label</i>	<i>No-Label</i>	
Key	100%	95%	Starts the <b>car</b>
Spoon	100%	100%	Used to eat cereal from the <b>bowl</b>
Shoe	100%	100%	Goes on the <b>baby</b> ’s foot
Car	100%	100%	Drives on the <b>road</b>
Cup	100%	100%	Sits on the <b>table</b>
Hat	100%	100%	The <b>man</b> wears it
Button	64%	79%	Goes on the <b>coat</b>
Ball	100%	100%	Drops through the <b>hoop</b>

<sup>a</sup>Average percentage of participants in each condition reported by parents to “definitely” understand each artifact label. Note that the percentages rise to at least 95% in each cell if the criterion is relaxed to “maybe” understands each label.



**FIGURE 1** Examples of hybrid artifacts. Children saw the stimuli in color.

Adult ratings have previously shown that, to an adult, these hybrids are atypical exemplars of the intended categories (Jaswal, 2003). A set of warm-up stimuli was also constructed, and consisted of pictures of typical dolls and shovels. All images were approximately 5 to 10 cm wide and 5 to 12 cm tall. Each image (and its left–right reverse) was printed in color, cut out, and mounted into a small stand so that it could stand on its own.

Each artifact was associated with a typical function, as shown in Table 1. For example, the spoon’s function was to eat cereal from a bowl, and the key’s function was to start a car. Color photographs of a bowl and a car were used to act out these functions. Photos were mounted onto 22 cm × 28 cm easels.

## Procedure

Children participated individually in the laboratory. They sat at a small table, with the researcher across from them. Each session began with a warm-up trial to introduce the task: On one easel, the experimenter displayed a photo of a bed and, using the doll stimulus, demonstrated and explained that a doll slept in the bed. This involved holding the doll up to the picture of a bed, rotating her as if she were lying down on it, and making snoring sounds. After this demonstration, the researcher handed the child the doll and encouraged him or her to “put the doll to sleep.” Next, on another easel, the researcher showed a photo of a bucket and, using the shovel stimulus, demonstrated and explained that a shovel could be used to scoop sand into or out of the bucket. This involved holding the shovel up to the picture of the bucket and making scooping motions as if dropping sand into or scooping it out of the bucket. The researcher handed the child the shovel and encouraged him or her to use it to scoop sand himself or herself. Children were then shown additional doll and shovel exemplars in alternating order, and they were asked to show what each did, until they succeeded in placing a doll with the bed and a shovel with the bucket

consecutively. Eleven children required at least one correction (5 in the label condition and 6 in the no-label condition).

Test trials were similar to the warm-up trial. For example, the researcher displayed a photo of a bowl of cereal on one easel and introduced the demonstration spoon, explaining, “See I can use the spoon to eat cereal!” She then held the spoon up to the picture of the cereal bowl, and pretended to scoop up cereal and eat it. The researcher then displayed a photo of a car on the other easel, and introduced the demonstration key, explaining, “See the key starts the car!” She held the key up to the picture of the car, turned it, and made motor sounds (“Vroom! Vroom!”). As in the warm-up task, following each demonstration, children were handed the demonstration stimulus and were encouraged to act out the activity they had just seen demonstrated (e.g., “Can you start the car?”). Following the demonstration of both functions, the experimenter offered a reminder: “So remember, we can use the spoon to eat cereal (pointing) and the key to start the car (pointing).” At this point, the demonstration exemplars were removed.

Children were then shown the three test exemplars (another typical spoon, another typical key, and either the key-like hybrid or the spoon-like hybrid), one at a time and in a pseudo-random order (described later). Their task was to act out, using the exemplar and the appropriate background photograph, whether each artifact was used to eat cereal or start the car. Because the demonstration exemplars had been removed, children had to rely on their memory of what had come before, or on their preexisting expectations about the functions of spoons and keys.

Half of the children participated in a label condition, where they heard the experimenter use a category label when introducing each test artifact (e.g., “Look at this spoon! Can you show me what this spoon does?”). The typical test artifacts were always called by their appropriate labels, and the hybrids were always called by labels that were the opposite of their appearance (e.g., the key-like hybrid was referred to as “this spoon”; see Figure 1). To establish baseline levels, the other half of the children participated in a no-label condition. They heard the experimenter use the phrase “this one” to introduce each test exemplar (e.g., “Look at this one! Can you show me what this one does?”). Regardless of children’s selections, the researcher responded neutrally (“Thank you!”), and proceeded to the next test artifact or the next set of artifacts.

