Journal of Experimental Child Psychology 144 (2016) 130–151

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

The influence of labels and facts on children's and adults' categorization

Megan Johanson [*,1], Anna Papafragou

University of Delaware, Newark, DE 19716, USA

ARTICLE INFO

Article history:
Received 26 August 2014
Revised 17 November 2015
Available online 28 December 2015

Keywords:
Categorization
Labels
Facts
Language and thought
Conceptual development
Preschool

ABSTRACT

Language has been assumed to influence categorization for both adults and children but the precise role and potency of linguistic labels in category formation remains open. Here we explore how linguistic labels help fit objects into categories when relevant perceptual information is either ambiguous or inconsistent with the labels. We also ask how the effects of labels compare to those of other types of information such as facts. We presented 4-year-old children and adults with tasks in which they had to categorize a perceptually ambiguous natural-kind stimulus with one of two equidistant standards (Exp. 1 and 2) or group an ambiguous natural-kind stimulus into a category with a perceptually dissimilar standard (Exp. 3). Participants had access to labels (e.g., "This one is a lorp/pim"), observable facts (e.g., "This one has a long/short beak"), or unobservable facts (e.g., "This one drinks water/milk") that grouped the ambiguous stimulus with one of the standards. Both children and adults followed label- and fact-driven category boundaries for perceptually ambiguous stimuli (Exp. 1 and 2), and continued to do so even when the labels or facts pointed to perceptually incongruent categories (Exp. 3). These findings suggest a strong causal role for both labels and facts in categorization and have implications about theories of how categorization develops in children.

* Corresponding author.
E-mail address: megan.johanson@gmail.com (M. Johanson).
[1] Current address: Crane Center for Early Childhood Research and Policy, The Ohio State University, Columbus, OH 43220, USA.

http://dx.doi.org/10.1016/j.jecp.2015.11.010
0022-0965/

Introduction

One of the most central processes in human cognition is categorization, that is, the grouping of discriminable properties, objects, or events into classes. Categorization is a complex process relying on a variety of cues, including perceptual information (Imai, Gentner, & Uchida, 1994; Landau, Smith, & Jones, 1998; Oakes & Rakison, 2003; Quinn, Norris, Pasko, Schmader, & Mash, 1999; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002), conceptual information (Booth & Waxman, 2002a,b; Booth, Waxman, & Huang, 2005; Waxman & Namy, 1997), and linguistic information (Casasola & Bhagwat, 2007; Ferry, Hespos, & Waxman, 2010; Fulkerson & Waxman, 2007; Landau & Shipley, 2001; Plunkett, Hu, & Cohen, 2008). In this study, we were interested in determining the circumstances under which linguistic cues (*labels*) and conceptual cues (*facts*) can influence categorization performance in both adults and young children.

Role of labels in categorization

Even though infants can form categories in the absence of linguistic labels (Casasola, Cohen, & Chiarello, 2003; Feigenson, Carey, & Spelke, 2002; Hespos & Baillargeon, 2001; Quinn, 2004; Quinn, Eimas, & Rosenkrantz, 1993; Spelke, 1990; Waxman & Markow, 1995), it is well known that the presence of labels can be helpful to young learners' categorization by drawing attention to shared features, relations, or actions. For example, applying the same label to different objects can make the similarities of these objects more salient to learners (Balaban & Waxman, 1997; Booth & Waxman, 2002b; Fulkerson & Waxman, 2007; Waxman, 1999; Waxman & Markow, 1995; Welder & Graham, 2006; cf. Nazzi & Gopnik, 2001; Oakes & Rakison, 2003). Moreover, young children expect that items that share a label should also share some perceptual features and that, conversely, items that look similar should share a label (Gershkoff-Stowe & Smith, 2004; Jones, Smith, & Landau, 1991; Waxman & Markow, 1995; Yamauchi & Markman, 2000; Yoshida & Smith, 2003; cf. Buresh & Woodward, 2007). Furthermore, the presence of a novel label in a match-to-sample task promotes categorical choices in 3- to 5-year-olds, as opposed to thematic choices (Markman & Hutchinson, 1988). These and related studies consistently support the idea that labels specify category membership even for young children (Gelman, 2003; Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Waxman, 2003; Yamauchi & Markman, 2000). On this view, "exactly what makes a dog a dog, or a lamb a lamb, may be unknown (and unknowable; Gelman, 2003), but a category label can serve as a placeholder that a reason exists" (Jaswal & Markman, 2007, p. 96). In short, labels may function as "invitations to form categories" (Waxman & Markow, 1995).

In most of the work documenting the effects of labels in categorization, labels were applied to perceptually congruent categories. To isolate the role of labels more clearly, and to show that labels can reveal deeper underlying conceptual similarities, it is important to look at cases where labels do not work in concert with perceptual information during categorization. Relevant work has shown that applying one label to two narrow perceptual categories helps infants and preschoolers to form a single broad category (Landau & Shipley, 2001; Plunkett et al., 2008), and applying two labels randomly to members of two perceptual categories disrupts category formation entirely (Plunkett et al., 2008). A more stringent test of the potency of labels would be given by a stimulus that is perceptually equidistant from two other stimuli belonging to different categories (and, thus, is perceptually "neutral"); applying the same label to this stimulus and to one of the foils would show whether labeling could affect the placement of category boundaries in the absence of perceptual support (cf. Carmichael, Hogan, & Walter, 1932). In addition, one could ask whether a single basic-level label, applied to stimuli that are perceptually dissimilar to each other, could lead to the creation of a unified (albeit perceptually incongruent) category. Empirical evidence from such tests, however, remains a topic of debate.

