Journal of Experimental Child Psychology 102 (2009) 182–195

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Compliance, conversion, and category induction

Vikram K. Jaswal *, Olivia K. Lima, Jenna E. Small

Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA

ARTICLE INFO

Article history:
Received 13 December 2007
Revised 22 April 2008
Available online 16 June 2008

Keywords:
Category induction
Categorization
Pragmatic demands
Testimony

ABSTRACT

When children hear an object referred to with a label that is moderately discrepant from its appearance, they frequently make inferences about that object consistent with the label rather than its appearance. We asked whether 3-year-olds actually believe these unexpected labels (i.e., conversion) or whether their inferences simply reflect a desire to comply with the considerable experimental demands of the induction task (i.e., compliance). Specifically, we asked how likely children would be to pass an unexpected label on to another person who had not been present during the labeling event. Results showed that children who used an unexpected label as the basis for inference passed that label on to another person about as often as they could remember it. This suggests that children's label-based inferences do reflect conversion rather than mere compliance.

Introduction

From an early age, children are highly receptive to testimony as a source of information. Among their first 50 words, for example, many have an idiosyncratic term they use specifically to request information from other people (e.g., "tha?" [Nelson, 1973]), and by 2.5 years of age, they make frequent use of "What's that?" questions to request the names of unfamiliar objects (Clark, 1991) or their functions (Kemler Nelson, Egan, & Holt, 2004). When testimony concerns something about which children have no expectations, such as the names or functions of novel objects, their deference to that testimony might not seem particularly surprising. But in some cases, specifically in studies on category induction, young children defer to testimony even when it conflicts with their expectations (e.g., Gelman & Markman, 1986; Graham, Kilbreath, & Welder, 2004; Jaswal, 2004; Sloutsky & Fisher, 2004).

* Corresponding author. Fax: +1 434-982-4766.
E-mail address: jaswal@virginia.edu (V.K. Jaswal).

0022-0965/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2008.04.006

For example, in Gelman and Markman's (1986) classic study, preschoolers were shown one drawing of a fish and one drawing of a dolphin. The experimenter explained that the fish "stays underwater to breathe," whereas the dolphin "pops above the water to breathe." Then children were shown a drawing of an animal that looked like a dolphin but which the experimenter unexpectedly referred to as a "fish," and they were asked how it breathed. Children resolved this conflict by favoring the label; even though the animal looked like a dolphin, they inferred that it breathed underwater like a fish.

The deference to unexpected labels observed in studies on category induction is somewhat surprising because there are other situations in which very young children resist or reject testimony that conflicts with their expectations. For example, Pea (1982) found that by 24 months of age, most children spontaneously objected when a speaker referred to a car as a "ball." Similarly, Koenig and Echols (2003, Study 1) found that nearly all 16-month-olds who heard an experimenter mislabel a series of objects (e.g., refer to a ball as a "duck") responded by providing the veridical label at least once. Koenig, Clement, and Harris (2004) showed that preschoolers discounted information from a speaker who mislabeled familiar objects: They were less likely to trust new information from a speaker who had mislabeled familiar objects than from a speaker who had labeled them correctly (see also Jaswal & Neely, 2006).

One reason why children in studies on category induction may be more deferential to unexpected labels than children in the studies just described has to do with the nature of the labeled stimuli. In many studies where children seem to reject a label, that label and children's expectations are highly discrepant. For example, in Koenig and Echols (2003), the experimenter mislabeled pictures of typical exemplars from familiar categories, sometimes even using labels that crossed an ontological boundary (e.g., a chair was referred to as a "cat"). In contrast, most category induction studies use stimuli that have been carefully selected so that the labels the experimenter uses, although unexpected, are still plausible. For example, Gelman and Markman (1986) used anomalies actually found in nature. One stimulus set involved a kaibab, a squirrel that has unusually long ears. Although a kaibab looks somewhat like a rabbit, it also has some squirrel-like features (e.g., a long tail) and so could possibly be a squirrel (see Gelman & Coley, 1990). Although they did not frame their research in terms of children's deference to an unexpected label, Sloutsky and Fisher (2004) demonstrated experimentally the role that the plausibility of the label plays: In several studies, they found that the more perceptually similar the item was to the labeled category, the more likely children were to make a label-based inference about that item (see also Graham et al., 2004).

Another reason why children in studies on category induction may seem relatively deferential—and the focus of the study presented here—has to do with the demand characteristics of the category induction task. In both category induction studies and mislabeling studies, an adult experimenter names an object using a label that conflicts with the object's appearance (e.g., a chair is called a "cat," an animal that looks like a dolphin is called a "fish"). In mislabeling studies, the children's spontaneous reactions to this mislabeling event are observed (e.g., Koenig & Echols, 2003; Pea, 1982), or the way in which they treat subsequent information from an inaccurate speaker versus an accurate speaker is measured (Koenig et al., 2004).

In contrast, in studies on category induction, an adult refers to an item with a label that does not match children's expectations and then asks children to make either a label-based inference or an appearance-based inference. Children's spontaneous reactions to the unexpected labels are not of interest and are rarely reported (see Jaswal, 2004), and there is no competing testimony that is consistent with their expectations to which children can turn. Because speakers normally provide information that is relevant to the topic at hand (Grice, 1975), a reasonable pragmatic assumption in the category induction task is that the adult believes that the unexpected label is relevant to the inference he or she is asking the children to make. As a result, children may make a label-based inference because they think this is what the experimenter wants them to do rather than because they believe that the labeled object is a member of the named category (e.g., Aronsson & Hundeide, 2002; Donaldson, 1978; Siegal, 1997; Siegal & Surian, 2004).