The order in which the four artifact sets were presented was counterbalanced according to a Latin Square design. The background photograph consistent with the function associated with a hybrid’s appearance appeared twice on the left and twice on the right for each child, and this was counterbalanced across children so that each photo appeared equally often on the left and right. Most children were presented with the four sets of artifacts shown in Table 1; however, due to fussiness, 2 children in the label condition did not complete the car–shoe set and 1 in the no-label condition did not complete the cup–hat set.

Coding of the children's responses was conducted offline, via videotape, and involved noting which of the two background photos children indicated when presented with each test exemplar. Children often made the movements (and sounds if applicable) that had earlier been demonstrated, but pointing gestures were also coded as responses. Occasionally, a child made one response and then switched; in these cases, the final response was coded. Two coders, blind to condition, each coded half of the sessions. To assess reliability, each coder also independently coded a random selection of one-fourth of the other coder's sessions; between-coder agreement was 99%.

Parents were asked to complete and return the MacArthur-Bates Communicative Development Inventory (Words and Sentences; Fenson et al., 1994). Our primary interest was in Part 1, which focuses on productive vocabulary. Parents were provided with a preaddressed, stamped envelope, and were asked to return the form as soon as possible after the session.

## RESULTS

Children's responses were scored as follows: A 1 was assigned if a child made a perceptually based inference by indicating the function associated with an artifact's appearance (e.g., a key or key-like object was used to start the car), and a 0 was assigned otherwise (e.g., a key or key-like object was used to eat cereal). Preliminary analyses showed no effects or interactions involving gender; subsequent analyses collapsed across this factor.

### Overall

Before analyzing the influence of vocabulary size on performance, we first confirmed that this study replicated the main findings from Jaswal and Markman (2007). As Table 2 shows, toddlers made perceptually based inferences about the typical test items over 86% of the time. On seeing a typical key, for example, chil-

TABLE 2  
Percentage of Perceptually Based Inferences About Hybrid Stimuli

	<i>Typical Test Items</i>		<i>Hybrid Test Items</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No-label <sup>a</sup>	86%*	14%	68%*	29%
Label <sup>a</sup>	89%*	14%	31%*	29%

<sup>a</sup>*n* = 20.

\*Significantly different from chance of 50%, *p* < .01.

dren inferred that its function was to start the car rather than to eat cereal, regardless of whether they heard it referred to as “this key” or “this one.” Although children were not perfect in making the appropriate inferences about the typical items, they did so more frequently than expected by chance of 50%,  $t_s > 11.70$ ,  $p_s < .0001$ ,  $d_s > 2.616$ , indicating that they clearly understood the task.

Crucially, as Table 2 also shows and as in Jaswal and Markman’s (2007) study, toddlers in the no-label condition were more likely than chance to make appearance-based inferences about the hybrid items,  $t(19) = 2.79$ ,  $p < .05$ ,  $d = .62$ , whereas those in the label condition were more likely to make the opposite, label-based inferences,  $t(19) = 2.95$ ,  $p < .01$ ,  $d = .66$ . For example, on seeing the key-like hybrid, those who heard it referred to as “this one” inferred that it was used to start a car, and those who heard it referred to as “this spoon” inferred that it was used to eat cereal. A one-way analysis of variance (ANOVA) confirmed that children in the no-label condition made significantly more perceptually based inferences than those in the label condition (68% vs. 31%),  $F(1, 38) = 16.41$ ,  $p < .001$ ,  $\eta_p^2 = .30$ .

The difference between the no-label and label conditions was also evident at the individual level. Table 3 shows the number of children who made zero, one, two, three, or four perceptually based inferences about the hybrid exemplars. (The table and analysis do not include 1 child in the no-label condition or the 2 children in the label condition who responded to only three of the four stimulus sets). As the table shows, the distribution of responses in the two conditions differed significantly, with more children in the no-label condition making more perceptually based inferences,  $\chi^2(4, N = 37) = 13.39$ ,  $p < .01$ , Cramer’s  $\phi = .60$ .