Early studies of children's inductive inferences about natural kinds (i.e., animals, insects, plants, and other kinds of entities found in nature; Gelman & Markman, 1986, 1987) showed that labels were used more reliably as markers of category membership than perceptual similarity when the two types of information conflicted, thereby supporting the hypothesis that labels point to a rich underlying category structure. In one such study, 4-year-olds saw two objects (e.g., a tropical fish and a dolphin) and heard both a label and a piece of information about each object (Gelman & Markman, 1986, Study 1). Then, a new object was presented that shared a label with one of the original objects but looked more similar to the other object (e.g., a shark that looked similar to the dolphin but was labeled a fish). Children were asked which piece of information from the original objects also applied to the new object. In this task, children extended the factual information from the object with the same label but different appearance 68% of the time (a proportion significantly different from chance), compared with 88% of the time when the test object shared both the same label and a similar appearance with one of the original objects. In an extension of this work, Gelman and Coley (1990) found that even 2-year-olds could look beyond salient perceptual features and use labels to identify category membership in an induction task.

More recent studies, however, have raised the possibility that earlier work (e.g., Gelman & Markman, 1986, 1987) may have overestimated the role of labels compared with perceptual appearances by presenting children with familiar objects and labels, thereby allowing them to draw on their existing knowledge about the structure of these categories. Davidson and Gelman (1990, Experiment 1) presented 4- and 5-year-olds with an induction task in which they viewed line drawings of novel animals. Children were given a fact about a target animal (e.g., "This zav has four stomachs") and were asked whether the fact also applied to a series of test animals that varied in perceptual similarity to the target. Children extended the fact to the perceptually similar animals regardless of whether they shared a label with the target animal ($M = .75$) or not ($M = .72$). Children rarely extended the fact to perceptually dissimilar animals even when labels matched ($M = .49$ for both cases). Thus, labels did not override perceptual incongruence for newly encountered stimuli but appeared to mitigate somewhat the categorization preferences created by perceptual factors (cf. Davidson & Gelman, 1990, Experiment 3; Gelman & O'Reilly, 1988).

In a similar study that used more stringent controls for perceptual similarity, Sloutsky, Lo, and Fisher (2001) found that 4- and 5-year-old children were more willing to accept that perceptually dissimilar animals shared the same underlying characteristic (e.g., "has yellow blood") if they also shared the same label ("a Bala"), compared with having different labels ("a Bala" vs. "a Guga"). Importantly, labels were found to be most influential when the two animals that formed the possible induction foils were highly perceptually similar to each other (and to the target), presumably because the labels were the only distinguishing feature. When one of the foils became perceptually very dissimilar to the target, children's willingness to extend the shared characteristic to that foil fell to chance level even when the foil and target shared the same label. In those cases, the shared label did increase the likelihood of making an inductive inference from the target to the perceptually mismatching foil relative to a no-label presentation, but the label did not trump perceptual similarity. By contrast, 11- and 12-year-olds and adults completely disregarded perceptual similarity in favor of labels when deciding which animals shared underlying characteristics (a third group consisting of 7- and 8-year-olds appeared to be a transitional group). Corroborating these findings, Deng and Sloutsky (2012) found that 4- and 5-year-olds, unlike adults, made more inferences about novel entities based on a salient perceptual feature (e.g., moving head) than on a category label when the two cues conflicted.

Summarizing the effects of labels on children's categorization, it seems clear that labels can give rise to a broad category that could potentially encompass two narrower perceptual categories in the stimuli (Landau & Shipley, 2001; Plunkett et al., 2008). What is less clear is whether labels might motivate learners to divide items (especially *novel* items) into discrete categories when there is no evidence for categorical structure in the visual stimuli (i.e., when a stimulus is perceptually equidistant from members of existing categories) or when the categorical structure of the visual stimuli is opposite to that suggested by the labels. Furthermore, it is not clear whether there are discontinuities in terms of the way in which adults and young children (especially around 4 years of age) bring labels to bear on categorization. In the experiments reported below, we addressed these issues.

Role of labels versus facts in categorization

A fuller understanding of the effects of labels in categorization requires comparing labels with other factors such as conceptual cues or *facts*. We know that young children are sensitive to (linguistically presented) facts; in one study, such facts ("this was given to me by my uncle", "this came from a place called Koba") were retained over long delays by both adults and 3- and 4-year-olds at least as well as novel labels ("this is called Koba") after a single exposure (Markson & Bloom, 1997). More pertinent, we know that facts can provide cues to category membership. For instance, applying enduring but unobservable properties to animals in an induction task (e.g., "this one helps us take care of sheep") helps 3- and 4-year-olds extend this information to members of the same subordinate-level category (even though not to other taxonomic levels; Waxman, Lynch, Casey, & Baer, 1997).

There are many different types of facts, and some are more informative for categorization than others (e.g., stable facts vs. temporary facts). Waxman and Booth (2000) found that 4-year-old children did not extend facts (e.g., "my uncle gave me this") to new objects in the same way as they extended noun labels (e.g., "this is a blicket") to new members of a category—presumably because there seems to be little reason for extending an incidental fact of possession to new objects (Bloom & Markson, 2001). In contrast, more stable facts about internal characteristics seem to be better candidates to rely on for categorization decisions. For instance, 4-year-olds were found to be significantly less likely to extend incidental less predictive facts (e.g., "eats a cupful/spoonful of food") to another object compared with deeper predictive facts (e.g., "eats plants/meat") (Gelman & Markman, 1986). Similarly, stable traits such as "this X *is* very daxy" were extended to new exemplars more than temporary traits such as "this X *feels* very daxy" (Graham, Welder, & McCrimmon, 2003).