Indeed, Siegal, Waters, and Dinwiddy (1988, Experiment 2) found that 4- to 6-year-olds recognize that people sometimes make responses to satisfy experimental demands rather than because they actually believe in what they are doing. In that study, children watched a video clip of a puppet performing a conservation of number task involving two rows of buttons. The puppet initially indicated

that one row had more buttons than the other row, but when the experimenter changed the spatial configuration of the two rows and asked a second time which row had more buttons, the puppet selected the other row. Interestingly, most children indicated that the puppet changed his response "just to please the grown-up," not because he "really thought [it] was true." In contrast, when the puppet indicated the same row of buttons both before and after the transformation, most children indicated that the puppet's response reflected his true belief rather than a desire to please the grown-up.

The distinction we wish to make is captured by the terms *compliance* and *conversion* (MacDonald, Nail, & Levy, 2004; Mugny, 1984; Nail, MacDonald, & Levy, 2000). When confronted with a message that conflicts with their own beliefs, children (and adults) may behave in a manner consistent with the message for social reasons (i.e., compliance) rather than because their beliefs have actually changed (i.e., conversion). To our knowledge, no one has yet considered the extent to which children's performance on a category induction task reflects compliance rather than conversion.

This is a crucial question because the category induction task has been used by a number of researchers interested in the structure of children's categories (e.g., Gelman, 2003; Jaswal, 2004; Sloutsky & Fisher, 2004). We will not describe the ongoing controversy about whether children's categories are theory based or similarity based because the current work was not designed to speak directly to this debate (but see Gelman & Waxman, 2007; Sloutsky, Kloos, & Fisher, 2007a, 2007b). Rather, our goal was to develop a new procedure that could be used to investigate whether children who use an unexpected label as the basis for inference actually believe that label.

In our new procedure, one experimenter (E1) provided an unexpected label and asked children to make an inference. She then left the room with the excuse that she needed to get something, and she was replaced by a second, ostensibly naive experimenter (E2). This second experimenter engaged the children in casual conversation while they waited for E1 to return. Of interest was whether children who used E1's unexpected label as the basis for inference about an object would pass that label on to E2 when she casually asked them about the name of that object.

Our use of two experimenters was motivated by several studies showing that young children can use someone's presence or absence to make inferences about that individual's likely knowledge (e.g., Akhtar, Carpenter, & Tomasello, 1996; Diesendruck & Markson, 2001; O'Neill, 1996; Tomasello & Haberl, 2003). For example, Diesendruck and Markson (2001) showed that 3-year-olds did not expect that a puppet would know a novel fact (e.g., which of two objects had been given to the experimenter by her uncle) unless that puppet had been present earlier when the experimenter stated that fact. Children did, however, expect that the puppet would know the name of a novel object—even if the puppet had not been present during the labeling event. Based on these previous studies, we reasoned that (a) children would not assume that E2 knew what had happened when she was out of the room but that (b) they would assume that she knew the conventional names of objects. Thus, when E2 asked about the name of an object, children would not feel pressure to provide the unexpected label E1 used earlier unless they actually believed that to be its conventional label.

Suppose, for example, that E1 refers to a key-like object as a "spoon" and children make a label-based inference, inferring that it is used to eat cereal rather than to start a car. If children later tell E2 that the key-like object is a "spoon," this would suggest that they actually believe it is a spoon. If, however, they tell her that it is a "key," this could suggest that their earlier spoon-like inference had been due to compliance.

It is also possible, however, that children will believe E1's assertion that the key-like object is a "spoon" but that they will later refer to it as a "key" because of memory limitations. After all, the label the experimenter uses is counterintuitive and, therefore, may be difficult to remember. Traditionally, studies on category induction have focused exclusively on the inferences children make immediately on hearing an unexpected label, not on their ability to retain that label. In our own previous work using the category induction procedure (Jaswal, 2004, 2006, 2007; Jaswal & Malone, 2007; Jaswal & Markman, 2007), we have informally asked children to name the objects after completing the study and have found that many who seemed deferential to the unexpected labels earlier nonetheless reverted back to naming the objects according to their appearance.

Thus, another group of children participated in a memory control condition. In this condition, children took part in the category induction task just as those in the experimental condition did. However, rather than having E2 later ask about the names of the objects, E1 later asked children what she had

called them. We expected that children in both conditions would be equally likely to make label-based inferences; the two conditions do not differ at this point in the procedure. The crucial comparison is the difference between the number of labels children in the memory control condition could remember and the number of labels children in the experimental condition passed on to E2. If those in the memory control condition remember more than those in the experimental condition pass on, this would suggest that at least some of the label-based inferences made by children in the experimental condition were due to compliance.

We chose to study 3-year-olds because this age is the focus of a number of previous studies using the category induction procedure (e.g., Gelman & Markman, 1987; Jaswal, 2004; Jaswal & Malone, 2007) and, as noted earlier, various theoretical claims have been made on the basis of the performance of children of this age (and younger) on this task (e.g., Gelman, 2003). It is important to note that there is already evidence showing that the category induction procedure does not always result in conversion. In Jaswal (2004), for example, 3-year-olds ignored the unexpected labels the experimenter used on approximately one-third of the trials, often expressing their skepticism explicitly (e.g., "That doesn't look like a spoon," "That's not a spoon—it's a key!"). Clearly, when children ignore or verbally reject an unexpected label, this reflects neither compliance nor conversion. Our primary interest, therefore, is in those cases where children do defer to the unexpected label the experimenter uses. Does this deference reflect a change in children's beliefs about what the labeled object actually is, or does it reflect mere compliance?