### By Vocabulary Size

Vocabulary inventories were returned for 15 of the 20 participants (75%) in each condition. Because we received these after children participated, it was not possi-

TABLE 3  
Number of Children Making Zero, One, Two, Three, or Four Perceptually Based Inferences About Hybrid Stimuli

	<i>No. of Perceptually Based Inferences</i>				
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
No-label	1	1	5	6	6
Label	7	4	4	3	0

*Note.* Excludes 1 participant in no-label condition and 2 participants in label condition who completed only three of the four trial blocks.

TABLE 4  
Average Number of Words in the Productive Vocabulary  
of Children in Each Condition

	<i>Below Median<sup>a</sup> ("Low")</i>			<i>Above Median ("High")</i>		
	<i>n</i>	<i>M</i>	<i>Range</i>	<i>n</i>	<i>M</i>	<i>Range</i>
No-label	6	292.0	52–401	9	538.4	420–664
Label	9	265.7	81–417	6	502.0	448–571

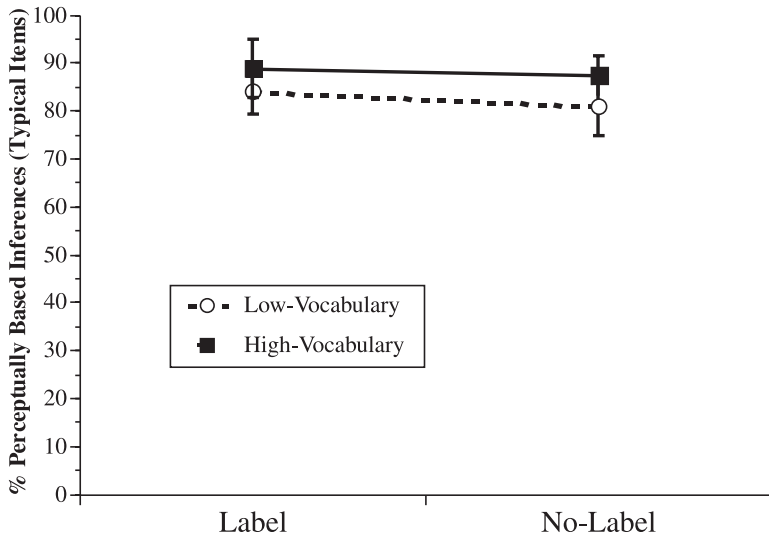
<sup>a</sup> Median productive vocabulary was 419 words.

ble to assign participants to a particular condition based on their vocabulary size. Instead, after receiving all of the forms, we grouped children according to whether their productive vocabulary size was below or above the median of 419 words (low-vocabulary vs. high-vocabulary).<sup>1</sup> Table 4 shows the means and ranges of productive vocabulary for children in each group. A  $2 \times 2$  (Condition  $\times$  Vocabulary Level) ANOVA on these data confirmed that children in the high-vocabulary group had significantly more words in their productive vocabularies than those in the low-vocabulary group,  $F(1, 26) = 48.52, p < .0001, \eta_p^2 = .65$ . There was no effect of condition and no interaction,  $F_s < 1$ .

Our primary interest was in what role, if any, vocabulary size played in toddlers' inferences about the typical and hybrid objects, and whether this varied depending on condition. To get at this question, we conducted a three-way ANOVA (Vocabulary Level  $\times$  Stimulus Type  $\times$  Condition) on the number of perceptually based inferences children made. Of most interest in this analysis was a significant three-way interaction,  $F(1, 26) = 5.76, p < .05, \eta_p^2 = .18$ . To explore this complex interaction further, we analyzed the inferences children made about the typical and hybrid stimuli separately.

The first analysis focused on how vocabulary size and condition were related to inferences children made about the typical items. As Figure 2 shows, regardless of whether children's vocabulary size was above or below the median, and regardless of whether they were in the label or no-label condition, children made perceptually based inferences about the typical exemplars at near-ceiling levels. This conclusion was supported by a  $2 \times 2$  (Condition  $\times$  Vocabulary Level) ANOVA on the data

<sup>1</sup>In her study investigating the recognition of shape caricatures, Smith (2003) analyzed results according to whether children had more or fewer than 100 object names in their productive vocabulary. We were unable to use the same cutoff because this study involved children who were, on average, 4 months older; as a result, only 2 children had fewer than 100 object names in their productive vocabulary (1 in the label condition and 1 in the no-label condition). Additionally, whereas Smith used object names in her criteria, we found the same pattern of results when analyses were based on all vocabulary items on the CDI, and so have chosen to report the latter here.