Currently, it is unclear whether facts and labels affect the way in which children (and adults) place category boundaries in similar ways. In the few cases in which facts have been explicitly contrasted with labels, facts generally appear to be less influential than labels. In one study, adults learned to categorize "aliens" faster after learning novel labels for them compared with facts (e.g., where the aliens lived), although both were applied consistently with perceptual information (Lupyan, Rakison, & McClelland, 2007). In another study, 2.5-year-olds in an induction task prioritized labels over surface appearance but ignored temporary facts ("This is sleepy"; Gelman & Coley, 1990). Nevertheless, neither of these studies contrasted labels with stable facts that have been shown to be particularly diagnostic of category membership (see, e.g., Gelman & Markman, 1986; Graham et al., 2003). In a more recent study that used just such facts, facts affected categorization in 5-year-olds and labels were ignored—but the study did not directly compare the two types of cue (Diesendruck & Peretz, 2013). In the experiments that follow, we take up the issue of how labels compare with different types of fact in shaping early categorization. Unlike past studies, we systematically compare children's and adults' use of facts (as well as labels) during categorization. We were especially interested in comparing the potency of facts and labels for category formation in situations where perceptual cues to category structure are ambiguous or misleading.

Goals and prospectus

In the following experiments, we investigated 4-year-old children's and adults' categorization preferences to explore the influence of labels and facts on categorization. We chose 4-year-olds because there is disagreement in the literature about whether children at this age consult labels in determining category membership, as adults typically do (e.g., Davidson & Gelman, 1990; Sloutsky et al., 2001). Recall that Gelman and Markman (1986) found that adults and 4-year-olds showed similar patterns of relying on category labels versus perceptual similarity when required to extend a characteristic to a novel exemplar of a natural kind. However, Deng and Sloutsky (2012) found that 4-year-olds were more likely to make category judgments based on perceptual features, and adults were more likely to use available labels.

We focused on natural kinds because of the importance of these categories for children's cognitive development and their prominent role in past work on categorization. Natural kinds are known to be perceptually more homogeneous compared with other categories (e.g., artifacts; Malt, Sloman, Gennari, Shi, & Wang, 1999), so they were a good test bed for comparing the potency of perceptual factors with that of labels and facts during categorization. More specifically, we focused on unfamiliar exemplars of natural kinds because they have been the source of divergent results in children's categorization (see previous sections). In all studies, we manipulated perceptual similarity systematically by morphing together two different, novel natural-kind stimuli (e.g., two birds) to create a continuum of stimuli at 10% intervals, representing various degrees of perceptual similarity to each of the original stimuli ("standards"). This allowed us to explore the categorization not only of stimuli with clear perceptual closeness to the standards but also of a stimulus that was completely perceptually ambiguous between the two standards. We asked whether labels and facts might motivate children (and adults) to group such a perceptually ambiguous stimulus with one of the standards in a label- or fact-consistent way (Experiments 1 and 2). We also asked whether labels and facts could motivate participants to group stimuli in ways opposite to perceptual similarity (Experiment 3).

Throughout our studies, we compared novel labels (e.g., "This one is a lorp/pim"), observable facts (e.g., "This one has a long/short beak"), and unobservable facts (e.g., "This one drinks water/milk") for these exemplars. We chose novel labels because they have been found to be comparable to known words without carrying information about familiar categories (see also Casasola & Bhagwat, 2007; Deng & Sloutsky, 2012; Markman & Hutchinson, 1988; Waxman & Markow, 1995). We chose observable and unobservable facts because both types captured plausible, stable, biological traits of the stimuli that could be indicative of categorical structure, at least for adults (cf. Gelman & Markman, 1986; Graham et al., 2003). Observable facts provided an enduring reference point against which participants could compare the different stimuli; they always referred to true physical features of exemplars in the display (e.g., a bird's short or long beak). These physical features involved gradable adjectives, so that the degree to which they could be said to apply to a stimulus could vary. Unobservable facts shared features with both observable facts and labels. Like observable facts, the unobservable facts provided stable biological properties of the ambiguous stimuli. Like labels, the unobservable facts were never tied to physical features but rather were tied to unseen behaviors or properties (e.g., diet, habitat) of the stimuli.

Of interest was, first, whether 4-year-old children and adults could use labels for categorization in ways that override ambiguous or inconsistent perceptual input or whether there would be developmental differences in the use of labels. Also of interest was whether the behavior of labels would generalize to conceptual cues to categorization such as facts. One possibility is that across age groups, participants would use both observable and unobservable facts to override perceptually neutral or inconsistent information (even though the degree to which they would rely on facts compared with labels remains open). An alternative possibility is that adults would use facts to categorize even in the absence of congruent perceptual input but that children would show some resistance toward using facts to override perceptual input during categorization. A particularly interesting variant of this outcome is a pattern of results in which one type of fact would affect categorization more heavily compared with the other. For instance, observable facts could be more accessible to children compared with unobservable facts (and labels) and could be used earlier to override perceptual ambiguity or incongruence.

Experiment 1: do labels and facts override perceptual ambiguity in categorization?

Method

Participants

Participants consisted of 64 children between the ages of 4;0 (years;months) and 5;3 (mean age = 4;6) and 48 adults. The children were recruited from local day care facilities in Newark, Delaware, in the eastern United States. The adults were recruited from introductory psychology classes at the University of Delaware (UD) and received course credit for participating.

Stimuli

Four sets of black-and-white morphed objects were created using FantaMorph (a commercial morphing program): two different sets of flowers, one set of birds, and one set of fish. Each set had two distinct standards that were morphed into five objects ("targets"). Each target was 10%, 30%, 50%, 70%, or 90% like one of the standards (see Fig. 1 for a sample set). The 10% and 30% targets were always more perceptually similar to Standard 1 than to Standard 2, and the 70% and 90% targets were always more perceptually similar to Standard 2 than to Standard 1. The 50% target was perceptually equidistant from Standards 1 and 2.