Preliminary study

We first obtained a baseline measure of children's expectations about the stimuli to be used in the experimental induction task. We wanted to ensure that when the items we used were referred to neutrally without labels, children would consistently base their inferences on their appearance. Furthermore, we wanted to ensure that under these conditions, they would later describe those items to another person using labels consistent with their appearance.

Method

Participants

Eight 3-year-olds (4 boys and 4 girls, mean age = 3 years 6 months, range = 3 years 2 months to 3 years 11 months) participated in a single laboratory visit. The children in this study, and in the main study, were recruited through a database of local families who had expressed an interest in participating in research. Most children were Caucasian and from middle- or upper middle-class backgrounds, and all were fluent English speakers.

Stimuli

Four sets of stimuli were prepared from a digital library of photo-objects (Hemera Technologies, Gatineau, QC, Canada). Photos representing prototypical exemplars of eight familiar categories were selected and arranged into pairs to form four stimulus sets: key–spoon, shoe–car, toothbrush–pen, and hat–cup. These stimuli served as "standards." For each stimulus set, two additional "typical" exemplars were created by modifying the color and/or size of the standards. Finally, a "hybrid" exemplar was created from each set; each hybrid exemplar looked mostly like one of the standard exemplars of that set but also shared some features of the other one. For example, the key-like hybrid looked mostly like a key, but it had some spoon-like features (e.g., a silver wide reflective base). In addition to the key-like hybrid, there was a shoe-like hybrid (with some features of a car), a hat-like hybrid (with some features of a cup), and a toothbrush-like hybrid (with some features of a pen). Fig. 1 shows the four hybrid items. It is important to note that these stimuli were not ambiguous in appearance. Although they had features of both categories, the preliminary study was designed to confirm that children would have a robust expectation that these hybrids were members of the category they most resembled.

Fig. 1. Hybrid artifacts. Children saw the stimuli in color.

Each stimulus was approximately 2 to 4 inches in width and 2 to 4 inches in height. Each one (and its left–right reverse) was printed in color and mounted onto a foam core base, which allowed it to stand independently.

Eight additional color photographs were selected to represent a scene or an object with which the function of each stimulus could be demonstrated. For example, a photograph of a car was used to demonstrate that a key could be used to start a car, and a photograph of a bowl was used to demonstrate that a spoon could be used to eat cereal. These photos were approximately 6 × 6 inches and were printed and mounted onto 8.5 × 11-inch easels. Table 1 shows the complete list of functions for each stimulus set.

Finally, a set of warm-up stimuli was constructed and consisted of exemplars of dolls and shovels and background photographs of a bed and a bucket. As is described below, the doll was put to "sleep" in the bed, and the shovel was used to pretend to scoop sand into the bucket.

Procedure

Children were tested individually at the lab. On arrival, they were greeted by two female experimenters (E1 and E2). These two experimenters played with children in the waiting room while their parents completed the consent forms (usually for approximately 10 min). We chose to have both experimenters interact with children before the session because we wanted to be sure that they were familiar and that they were comfortable with both of them. This was especially important because children later needed to provide verbal responses to questions posed by E2. In pilot work, we found that 3-year-olds were extremely reticent when they were not familiar with E2. When parents had completed the consent forms and children appeared to be comfortable, E1 said goodbye to E2 and led children to the experimental room.

E1 began the session with a warm-up trial to familiarize children with the procedure. First, she showed the background photo of a bed. She explained aloud and demonstrated that the doll stimulus

Table 1
Stimulus sets

Stimulus set	Function[a]
Key	Starts the **car**
Spoon	Eats the **cereal**
Hat	Goes on the **man's head**
Cup	Goes on the **table**
Toothbrush	Brushes the **lady's teeth**
Pen	Writes on the **paper**
Car	Goes on the **street**
Shoe	Goes on the **baby's foot**

[a] Bold type indicates the background photograph used during the inference task.

could sleep in the bed. Next, she showed the background photo of the bucket and explained and used the shovel stimulus to pretend to scoop sand into the bucket. Finally, E1 presented children with another doll and shovel, one at a time, and asked where each went. Correct selections were praised and errors were corrected. Four children needed to be corrected and were asked to make inferences about an additional doll and shovel.

After children had successfully placed a doll with the bed and a shovel with the bucket, E1 removed the easels from the table, revealing the exemplars of the dolls and shovels that had been used during that warm-up trial. She explained that she needed to go to a different room to get the next set of stimuli but that she would call E2 to wait with the children. She opened the door and called for E2, who appeared and agreed to stay. E1 left the room and closed the door behind her.

Because children had interacted with both experimenters in the waiting room, they might infer that E2 somehow knew what had happened in the room even though she had not been there. In an attempt to prevent children from coming to this conclusion, each time E2 was asked to wait with them, she seemed surprised and ignorant about the game children had been playing and the items sitting on the table. She casually asked about the names of each one (“Wow! Look at these! These are neat pictures! What’s this one?”). All children used the correct labels (“doll” and “shovel”) when referring to the warm-up stimuli. Shortly after children had provided a name for each item, E1 returned, carrying a new set of stimuli, and E2 left, closing the door behind her. E1 then began the test trials.