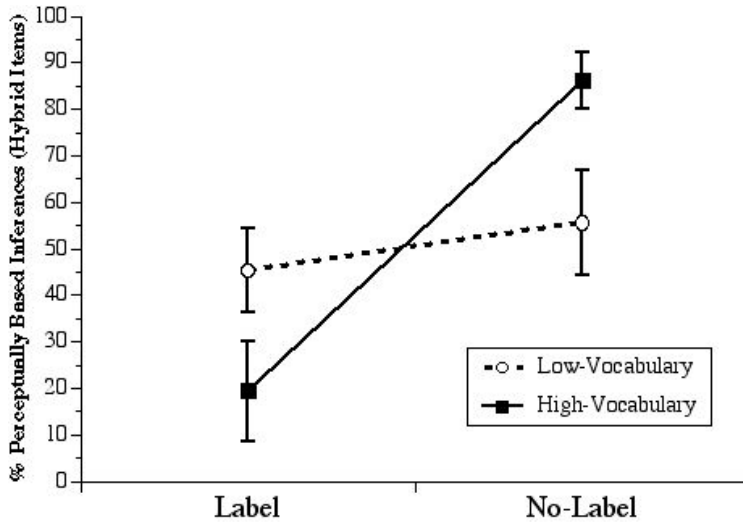


**FIGURE 2** Average percentage of perceptually based inferences children made about typical items, as a function of condition and of whether their productive vocabulary size was above (high-vocabulary) or below (low-vocabulary) the median. Error bars show *SEM*.

shown in Figure 2, which showed no significant effects or interactions,  $F_s < 1.04$ . Consistent with Smith's (2003) research using lifelike objects, vocabulary size did not affect children's ability to make category-appropriate inferences about the typical items.

The second analysis investigated the relation among vocabulary size, condition, and the inferences children made about the hybrid items. As Figure 3 shows, labeling had an effect on the performance of children in the high-vocabulary group, but not on the performance of those in the low-vocabulary group. A  $2 \times 2$  (Condition  $\times$  Vocabulary Level) ANOVA on the data in Figure 3 revealed a significant effect of condition,  $F(1, 26) = 17.70, p < .001, \eta_p^2 = .41$ , and a significant interaction,  $F(1, 26) = 9.56, p < .01, \eta_p^2 = .27$ . The Condition  $\times$  Vocabulary Level interaction was analyzed using simple main effects analyses.

Based on Smith's (2003) results using atypical, caricature-like objects, we had expected that, left to their own devices, children with low vocabularies might find our hybrid stimuli ambiguous with respect to category membership. Indeed, considering just the children in the no-label condition, those with vocabulary sizes below the median were apparently ambivalent about the category to which each hybrid belonged. They made significantly fewer perceptually based inferences about the hybrids than those with vocabulary sizes above the median (56% vs. 86%),  $F(1, 26) = 5.68, p < .05, \eta_p^2 = .18$ .



**FIGURE 3** Average percentage of perceptually based inferences children made about hybrid items, as a function of condition and of whether their productive vocabulary size was above (high-vocabulary) or below (low-vocabulary) the median. Error bars show *SEM*.

We had also expected that when they heard the speaker label the hybrids, children in the low-vocabulary group would make more label-based (and therefore fewer perceptually based) inferences. After all, the speaker's labels should have helped to disambiguate otherwise unrecognizable stimuli. In fact, however, labeling had no effect on children in the low-vocabulary group: Low-vocabulary children in the label condition were as likely to make perceptually based inferences as low-vocabulary children in the no-label condition (45% vs. 56%),  $F(1, 26) = .63, p = .44, \eta_p^2 = .02$ .

Where labeling did have a clear effect was on the inferences made by children with a large vocabulary. In this group, and as Figure 3 clearly shows, those who heard the hybrids referred to neutrally were significantly more likely to make perceptually based inferences than those who heard them referred to with labels (86% vs. 19%),  $F(1, 26) = 26.69, p < .0001, \eta_p^2 = .51$ . Interestingly, among the children in the label condition, those with low vocabularies were marginally more likely to make perceptually based inferences than those with high vocabularies (45% vs. 19%),  $F(1, 26) = 4.06, p = .06, \eta_p^2 = .14$ . In short, children who were most likely to have an expectation about the category to which each hybrid belonged—those with large vocabularies—were also the most likely to defer to the speaker's counter-intuitive labels.