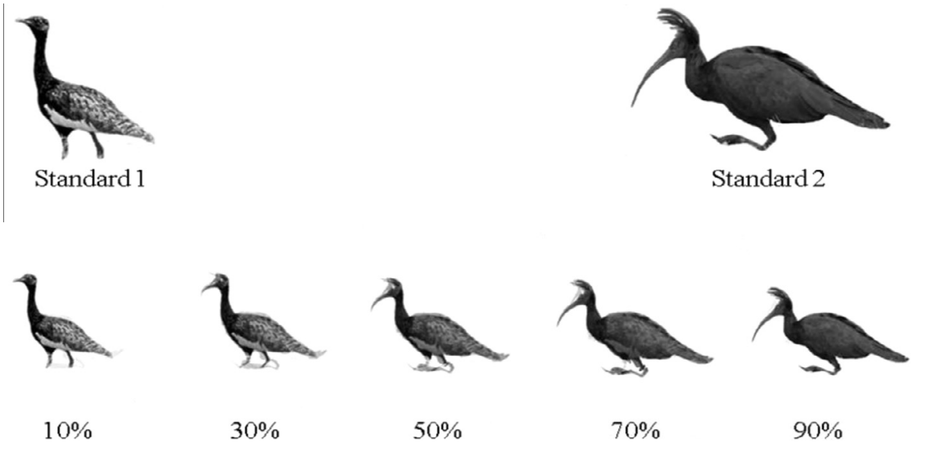

Fig. 1. Sample set of standards and morphed objects (targets) from Experiment 1.

To ensure that the targets could be perceptually grouped in accordance with the FantaMorph similarity ratings, a separate group of 18 adults (all undergraduate students at UD recruited in the same way as the main sample) rated the perceptual similarity of all targets to the standards. Triads were created for each morphed object with the two standards on top and a target below them (see Fig. 2). Participants were tested in small groups and were given an answer sheet to record their ratings. Stimuli were presented on a projector screen. Participants viewed all possible triads in one of two randomized orders and were asked to mark where the target fell on a 9-point scale from 10% to 90%, with Standard 1 at 0% and Standard 2 at 100%.

The ratings were averaged across adults for targets at each 10% interval (as determined by FantaMorph) and were subtracted from the expected rating, giving us a difference score. See Table 1 for mean ratings and standard deviations, and mean difference scores at each 10% interval, collapsed across sets. As shown in Table 1, participants were very accurate in determining the correct rating for the targets. We compared each mean rating score with the expected rating using a *t*-test. The only cases where the mean ratings were different from the expected ratings were the 10% target, $t(17) = 3.37$, $p = .004$ (all *t*-tests two-tailed), and the 90% target, $t(17) = -3.98$, $p = .001$, but even so these objects were rated as unambiguously closer to the appropriate standard. Overall, the ratings of the morphed objects were judged to be close to the FantaMorph ratings and to be appropriate stimuli for the following experiments.

Furthermore, a separate group of 10 adults (all undergraduate students at UD recruited in the same way as the main sample) were asked to name each standard. This was done to ensure that the stimuli would be unfamiliar to participants. Results revealed that each standard was labeled consistently by at most 4 participants. Even where such limited agreement was observed, the label was either a basic category label that could not distinguish between the two standards within a set (e.g., *bird* in Fig. 1) or an incorrect label (e.g., *sunflower* for one of the flowers). Therefore, we conclude that the standards (and, a fortiori, the targets derived from them) were truly novel stimuli.

Procedure

Children were tested individually in front of a computer screen displaying the stimuli. Adults were tested in small groups of 4 to 6. They watched the stimuli displayed on a larger screen and entered their responses on an answer sheet.

Participants were randomly assigned to one of four conditions: Label, Observable Fact, Unobservable Fact, or No Cue. Within each condition, participants viewed four stimuli sets. Each set consisted of five trials, and each trial consisted of a triad of the two standards for the set and one target

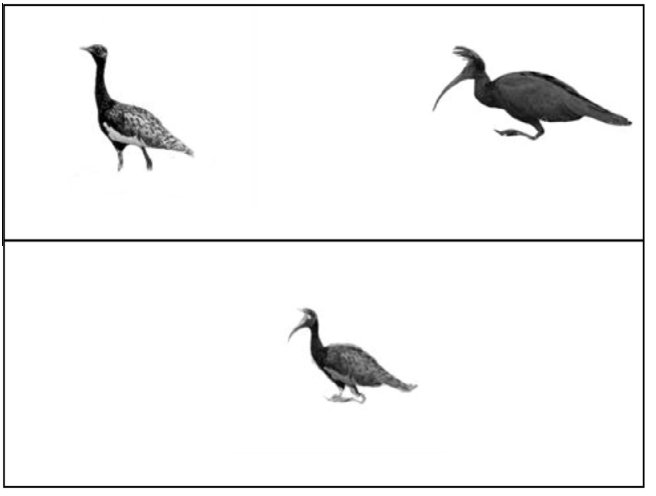

Fig. 2. Sample 50% trial from the bird set in Experiment 1.

Table 1
Mean rating (and standard deviation) and mean difference collapsed across sets for each 10% interval.

Expected rating	Mean rating	Mean difference
10	11.70 (4.44)	−1.70
20	19.58 (11.68)	0.42
30	28.19 (13.87)	1.81
40	39.03 (11.65)	0.97
50	49.17 (12.19)	0.83
60	58.33 (12.10)	1.67
70	68.06 (13.80)	1.94
80	79.03 (11.28)	0.97
90	84.03 (10.16)	5.97

(the 10%, 30%, 50%, 70%, or 90% target; see Fig. 2). All five trials for each set were presented sequentially. After all trials within a set were completed, the experimenter moved on to the next set.