Test trials resembled the warm-up trial. For each of the four stimulus sets, children watched as E1 demonstrated and explained that each standard item could be used for a particular function. For example, E1 showed and explained that a key could be used to start a car and that a spoon could be used to eat cereal from a bowl. Following the demonstration, children were presented with the three test items for that stimulus set (two typical and one hybrid), one at a time, in a pseudo-random order such that the hybrid item was presented first, second, and third at least once each across the four trial blocks. E1 referred to each of the test items neutrally without labels: “Look at this! Can you show me where this one goes?” Hybrid items were introduced in exactly the same way as the typical ones; that is, no special attention was drawn to them. Children received neutral feedback following a response (“Okay!”).

After children responded to the three test items of a given set, E1 called E2 to wait with them while she went to get the next set of stimuli. As in the warm-up trial, E2 expressed an interest in the items sitting on the table, casually asking about the names of each one. The order in which E2 asked about the names was the same as the order in which children had been asked by E1 to make an inference about each one. Thus, they were asked about the hybrid item first, second, and third at least once across the four trial blocks. When all of the stimuli had been named, E1 returned with the next set of stimuli and E2 left. E1 then repeated the procedure using the new stimulus set. The four sets of stimuli were presented in one of four orders according to a Latin square design.

Coding in this study and in the primary study was conducted from videotape and consisted of noting which inference children made about each test item and what they called each one when asked about its name. A second coder coded a random selection of 25% of sessions from the preliminary study and the primary study. Reliability was excellent, with 100% agreement on children’s selections during the inference task and 98% agreement on their verbal responses during the naming task.

Results and discussion

Children made appearance-based inferences about the typical and hybrid items on 98% and 100% of trials, respectively. When E2 later asked about the names of those items, children responded by providing labels that matched their appearances on 97% of the trials for the typical items and 100% for the hybrids. For example, children used both the typical key and the key-like object to start the car and later called both “keys,” and they used the typical spoon to eat cereal from the bowl and called it a “spoon.”

From these results, we can be sure that the hybrid objects were seen as members of the categories they were designed to resemble. Even though the hybrids had features of two categories, children did not see them as ambiguous blends of the two. They always made inferences about the hybrids consistent with the particular category they were designed to resemble most, and they always labeled them

using the label of the particular category they were designed to resemble most. Clearly, children had robust expectations about the category to which each hybrid belonged.

Primary study

The primary study was designed to investigate whether children's label-based inferences reflect compliance or conversion. The procedure was similar to that used in the preliminary study except that E1 referred to each hybrid item using a label that conflicted with the one that children in the preliminary study had always used. For example, children in the preliminary study always referred to the key-like object as a "key," but E1 referred to it as a "spoon." Previous work with a similar set of stimuli has shown that under these conditions, 3-year-olds often base their inferences on the unexpected labels (Jaswal, 2004; Jaswal & Malone, 2007).

In the experimental condition, E1 then left the room and E2 entered and casually asked children what each object was called. In the memory control condition, E1 also left the room but then returned immediately and asked children what she had called each object. Our interest was in whether some of the children's label-based inferences reflected compliance due to the experimental demands of the task. If so, children in the memory control condition should remember more unexpected labels than those in the experimental condition pass on to E2.

We also included a trial at the end of the study to ensure that our procedure could identify a case of compliance. As noted in the Introduction, previous category induction studies have used labels that are only moderately discrepant from children's expectations. On this final trial, we purposely violated this convention by using a label that crossed the ontological boundary between natural kinds and artifacts. The experimenter referred to a picture of a typical horse using the name of an artifact (e.g., "key"). Children chose between making an inference that was consistent with the label but was not appropriate for a horse (e.g., using it to start a car) and one that was neither consistent with the label nor appropriate for a horse (e.g., using it to eat cereal). We expected that some 3-year-olds would make a label-based inference simply to satisfy the experimental demands of the task. But we doubted that they actually would believe that the horse was, for example, a key. Thus, we expected that children in the memory control condition would remember that E1 called the horse a "key," but we did not expect that children in the experimental condition would tell E2 that the horse was a "key."

Method

Participants

Participants were 32 3-year-olds (16 boys and 16 girls, mean age = 3 years 6 months, range = 3 years 0 months to 4 years 0 months). None had participated in the preliminary study. One additional child was excluded because of parental interference.

Stimuli and procedure

The stimuli and procedure were similar to those used in the preliminary study. As in the preliminary study, there were four trial blocks, each consisting of three inference test trials (two involving typical items and one involving a hybrid). In the primary study, however, E1 introduced each test item by referring to it twice by name. She referred to the typical test items using labels that matched their appearance and referred to the hybrid test items with labels that did not (e.g., the key-like object was referred to as a "spoon" [see Fig. 1]). As in the preliminary study, children were asked to make an inference about each test item. At the end of each stimulus set, E1 left the room, claiming that she needed to get the next set of stimuli.

Children participated in one of two conditions. In the experimental condition ($n = 16$), E2 entered and asked children what each item in that set was called. In the memory control condition ($n = 16$), E1 returned and asked children "What did I call this?" for each item in that set. The average time between the conclusion of the inference task and the start of the naming task in both conditions was approximately 15 s.