## GENERAL DISCUSSION

The results from this study offer several important insights into the relation among vocabulary size, object recognition, and toddlers' willingness to defer to adult testimony about category membership. First, children were able to recognize the typical test items regardless of the size of their vocabularies, but those with larger vocabularies were better able to recognize the hybrid items than those with smaller vocabularies. Specifically, toddlers with larger vocabularies who heard the hybrid objects referred to neutrally were more likely than those with smaller vocabularies to infer that the hybrids had functions consistent with the category they were designed to resemble.

This finding is consistent with the results from Smith's (2003) study, which showed that children with smaller vocabularies readily recognized lifelike replicas, but were much less likely than those with larger vocabularies to recognize highly stylized, 3-D caricatures. As Smith argued, as children learn more categories, they learn to focus on those perceptual aspects of an object that are normally most relevant to determining its category membership (e.g., its overall shape and configuration of parts), and to background those aspects that may be more idiosyncratic (e.g., for an artifact, its particular color). In the context of this study, for example, recognizing the button-like hybrid as a button required children to focus on its circular shape and holes, and to ignore its color and basket-ball-like markings.

Given that children with larger vocabularies were better able to recognize the hybrids, they likely also experienced more conflict when they heard the experimenter refer to the hybrids using labels that did not match their appearance. After all, hearing a key-like object referred to as a spoon is surprising only if you think the object is a key. One might therefore have expected that children with larger vocabularies would be more reluctant to accept the labels than children with smaller vocabularies. In fact, however, results showed that this was not true. If anything, there was a marginal trend for children with larger vocabularies to more often make inferences in line with the experimenter's unexpected labels than children with smaller vocabularies.

One unexpected finding was that children in the low-vocabulary group who heard the hybrids labeled did not seem to use those labels to disambiguate the categories to which the hybrids belonged. One explanation could be that children with smaller vocabularies had not yet learned that the label given to an object may be a more reliable source of information about its category membership than its appearance (e.g., Mareschal, 2003; Sloutsky & Fisher, 2004). This explanation seems unlikely, however, because these children did not respond by making primarily perceptually based inferences as would be predicted on that account. Instead, like the low-vocabulary children in the no-label condition, they remained, on average, am-

bivalent. A more likely explanation may be that these children were less attentive to the speaker when she provided the unexpected labels. As a result, like the low-vocabulary children in the no-label condition, they simply remained uncertain about the identity of the hybrids. Unfortunately, the design of this study did not include an independent measure of whether children actually heard and processed the labels before they made their response.

As in this study, Jaswal and Markman (2007) found that, as a group, 24-month-olds who heard hybrid objects referred to neutrally tended to make inferences consistent with their appearance, whereas those who heard them referred to with labels tended to make inferences consistent with their labels. That study, however, was limited in that it did not investigate whether some children were more likely to recognize the hybrids and others more likely to defer to the experimenter's labels. The present study is important because it shows clearly that even though children with larger vocabularies were more likely than those with smaller vocabularies to recognize the hybrids, they readily deferred to the experimenter's labels (and if anything, did so more readily than children with smaller vocabularies). As Jaswal and Markman suggested, then, a simple category label can indeed lead toddlers to give up a compelling, perceptually based classification in favor of a label-based one.

Given that children defer to labels that conflict with their expectations, one important question concerns why. The most likely explanation is that even very young children recognize the authority of adults on matters of naming (and indeed, on most matters in their lives). When they see an object from an unfamiliar category, for example, rather than assuming that they can call it whatever they like, they often seek out its name from an adult. Among their first 50 words, many children have an idiosyncratic term they use specifically to request labels (e.g., "tha?"; Nelson, 1973), and by 2.5 years of age, they make frequent use of "What's that?" questions to elicit testimony about an object's name (Clark, 1991) or function (Kemler Nelson, Egan, & Holt, 2004).