In the Label condition, participants viewed each trial and heard novel labels given to each of the standards and the target. For example, in one trial of the bird set (see Fig. 2), the experimenter pointed to Standard 1 and said, "Look! This one is called a lorp!" and then pointed to Standard 2 and said, "Look! This one is called a pim!" Next, a target from that set appeared at the bottom of the screen and was given a label by the experimenter that grouped it with one of the two standards (see Fig. 2). For instance, the experimenter said, "Look [pointing to the target]! This one is another lorp/pim!" For the unambiguous trials (10%, 30%, 70%, and 90%) within each set, the label for the target always lined up with the perceptually matching standard. For the ambiguous trial (50%), the target label matched Standard 1 half of the time and matched Standard 2 the other half of the time. Immediately after presenting the label for the target, the experimenter pressed a key and made the target disappear. As the target disappeared, the experimenter said, "Oops! It went away!" After approximately 2 seconds, the experimenter pressed another key making the target reappear and said, "Here it is again! Which one of these [pointing to each of the standards] does it go with?" Although previous studies did not involve a delay between providing information and making a categorization decision, we removed the stimuli from the screen to discourage the children from continued looking at the relevant stimuli after the experimenter spoke and to encourage them to process the information more deeply.

In the Observable Fact condition, the procedure was the same as in the Label condition except that the experimenter gave an observable fact about each standard and target. For instance, in the bird trial described above, the experimenter pointed to Standard 1 and said, "Look! This one has a short beak!" Next, the experimenter pointed to Standard 2 and said, "Look! This one has a long beak!" When the target appeared, the experimenter said, "Look [pointing to the Target]! This is another one with a short/long beak!"

The procedure in the Unobservable Fact condition was similar but with the introduction of unobservable facts for each standard and target. For example, in the bird trial, the experimenter pointed to Standard 1 and said, "Look! This one drinks milk!" Next, the experimenter pointed to Standard 2 and said, "Look! This one drinks water!" When the target appeared, the experimenter said, "Look [pointing to the Target]! This is another one that drinks milk/water!"

In the No Cue condition, the procedure was again the same except that the experimenter offered no information about either the standards or the targets. Instead, the experimenter simply pointed to each object and said, "Look at this one!" See Table 2 for a summary of the presentation format for all four conditions.

In all conditions, the trials in each set were presented in a pseudorandomized order such that the ambiguous trial was always third, but the position of the unambiguous trials varied randomly for each set. For half of the participants, we created a different list in which the left–right position of the standards for each trial was switched. Finally, we created a reversed version of each of those two lists for a total of four lists. Participants were randomly assigned to one of the lists.

The pairs of novel words used in the Label condition were *lorp/pim* (bird set), *hep/moof* (fish set), *dax/blick* (first flower set), and *fliff/sned* (second flower set). The pairs of facts used in the Observable Fact condition were "has a long/short beak" (bird set), "has black/white stripes" (fish set), "has light/dark petals" (first flower set), and "has full/skinny petals" (second flower set). The pairs of facts used in the Unobservable Fact condition were "drinks milk/water" (bird set), "lives in the river/sea" (fish set), "smells good/bad" (first flower set), and "grows in the dark/sunlight" (second flower set).

We expected that children and adults would consistently match the unambiguous targets with the perceptually matching standard because of the shared perceptual features (No Cue condition) or the combination of the shared perceptual features and the additional information (Label, Observable Fact, and Unobservable Fact conditions). Of interest was how the participants would treat the ambiguous (50%) target. Because there was no perceptual information or additional cue to rely on, we expected performance to be at chance for both adults and children in the No Cue condition. However, in the Label, Observable Fact, and Unobservable Fact conditions, we expected that adults would use the labels and facts to make their grouping decision. For example, if an ambiguous object received the same label or fact as one of the standards, it should be grouped with that standard. Of interest was whether children would be able to take advantage of the labels and facts for the purposes of this task or whether they would be less consistent than adults in their use of labels for categorization.

After viewing all five test trials in a set, the participants would see the ambiguous target for that set again and hear, "This is the mystery flower/bird/fish! We saw this one before." Then participants were asked, "Which one does it go with?" (*test repeat* trial) and "Do you remember what we told you about this flower/bird/fish?" (*cue recall* trial). Participants in the No Cue condition were asked only the first question. The purpose of these questions was to ascertain whether participants retained information from the main phase of the experiment. The more participants could retain the grouping information over this short delay, the more confident we could be that they were actually forming categories that could survive beyond individual trials. Because the ambiguous target was always presented third within each set and each set consisted of five trials, two test trials always intervened between the main 50% trial and these additional questions.

Coding

Correct responses in the Label, Observable Fact, and Unobservable Fact conditions were those that conformed to either the combination of perceptual similarity and the label/fact given (unambiguous trials) or the label/fact alone (ambiguous trials). For the No Cue condition, correct responses for the unambiguous trials were those that conformed to perceptual similarity; for the ambiguous trials, correct responses were chosen in accordance with what would be the correct answer in the previous

Table 2
Sample presentation format for all conditions in Experiment 1.

	Experimenter points to Standard 1	Experimenter points to Standard 2	Target appears	Target disappears	Target reappears
Label	"Look! This one is called a lorp!"	"Look! This one is called a pim!"	"Look! This is another lorp/pim!"	"Oops! It went away!"	"Which one of these does this one go with?"
Observable Fact	"Look! This one has a short beak!"	"Look! This one has a long beak!"	"Look! This is another one with a short/long beak!"	"Oops! It went away!"	"Which one of these does this one go with?"
Unobservable Fact	"Look! This one drinks milk!"	"Look! This one drinks water!"	"Look! This is another one that drinks milk/water!"	"Oops! It went away!"	"Which one of these does this one go with?"
No Cue	"Look at this one!"	"Look at this one!"	"Look at this one!"	"Oops! It went away!"	"Which one of these does this one go with?"

three conditions (essentially, these corresponded arbitrarily to Standard 1 half of the time and to Standard 2 the other half of the time). For the cue recall trials, if a participant incorrectly recalled or failed to recall the cue, it was counted as zero for that item.