After the final inference of the fourth trial block in both the experimental and memory control conditions, E1 held up a picture of a typical exemplar of a horse (3×3 inches); referred to it twice either

as a "hat," a "key," a "car," or a "toothbrush"; and asked children to make an inference about it. Clearly, none of these labels is appropriate for a horse. The particular label E1 used and the particular inference choices available depended on which stimulus set had comprised the fourth trial block. For example, if the fourth trial block had involved keys and spoons, the background photo showing a van and the one showing a bowl of cereal remained on the table. Children heard E1 refer to the horse as a "key," and were asked to decide whether it was used to eat cereal or to start the car. E1 then left the room and either she was replaced by E2, who asked about the names of all the objects from the fourth trial block and the horse (in the experimental condition), or E1 returned to ask what she had called each object (in the memory control condition).

Results and discussion

Horse trial

We describe results from the horse trial first because it was designed to show that children might make a label-based inference because of compliance and also to ensure that our procedure could identify a case of compliance. Because this involved a single trial, we report results in terms of the number of children who responded in a particular way. Of the 16 children in the experimental condition, 7 made a label-based inference even though E1 used a label that crossed an ontological boundary (e.g., referred to a horse as a "key"), 1 made the opposite inference, and 8 simply refused to respond. Importantly, however, no children used the inappropriate label when E2 entered and asked what the horse was called; all 16 referred to it as a "horse."

The memory control condition showed that this was not because children had forgotten the label E1 used. In this condition, 10 of 15 children made a label-based inference about the horse (1 child did not receive this trial due to experimenter error). When E1 left the room and then returned to ask what she had called it earlier, 7 of those 10 children remembered, saying that she had called it, for example, a "key." The other 3 misremembered, claiming that she had called it a "horse."

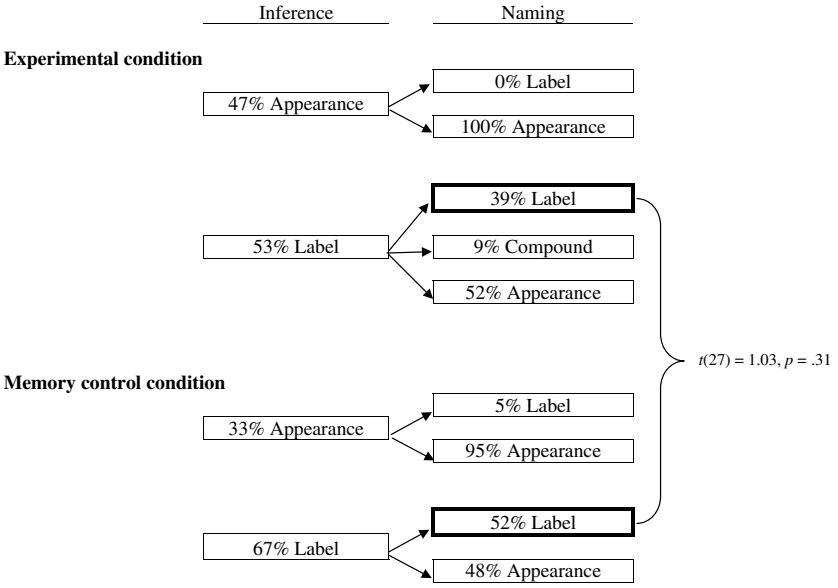

Fig. 2. Distribution of responses to the hybrid stimuli on the inference and naming tasks in each condition of the primary study. E1 asked the inference questions in both the experimental and memory control conditions. E2 asked the naming questions in the experimental condition, and E1 did so in the memory control condition.

The results of this horse trial are important because they demonstrate that the pragmatic demands of the induction task can lead children to make a label-based inference even when the label is clearly inappropriate. Although some children objected and refused to respond when the experimenter mislabeled the horse, approximately half went along with the inappropriate label she used. The results of this trial are also important because they demonstrate that our procedure can identify such cases of compliance: Children in the memory control condition tended to remember the bizarre label E1 used, but those in the experimental condition never passed that label on to E2.

Typical items

Participants received eight trials involving typical items, and so results are reported in terms of the average percentage of trials on which children responded in a particular way. As in the preliminary study, in both the experimental and memory control conditions, children nearly always (at least 99% of trials) inferred that the typical items had functions consistent with their appearance. Those in the experimental condition later described them to E2 using labels consistent with their appearance on 99% of trials. Those in the memory control condition later reported to E1 that she had called them by labels consistent with their appearance on 94% of trials.

Hybrid items

Participants received four trials involving hybrid items, and so results are reported in terms of the average percentage of trials on which children responded in a particular way. Recall that children heard each of the four hybrids referred to with labels that did not match their appearance. As Fig. 2 shows, most inferences in both conditions were consistent with those labels rather than with the appearance of the hybrid. Children made label-based inferences about the hybrids on an average of 53% (*SD* = 33%) of the trials in the experimental condition and 67% (*SD* = 33%) of trials in the memory control condition. Although children were slightly more likely to make label-based inferences in the memory control condition than in the experimental condition, the two conditions did not differ from each other, $t(30) = 1.22$, $p = .23$. Consistent with a good deal of previous research using the category induction task, then, children in both conditions were often initially deferential to the unexpected labels they heard the experimenter use even though those labels conflicted with what we knew from the preliminary study their expectations to be (e.g., Gelman & Markman, 1986, 1987; Jaswal, 2004; Sloutsky & Fisher, 2004).