In short, children seem to recognize a "division of linguistic labor," a phrase Putnam (1973, 1975) coined to explain an important aspect of adult word use. Adults, Putnam argued, implicitly recognize that they do not know the criteria for every term they use. For example, although most adults know the word *gold*, few actually know what makes something gold versus fool's gold, and fewer still would be able to perform a test to distinguish between the two. When such a distinction needs to be made (e.g., for appraisal of a piece of jewelry), we rely on the expertise of other people—people who have specialized knowledge in the relevant domain. Of course, children are universal novices (Brown & DeLoache, 1978): Compared to the average 2-year-old, the average adult is an expert in every domain. Thus, it is adaptive for children to be receptive to information that adults provide (Csibra & Gergely, 2006), sometimes even when that information conflicts with their own expectations, as in this study.

Adults normally provide what they believe to be truthful testimony to children and to each other (Coady, 1992; Grice, 1975). For a variety of reasons, however, including error, ignorance, or deception, they sometimes say things that are false (Perner, 1988). Thus, when a speaker says something that conflicts with one's own expectations, a sophisticated listener should initially respond skeptically. Indeed, using a procedure and stimuli similar to those used here, Jaswal (2004) showed that 4-year-old children were often skeptical about the unexpected labels the speaker used. They would often verbally reject the speaker's claim that, for example, the key-like object was a spoon (e.g., "That's not a spoon; it's a key!"). When the speaker made clear, however, that she or he really meant to use the unexpected labels (e.g., "You're not going to believe this, but this is actually a spoon"; Jaswal, 2004) or asserted some special naming rights over the object (e.g., "This is a spoon I made"; Jaswal, 2006), even these older children could be persuaded to defer to the adult's putative expertise.

Two-year-olds in this study did not require any kind of additional support to accept the experimenter's labels. Before concluding that children this age are completely credulous, however, it is important to note that the hybrid stimuli used here were designed to be plausible (but unlikely) members of the named categories. Even though the button-like object looked primarily like a button, for example, it shared the general shape and coloration of a basketball. Previous research has shown that when more obvious misnaming events occur—for example, when a speaker refers to a shoe as a "ball" (two kinds that share relatively little in common perceptually)—many children as young as 16 months spontaneously attempt to correct the speaker (Koenig & Echols, 2003; Pea, 1982). If the discrepancy between the appearance of the hybrids and their labels had been more extreme in this study, children might have responded more skeptically.

We do not yet know whether children younger than 24 months would as readily trust a speaker's unexpected testimony about the category to which a hybrid belongs. On the one hand, some evidence suggests that they may. In Graham, Kilbreath, and Welder (2004), for example, 13-month-olds were shown two moderately dissimilar novel objects and either heard them called by the same novel label or referred to neutrally. Only those infants who heard them referred to by the same name treated them as members of the same category. This procedure differs from the one used in the present study, however, in that it involved completely novel objects. Accepting that two novel objects belong to the same novel category is arguably less challenging than accepting that an object that looks like a member of one familiar category is actually a member of another.

When reclassification is required, Mervis (1987; Mervis, Pani, & Pani, 2003) has proposed that children younger than 2 years of age may not immediately defer to adult authority. In her diary data, she found that although her son readily learned words for objects from unfamiliar categories, he was initially unwilling to reclassify an object simply on the basis of hearing an adult label it. For example, he used

the word *duck* to refer to a pelican, even when adults had called it a pelican. Only on being told that a pelican was “a special kind of birdie” with a large beak did he begin using *pelican* appropriately. Mervis argued that the additional information enabled him to understand an otherwise apparently arbitrary category reassignment, thereby making it more acceptable.

Eventually, Mervis (1987; Mervis et al., 2003) suggested, children do accept category reassignments on the basis of adult labeling patterns alone. For her son, this seemed to happen between 20 and 23 months: At 23 months, he learned that a puffin was more appropriately referred to as a *puffin* rather than a *duck* simply on the basis of hearing an adult use the appropriate label. At this point, she proposed that he had acquired an authority principle, or an understanding that “there exist people (authorities) who know more about forms, functions, and form-function correlations than you do. When these people label an object, they probably are referring to a valid category” (Mervis et al., 2003, p. 248).

There are a number of reasons other than an undeveloped authority principle that a child may not immediately accept a reclassification on the basis of testimony alone.<sup>2</sup> For example, the child may not be paying attention when the label is provided (as may have been the case with the low-vocabulary children in the label condition of this study). Or the child may persist with his or her (incorrect) term because it is easier to articulate or is easier to retrieve from memory. Thus, the development of an appropriate procedure to investigate experimentally the willingness of children younger than 24 months to accept a reclassification on the basis of category labels remains an important direction for future research.