Results

Unambiguous trials

We looked at the percentages of correct responses for the unambiguous trials (10%, 30%, 70%, and 90%). The results were all at or near ceiling for both children ($M_{\text{Label}} = .96$, $M_{\text{Obs\_Fact}} = .98$, $M_{\text{Unobs\_Fact}} = .95$, $M_{\text{No\_Cue}} = .95$) and adults ($M_{\text{Label}} = 1.00$, $M_{\text{Obs\_Fact}} = 1.00$, $M_{\text{Unobs\_Fact}} = .99$, $\text{M}_{\text{No\_Cue}} = .98$), demonstrating that participants could successfully categorize unambiguous objects.

Ambiguous trials (main phase)

Next, we looked at the percentage of correct responses on the ambiguous (50%) trials. We conducted an analysis of variance (ANOVA) using the percentage of correct responses as the dependent variable and age group (adults or children) and condition (Label, Observable Fact, Unobservable Fact, or No Cue) as between-participants factors (see Fig. 3). The ANOVA revealed a significant effect of age group, $F(1, 104) = 6.24$, $p = .01$, $\eta_p^2 = .06$ ($M_{\text{adults}} = .81$ and $M_{\text{children}} = .71$), and condition, $F(3, 104) = 22.01$, $p < .0001$, $\eta_p^2 = .39$ ($M_{\text{Label}} = .80$, $M_{\text{Obs\_Fact}} = .88$, $M_{\text{Unobs\_Fact}} = .86$, and $M_{\text{No\_Cue}} = .48$). The condition effect was caused by a significant difference between performance in the No Cue condition and performance in the Label condition, $F(1, 54) = 22.62$, $p < .0001$, $d = 1.28$, the Observable Fact condition, $F(1, 54) = 45.69$, $p < .0001$, $d = 1.81$, and the Unobservable Fact condition, $F(1, 54) = 46.33$, $p < .0001$, $d = 1.82$ (no other comparisons were significant). There was also a marginal age group by condition interaction, $F(3, 104) = 2.499$, $p = .06$, $\eta_p^2 = .07$.

As expected, performance in the No Cue condition did not differ from chance for either adults or children (*ps* > .05). Additional *t*-tests showed that both adults and children performed differently from chance (.50) in all other conditions (adults: Label, $t(11) = 13.40$, $p < .0001$; Observable Fact, $t(11) = 6.50$, $p < .0001$; and Unobservable Fact, $t(11) = 8.86$, $p < .0001$; children: Label, $t(15) = 2.55$, $p = .02$; Observable Fact, $t(15) = 6.82$, $p < .0001$; and Unobservable Fact, $t(15) = 6.45$, $p < .0001$). Thus, both adults and children were able to use all three types of cues to adjust category boundaries.

Test repeat and cue recall trials

Recall that at the end of each set, participants viewed the ambiguous target from that set and needed to answer the test question a second time. We conducted an ANOVA using the percentage of correct responses on the test repeat trial as the dependent variable with age group (adults or

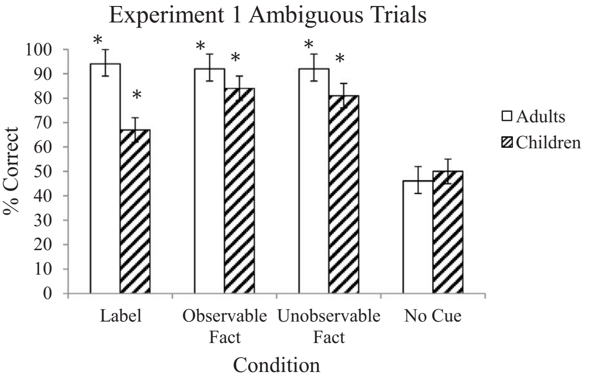

Fig. 3. Percentages of correct responses for the ambiguous trials in Experiment 1. Asterisks (∗) indicate means that are significantly different from chance ($p < .05$). Error bars represent standard errors of the means.

children) and condition (Label, Observable Fact, Unobservable Fact, or No Cue) as independent variables. This analysis returned a significant effect of age group, $F(1, 104) = 14.89$, $p = .0002$, $\eta_p^2 = .12$ ($M_{adults} = .71$ and $M_{children} = .54$), and condition, $F(3, 104) = 4.17$, $p = .008$, $\eta_p^2 = .11$ ($M_{Label} = .61$, $M_{Obs\_Fact} = .73$, $M_{Unobs\_Fact} = .66$, and $M_{No\_Cue} = .51$), but no age group by condition interaction. Further comparisons revealed that the effect of condition was due to a significant difference in performance between the Observable Fact and No Cue conditions, $F(1, 54) = 9.05$, $p = .004$, $d = 0.80$, and a marginal difference in performance between the Unobservable Fact and No Cue conditions, $F(1, 54) = 3.75$, $p = .06$, $d = 0.52$. No other comparisons were significant. Thus, performance in the Label condition after a short delay was not significantly different from performance in the No Cue condition.

In addition, we compared the mean performance in each condition with chance. Whereas performance in the Observable Fact condition, $t(27) = 4.50$, $p < .0001$, and the Unobservable Fact condition, $t(27) = 2.92$, $p = .007$, differed from chance (.50), performance in the Label condition, $t(27) = 1.84$, $p = .08$, *ns*, and the No Cue condition, $t(27) = 0.18$, $p = .86$, *ns*, did not. Whatever the effects of labels on categorization, these effects seem to be lost a few trials after the test trial (unlike effects of observable or unobservable facts).