The critical and novel part of our procedure was the naming task. In the experimental condition, an ostensibly naive E2 replaced E1 and casually asked children what the hybrid from a particular set was called. In the memory control condition, E1 asked explicitly, "What did I call this?" We describe results for naming trials that followed appearance-based inferences first and then turn to naming trials that followed label-based inferences.

As Fig. 2 shows, when children ignored the unexpected label E1 used and made an appearance-based inference, they nearly always later referred to the hybrid with the label consistent with its appearance. They did so 100% of the time in the experimental condition and 95% of the time in the memory control condition. Consider, for example, a trial on which E1 referred to a key-like object as a "spoon." Children in the experimental condition who ignored E1's unexpected label and used the object to start a car rather than to eat cereal always told E2 that it was a "key." Interestingly, children in the memory control condition who ignored E1's unexpected label later mistakenly claimed that she had called it a "key." We return to this intriguing error in the General Discussion.

Our primary interest was in how children would name the hybrids when they had earlier made a label-based inference. Because the memory control condition provides a baseline of how many of the unexpected labels children could actually remember, we describe results from this condition first. As Fig. 2 shows, on those trials when children in the memory control condition made a label-based inference, they subsequently recalled the unexpected label E1 had used, on average, 52% of the time. The other 48% of the time, they incorrectly reported that she had referred to the hybrid using a label consistent with its appearance. Given that children in the memory control condition could remember only half of the unexpected labels E1 used, we would not expect those in the experimental condition to pass on to another person more than 50%. However, if they pass on significantly fewer than 50%, this would suggest that some of their initial label-based inferences were due to compliance rather than conversion.

Table 2
Number of children reporting zero, one, two, or three of the four unexpected labels to E2 (experimental condition) or to E1 (memory control condition)

	Number of initial label-based inferences	Number of unexpected labels reported			
		0	1	2	3
Experimental condition	0	2			
	1	3			
	2	1	3	1	
	3	0	1	2	
	4	0	2	0	1
Memory control condition	0	1			
	1	1			
	2	2	1	2	
	3	0	1	2	
	4	0	2	2	2

As Fig. 2 shows, on those trials when children in the experimental condition made a label-based inference, they subsequently passed the unexpected label on to E2, on average, 39% of the time, used a compound noun that combined the label and appearance (e.g., "shoe–car") 9% of the time, and used a label consistent with the hybrid's appearance the remaining 52% of the time. Although children were slightly less likely to report the counterintuitive label in this condition than in the memory control condition (39% vs. 52%), this difference was not significant, $t(27) = 1.03$, $p = .31$, Cohen's $d = .38$.[1] A power analysis indicated that it would take more than 107 participants in each condition for an effect of this size to become significant, and even then the magnitude of the difference between the two conditions (13%) would appear to be quite small.

In short, children in the experimental condition passed on roughly as many unexpected labels to E2 as those in the memory control condition could remember. This suggests that when 3-year-olds made a label-based inference about a hybrid, they usually did so because they actually believed that label, not simply because of the experimental demands of the situation.

Individual response patterns

One important question concerns the extent to which the group data in each condition are representative of individual response patterns. Table 2 shows the number of children reporting zero, one, two, or three of the four unexpected labels during the naming task to E2 (in the experimental condition) or to E1 (in the memory control condition) as a function of the number of label-based inferences they made initially. There are two important features to note from the table. First, during the initial inference task, most children made label-based inferences on at least half of the trials (11 of 16 in the experimental condition and 14 of 16 in the memory control condition). In addition, the distribution of the number of children making each number of label-based inferences was similar across the conditions, $\chi^2(4, N = 32) = 2.33$, $p = .67$. Thus, most children were neither completely credulous nor completely skeptical initially, consistent with the findings from other category induction studies with 3-year-olds (Gelman & Markman, 1987; Jaswal, 2004).

Second, 10 of 11 children (91%) in the experimental condition who made two or more label-based inferences passed on at least one of those unexpected labels to E2. Similarly, 12 of 14 children (86%) in the memory control condition who made two or more label-based inferences recalled at least one of those labels later. Thus, results were not driven by a small subset of children who both were willing to believe anything E1 said and had excellent memories.

[1] Note that the degrees of freedom for this *t* test depended on the number of children who made a label-based inference. Because 14 children in the experimental condition and 15 children in the memory control condition did so, the degrees of freedom for this test was $df = 27$.

General discussion

There are at least two reasons why children participating in a category induction study might make an inference consistent with an unexpected label they just heard: Children might actually believe that label, or they might simply feel pressure to comply with the label because of the considerable experimental demands of the task. The study reported here is important because it is the first to acknowledge and explicitly address these possibilities. The results suggested that 3-year-olds' label-based inferences about the hybrid stimuli reflected conversion rather than compliance. Children in the experimental condition passed on roughly as many of the unexpected labels for the hybrid items to another, ostensibly naive person as children in the memory control condition could remember.

Our horse trial demonstrated that the demands of the induction task could lead children to make a label-based inference and that our procedure could identify a case of compliance. On this trial, the experimenter referred to a horse with the label of an inanimate object, violating the convention used in previous studies of category induction, where the unexpected label has typically been from the same ontological category as the named object. Nevertheless, approximately half of the children made a label-based inference about the horse. Of these children, most in the memory control condition could later remember the inappropriate label the experimenter had used, but none in the experimental condition transmitted it to another person. Thus, the initial label-based inferences made about the horse on this trial most likely reflected compliance due to the experimental demands of the induction task.