The ability to transmit nonobvious information between individuals and across generations is a uniquely human skill. As Gelman, Hollander, Star, and Heyman (2000) noted, “It is difficult to imagine how a nonlinguistic species could convey that a legless lizard really is a lizard, even though it looks outwardly just like a snake. With language, however, such a concept is elegantly expressed (e.g., ‘This is a lizard’)” (p. 203). Of course, a listener must be willing to accept the speaker’s assertion that the animal is actually a lizard even though it looks like a snake. A willingness to do so is important because it opens up a rich knowledge base from which inferences can be drawn. Although knowing whether an animal is a lizard or snake might not affect whether you would like to touch it or have it as a pet, when it comes to reasoning about that animal’s physiology or behavior, for example, knowing the category to which it really belongs may be crucial. The study reported here has shown that 24-month-olds are willing to trust the classifications provided by adult labeling patterns and to use them as the basis for inference, even when those labeling patterns conflict with the children’s own expectations.

---

<sup>2</sup>I thank an anonymous reviewer for pointing out these alternative possibilities.

## ACKNOWLEDGMENTS

Thanks to the children and parents who participated, to Lauren Hayden, Jennifer Luu, Lauren Malone, Brittany McLaughlin, Molly Wallace, Anna Walters, Cristina Wilcox, and especially Lindsay Goldman for assistance in data collection and coding. Thanks also to Jennifer Henderlong Corpus for helpful discussion. Portions of this work were presented at the International Conference on Infant Studies, June 2006, in Kyoto, Japan, and the Eastern Psychological Association Conference, April 2006, in Baltimore, MD.

## REFERENCES

- Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 3–35). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Clark, E. V. (1991). Acquisitional principles in lexical development. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 31–71). Cambridge, UK: Cambridge University Press.
- Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford, UK: Oxford University Press.
- Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 249–274). Oxford, UK: Oxford University Press.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(3, Serial No. 242).
- 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., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, 57, 396–404.
- Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. *The Psychology of Learning and Motivation*, 39, 201–263.
- 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.
- Gopnik, A., & Sobel, D. M. (2000). Detectingblickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, 71, 1205–1222.
- 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: Vol. 3. Speech acts* (pp. 41–58). New York: Academic.
- Harris, P. L. (2002). What do children learn from testimony? In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 316–334). Cambridge, UK: Cambridge University Press.
- Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, 77, 505–524.
- Jaswal, V. K. (2003). *A developmental study of the division of linguistic labor*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

- Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development, 75*, 1871–1885.
- Jaswal, V. K. (2006). Preschoolers favor the creator's label when reasoning about an artifact's function. *Cognition, 99*, B83–B92.
- Jaswal, V. K., & Malone, L. S. (in press). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development*.
- Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development, 8*, 93–111.
- Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8*, 113–139.
- Kemler Nelson, D. G., Egan, L. C., & Holt, M. B. (2004). When children ask, "What is it?" what do they want to know about artifacts? *Psychological Science, 15*, 384–389.
- Kemler Nelson, D. G., Russel, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development, 71*, 1271–1288.
- 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, UK: Oxford University 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, UK: Oxford University Press.
- Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–233). Cambridge, UK: Cambridge University Press.
- Mervis, C. B., Pani, J. R., & Pani, J. R. (2003). Transaction of child cognitive-linguistic abilities and adult input in the acquisition of lexical categories at the basic and subordinate levels. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 242–274). Oxford, UK: Oxford University Press.
- Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition, 80*, B11–B20.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1–2, Serial No. 149).
- Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two- and three-year-olds. *Journal of Child Language, 9*, 597–626.
- Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 141–172). Cambridge, UK: Cambridge University Press.
- Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy, 70*, 699–711.
- Putnam, H. (1975). The meaning of "meaning." In K. Gunderson (Ed.), *Language, mind, and knowledge* (Vol. 7, pp. 131–193). Minneapolis: University of Minnesota Press.
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*, 463–475.
- Robinson, E. J. (1994). What people say, what they think, and what is really the case: Children's understanding of utterances as sources of knowledge. In C. Lewis & P. Mitchell (Eds.), *Origins of an understanding of mind* (pp. 355–381). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- 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.
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition, 73*, 1–33.

- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*, 166–188.
- Smith, L. B. (2003). Learning to recognize objects. *Psychological Science*, *14*, 244–250.
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.