We also analyzed the ability of participants to correctly recall the cue they were given about each ambiguous target (Label, Observable Fact, or Unobservable Fact). Because there was no cue to be recalled in the No Cue condition, this analysis exludes participants in that condition. Using age group (adults or children) and condition (Label, Observable Fact, or Unobservable Fact) as variables, we ran an ANOVA with the percentage of correct cue recall responses as the dependent variable. The analysis revealed a main effect of age group, $F(1, 78) = 17.77$, $p < .0001$, $\eta_p^2 = .19$ ($M_{adults} = .77$ and $M_{children} = .56$), and a main effect of condition, $F(2, 78) = 7.40$, $p = .0011$, $\eta_p^2 = .16$ ($M_{Label} = .53$, $M_{Obs\_Fact} = .76$, and $M_{Unobs\_Fact} = .69$), but no age group by condition interaction. Performance in the Label condition was significantly lower than performance in either the Observable Fact condition, $F(1, 54) = 11.72$, $p = .0012$, $d = 0.92$, or the Unobservable Fact condition, $F(1, 54) = 4.06$, $p = .05$, $d = 0.54$, but performance did not differ between the Observable and Unobservable Fact conditions. Again, this suggests that labels are more susceptible to fading in memory than either observable or unobservable facts.

Discussion

Our results indicate that labels, observable facts, and unobservable facts help adults and 4-year-old children to adjust category boundaries when perceptual information is indeterminate. However, labels and the categorization preferences they led to were less likely to be retained after a memory delay by both adults and children compared with facts and the categorization decisions made on the basis of factual information. At least for facts, then, the responses given in the main phase of the task were based on genuine categorization commitments that were retained at later phases of testing.[2]

Our current results have two limitations. First, they cannot lead to firm conclusions about whether adults and children treated individual types of cues differently. Recall that there was a marginal interaction between age and condition; inspection of Fig. 3 suggests that there may be a wider gap between children's and adults' performance in the Label condition compared with the other conditions. Further evidence is needed, however, to evaluate this possibility. Second, it remains possible that children sometimes forgot the cues for the target and/or standards during the brief delay between the training and test trials for each item. Labels (and the categorization patterns they gave rise to) seem to be particularly vulnerable to memory decay compared with other cues in both adults and children (cf. the test repeat and cue recall trials). It is currently unknown whether, in children, this decay began earlier in the course of a trial compared with adults and affected categorization performance in the main phase of the experiment. In the next experiment, we used a modified version of Experiment 1 to address these issues.

Experiment 2: further explorations of the role of labels and facts in categorization

In Experiment 2, we replicated Experiment 1 but eliminated the memory demands of the previous study by leaving the target on the screen from training to test. In addition, we asked participants to repeat the cue for each object immediately after it was given to ensure their understanding. Most crucially, we checked participants' memory for each of the cues after they made their categorization decision and analyzed only responses on which memory for cues was accurate.

Method

Participants

Participants were a new group of 30 children between the ages of 3;10 and 5;2 (mean age = 4;6) and 30 adults. The children were recruited from day care facilities in the Newark area. Adult participants were recruited from UD introductory psychology courses and received course credit for participating. One additional child who completely failed memory controls on the critical trials was excluded.

Stimuli

The stimuli were the same as in Experiment 1.

Procedure

We omitted the test repeat and cue recall trials as well as the No Cue condition from Experiment 1. Participants were randomly assigned to one of three conditions: Label, Observable Fact, or Unobservable Fact. The procedure for each condition was the same as in Experiment 1 except for the changes noted below. First, the target never disappeared from the screen between training and test. Second, immediately after the experimenter gave the label or fact for each standard and target, participants were asked to repeat it before the experimenter moved on to the next object on the slide (e.g., the experimenter would point to a standard and say, "This one is called a lorp! Can you say lorp?"). Finally, after participants made a categorization choice within a set, the experimenter pointed to each of the

[2] In further control manipulations, we replicated the Label condition (without the test repeat and cue recall trials) with new groups of 4-year-olds and adults ($n = 12$ per manipulation and age group). Results showed that children's performance with labels in Experiment 1 was not tied to the specific test question or to the specific novel labels chosen. For instance, when a different test question was used ("Can you find *another one* of the same kind?"; see Waxman & Namy, 1997), results in the ambiguous trials remained the same as in the original Label condition ($M_{adults} = .95$ and $M_{children} = .68$). Similarly, when participants were introduced to a game in which they needed to help an astronaut learn about a planet called Mars and heard each object being described as a "Martian X" (e.g., Martian robin/Martian duck), performance was unaffected ($M_{adults} = .94$ and $M_{children} = .68$). Individual ANOVAs comparing each of these manipulations with the original Label condition revealed only an effect of age group, $F(1, 48) = 19.64$, $p < .0001$, $\eta_p^2 = .29$, and $F(1, 48) = 22.38$, $p < .0001$, $\eta_p^2 = .32$, respectively, but no effect of manipulation and no interactions.

standards and target again and asked, for example, whether it was "a lorp or a pim" (and similarly for facts). These post-test questions served as memory controls to determine whether participants remembered the labels or facts given by the experimenter earlier.