That some children were deferential on the horse trial is consistent with other research showing that children sometimes respond in particular ways so as to "please the grown-up" (e.g., Siegal et al., 1988). It is also consistent with research from the child eyewitness testimony literature, which has demonstrated that very young children may respond in a particular way because they have as their primary goal a desire to affiliate with the adult interlocutor (e.g., Ceci & Bruck, 1993). This may conflict with the experimenter's goal, which of course is to obtain responses that reflect children's true beliefs (Aronsson & Hundeide, 2002; Siegal & Surian, 2004).

In contrast to children's label-based inferences on the horse trial, their label-based inferences about the hybrids seemed to reflect conversion. Children in the experimental condition were willing to pass on a number of unexpected labels to a second experimenter who had not been present during the labeling episode. We reasoned that children would not report those labels to E2 unless they actually believed them to be the conventional terms for those objects (e.g., Diesendruck & Markson, 2001). One objection could be that children did not see E1 and E2 as independent. If children thought that E2 somehow knew the unexpected labels E1 had used, they might have felt the same pressure to comply even though the naming task was conducted by E2. We think that this is unlikely for three reasons. First, the same procedure was used on the trial involving the horse, and children in the experimental condition always told E2 that it was a "horse" rather than the unusual label E1 had used earlier. If children had felt pressure to provide E2 with the unusual label E1 had used, we would have expected them to report that label. Second, we attempted to make it clear that E2 was not familiar with the stimuli. She feigned surprise each time she entered the room and saw a new set of objects ("Wow! Look at these!"). Finally, as noted in the Introduction, a number of studies have shown that very young children readily make use of a person's presence or absence to make inferences about their likely knowledge (e.g., Akhtar et al., 1996; Diesendruck & Markson, 2001; O'Neill, 1996; Tomasello & Haberl, 2003).

One question concerns why children seemed to show evidence of compliance on the horse trial and conversion on the hybrid trials. This difference likely stems from the fact that the label E1 provided during the horse trial was highly discrepant from children's expectations; it crossed the natural kind–artifact ontological boundary. Although many children were willing to play along with the experimenter when she referred to the horse as a "key," for example, they did not actually believe that it was a key. In contrast, the labels E1 provided during the hybrid trials were unexpected but plausible; they named other artifacts that were roughly the same size and shape as the depicted artifact.

In most studies on category induction, including the one reported here, children are generally deferential to the moderately discrepant labels they hear an adult use. There are a number of other sit-

uations in which children also defer to an adult's label rather than favoring their own. For example, even if children overextend the word *dog* to all four-legged animals, they eventually do learn to refer to non-dogs appropriately (e.g., Mervis, Pani, & Pani, 2003). Similarly, children sometimes coin new words to fill gaps in their lexicon (e.g., "don't broom my mess" for "don't sweep my mess"), but these innovations fall out as soon as children learn the appropriate vocabulary (Clark, 1982, 1991).

This raises an intriguing question: Why should children weight what someone else says more heavily than their own expectations? We suspect that this deference reflects the fact that children recognize from an early age that language is conventional (e.g., Clark, 1991; Henderson & Graham, 2005) and that they have a general inclination to trust what adults tell them (e.g., Csibra & Gergely, 2006; Dawkins, 2004; Harris, 2002; Reid, 1764/1997). Despite this general inclination to trust adults, children are not entirely credulous: They may respond skeptically if an adult says something that is highly discrepant with their expectations (e.g., Koenig & Echols, 2003; Pea, 1982), if an adult expresses uncertainty about something that is moderately discrepant (Jaswal & Malone, 2007), or if an adult has been wrong in the past (Jaswal & Neely, 2006).

The memory control condition showed that children had difficulty in remembering the unexpected labels. Children in this condition usually claimed that E1 had used a label consistent with a hybrid's appearance rather than the one she had actually used just moments before. These errors are understandable in the context of the considerable information-processing demands of our task. First, by design, the hybrids did not look like what the experimenter claimed they were. For example, the key-like object had some features of a spoon, but it really did look like a key (as the preliminary study showed). Thus, children needed to remember that an object that looked very much like a key was actually a spoon, a task made even more challenging because the experimenter referred to it as a "spoon" just twice. Second, the experimenter referred to the typical items with labels that did match their appearance, so children needed to keep track of which items in a particular set had been referred to as expected and which ones had not. Finally, children may have been distracted by the fact that the experimenter left the room after the inference task and before the naming task—a procedural decision we made to keep the memory control condition as close to the experimental condition as possible.

Other studies in which children learn something that is counterintuitive have also shown that children sometimes later "misremember" this information (e.g., Kaiser, Proffitt, & McCloskey, 1985; Gopnik & Astington, 1988; Karmiloff-Smith, 1992; Moses & Flavell, 1990). For example, Gopnik and Sobel (2000, Study 1) found that 3- and 4-year-olds who watched as one novel object activated a "blicket detector" and a second perceptually identical object did not sometimes later mistakenly claimed that both had activated the detector. Gopnik and Sobel suggested that children in their study made the reasonable assumption that perceptually identical objects had similar causal powers even though they had just seen that they did not in the blicket detector case. They argued that "young children will reinterpret and even misremember events that took place only a moment or so before in the light of their higher level beliefs and expectations" (p. 1221).

In the current study, children in the memory control condition presumably had the reasonable expectation that an object's appearance is a good cue as to what it is (e.g., Bloom, 2000), and this expectation overrode the brief exposure children had to the unexpected label. A mechanistic account might involve something like the following.[2] Hearing E1 refer to the hybrid with an unexpected label primed a weak representation of the named category. This representation was active during the inference task, and children either relied on it or relied on the competing perceptually based representation, depending on how confident they were in their own judgment about the item's category membership (and possibly in how credible they viewed E1 as being [see Jaswal & Malone, 2007]). By the time the naming task took place, the effect of the priming had largely dissipated, often leaving children with the more salient perceptually based representation.

Of course, this does raise the question of what would be needed for children to retain the unexpected labels or, indeed, to retain other kinds of counterintuitive information. The study described here was not designed to address this question, but we expect that if children had been exposed to the unexpected labels more frequently, they might have been better able to remember them. In addi-

[2] We thank an anonymous reviewer for suggesting this possibility.

tion, studies in science education with older children suggest that explicitly noting the discrepancy between the information being conveyed and the children's expectations can facilitate learning (e.g., Chi, 1992).

The study reported here is important because it validates a crucial assumption that has been made by researchers who use the category induction procedure to study the nature of children's categories. When children make a label-based inference about stimuli that are moderately discrepant from their expectations, they seem to do so because they actually believe that label, not just because they are complying with the considerable experimental demands of the task.

Acknowledgments

We thank the children, parents, and teachers who participated in these studies. Thanks also go to Jennifer Bailey, Heather Burns, Lindsay Goldman, and Rachel Sylla for assistance in collecting and coding data as well as to Dave McKercher for helpful comments on a previous draft. O.K.L. was supported by Institute of Education Sciences (IES) Grant R305B040049 to the University of Virginia. The research was supported by National Institute of Child Health and Human Development (NICHD) Grant HD-053403 to V.K.J.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635–645.

Aronsson, K., & Hundeide, K. (2002). Relational rationality and children's interview responses. *Human Development, 45*, 174–186.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*, 403–439.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In *Cognitive models of science*. In R. N. Giere (Ed.). *Minnesota Studies in the Philosophy of Science* (Vol. 15, pp. 129–186). Minneapolis: University of Minnesota Press.

Clark, E. V. (1982). The young word maker: A case study of innovation in the child's lexicon. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. Cambridge, UK: Cambridge University Press.

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.

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.

Dawkins, R. (2004). What use is religion? (Part 1). *Free Inquiry, 24*(4), 13–14.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630–641.

Donaldson, M. (1978). *Children's minds*. New York: Norton.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

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

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

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

Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks: Comments on Sloutsky, Kloos, and Fisher (2007). *Psychological Science, 18*, 554–555.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Development, 59*, 26–37.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: 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 *Speech acts*. In P. Cole & J. L. Morgan (Eds.). *Syntax and Semantics* (Vol. 3, pp. 41–58). New York: Academic Press.

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.

Henderson, A. M. E., & Graham, S. A. (2005). Two-year-olds' appreciation of the shared nature of novel object labels. *Journal of Cognition and Development, 6*, 381–402.

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. (2007). The effect of vocabulary size on toddlers' receptiveness to unexpected testimony about category membership. *Infancy, 12*, 169–187.
Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development, 8*, 263–283.
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.
Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 17*, 757–758.
Kaiser, M. K., Proffitt, D. R., & McCloskey, M. (1985). The development of beliefs about falling objects. *Perception and Psychophysics, 38*, 533–539.
Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.
Kemler Nelson, D. C., Egan, L. C., & Holt, L. (2004). When children ask, "what is it?" what do they want to know about artifacts? *Psychological Science, 15*, 384–389.
Koenig, M., Clement, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.
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.
MacDonald, G., Nail, P. R., & Levy, D. A. (2004). Expanding the scope of the social response context model. *Basic and Applied Social Psychology, 26*, 77–92.
Mervis, C. B., Pani, J. R., & Pani, A. M. (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). New York: Oxford University Press.
Moses, L. J., & Flavell, J. H. (1990). Inferring false beliefs from actions and reactions. *Child Development, 61*, 929–945.
Mugny, G. (1984). Compliance, conversion and the Asch paradigm. *European Journal of Social Psychology, 14*, 353–368.
Nail, P. R., MacDonald, G., & Levy, D. A. (2000). Proposal of a four-dimensional model of social response. *Psychological Bulletin, 126*, 454–470.
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).
O'Neill, D. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development, 67*, 659–677.
Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two- and three-year-olds. *Journal of Child Language, 9*, 597–626.
Reid, T. (1997). *An inquiry into the human mind on the principles of common sense*. University Park: Pennsylvania State University Press. Original work published 1764.
Siegal, M. (1997). *Knowing children: Experiments in conversation and cognition* (2nd ed.). East Sussex, UK: Psychology Press.
Siegal, M., & Surian, L. (2004). Conceptual development and conversational understanding. *Trends in Cognitive Sciences, 8*, 534–538.
Siegal, M., Waters, L. J., & Dinwiddy, L. S. (1988). Misleading children: Causal attributions for inconsistency under repeated questioning. *Journal of Experimental Child Psychology, 45*, 438–456.
Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology, 133*, 166–188.
Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007a). What's beyond looks? Reply to Gelman and Waxman. *Psychological Science, 18*, 556–557.
Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007b). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*, 179–185.
Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology, 39*, 906–912.