Results

We excluded any test trials on which participants did not correctly recall the cues for both the standards and the target in the memory control trials. Most of these exclusions occurred in the children's data, particularly in the Label condition. Specifically, for children, exclusions were as follows. For unambiguous trials, we excluded 61 Label trials, 27 Observable Fact trials, and 36 Unobservable Fact trials (out of a total of 480 trials per condition). For ambiguous trials, we excluded 17 Label trials, 7 Observable Fact trials, and 11 Unobservable Fact trials (out of a total of 120 trials per condition). In total, we excluded 8.8% of the child data. Statistical analysis confirmed that condition affected exclusions, $F(2, 27) = 4.72$, $p = .0174$, $\eta_p^2 = .26$, with exclusions overall higher in the Label condition compared with the Observable Fact condition, $F(1, 18) = 9.68$, $p = .006$, $d = 1.39$, but not the Unobservable Fact condition, $F(1, 18) = 4.05$, $p = .0595$, $d = 0.90$.

For adults, exclusions were as follows. For unambiguous trials, we excluded 2 Label trials, 4 Observable Fact trials, and 8 Unobservable Fact trials (out of a total of 480 trials per condition). For ambiguous trials, we excluded 1 Label trial, 3 Observable Fact trials, and 1 Unobservable Fact trial (out of a total of 120 trials per condition). In total, we excluded 1% of the adult data.

Unambiguous trials

We looked at the percentage of correct responses for the unambiguous trials (10%, 30%, 70%, and 90%). The results were all at or near ceiling for both children ($M_{Label} = .97$, $M_{Obs\_Fact} = .99$, and $M_{Unobs\_Fact} = .99$) and adults ($M_{Label} = 1.00$, $M_{Obs\_Fact} = 1.00$, and $M_{Unobs\_Fact} = .99$), demonstrating that participants could successfully categorize unambiguous objects.

Ambiguous trials

We performed an ANOVA using age group and condition as independent variables and correct categorization as the dependent variable. There was no main effect of age group or condition and no age group by condition interaction ($ps > .05$; see Fig. 4). Moreover, each of the groups performed differently from chance (all $ps < .05$).

Next, we compared performance on the ambiguous trials in Experiments 1 and 2 using a 2 (Age Group: adult or child) × 3 (Condition: Label, Observable Fact, or Unobservable Fact) × 2 (Experiment: 1 or 2) ANOVA. This analysis revealed a significant effect of age group, $F(1, 131) = 7.62$, $p = .007$, $\eta_p^2 = .05$ ($M_{adults} = .95$ and $M_{children} = .88$) and a main effect of experiment, $F(1, 131) = 25.94$, $p < .0001$, $\eta_p^2 = .17$ ($M_{Exp\_1} = .85$ and $M_{Exp\_2} = .99$). Crucially, there was also an age group by experiment interaction, $F(2, 131) = 6.83$, $p = .01$, $\eta_p^2 = .05$. The interaction was driven by the fact that there was no effect of age group for Experiment 2 but a significant effect of age group for Experiment 1, $F(1, 82) = 10.47$, $p = .002$, $d = 0.52$ ($M_{adults} = .92$ and $M_{children} = .78$). There was no main effect of condition or interactions with condition.[3]

Discussion

In Experiment 2, both 4-year-old children and adults used labels to place category boundaries when perceptual information offered no determinate evidence about the nature of categories.

[3] If no ambiguous trials are excluded in Experiment 2 (i.e., if memory of cues is not used to filter the data), there is an apparent disadvantage in children's use of labels. An ANOVA with condition and age as factors conducted on all of the ambiguous trial data yielded no main effect of condition, a main effect of age, $F(1, 59) = 7.27$, $p < .0001$, $\eta_p{}^2 = .12$ ($M_{adults} = .97$ and $M_{children} = .86$), and an age by condition interaction, $F(1, 54) = 3.92$, $p = .0258$, $\eta_p{}^2 = .13$; adults and children seemed to differ in their use of labels ($M_{adults} = 1.00$ and $M_{children} = .75$) but not in their use of facts. Analysis of the unambiguous trials without exclusions did not change the results, presumably because children relied less on other cues when perceptual information was unambiguous. These results highlight the importance of memory controls for interpreting the results of a task where children need to remember novel labels and bind them with novel exemplars to succeed in categorization (see also Experiment 1).

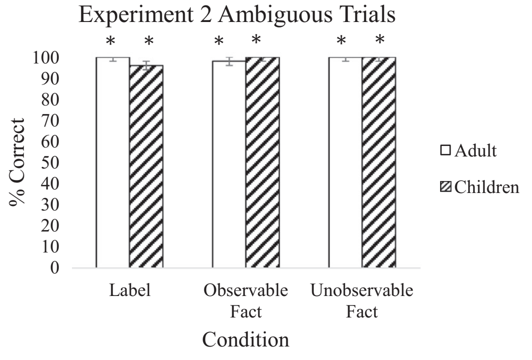

Fig. 4. Percentages of correct responses for the ambiguous trials in Experiment 2. Asterisks (*) indicate means that are significantly different from chance ($p < .05$). Error bars represent standard errors of the means.

Furthermore, the potency of labels was equal to that of observable and unobservable facts in terms of their ability to indicate category membership. Performance in Experiment 2 was near ceiling for 4-year-olds and adults in all conditions; furthermore, performance was significantly higher in Experiment 2 than in Experiment 1. In sum, our data indicate no developmental differences in categorization patterns and no difference between the use of labels and facts as cues for the formation of categories.

Experiment 3: do labels and facts override perceptual ambiguity to establish perceptually incongruent categories?

Experiments 1 and 2 found that labels, observable facts, and unobservable facts can all lead participants to adjust category boundaries for ambiguous stimuli. Experiment 3 asked whether the same cues can help participants to adjust category boundaries in a way that is perceptually incongruous—in other words, whether labels and facts can override perceptual similarity during categorization.

The experiment consisted of a categorization task with stimuli from the same four morphed sets as in Experiments 1 and 2. Similar to our earlier tasks, we used facts and labels as cues and then asked participants which of the standards each target "went with." In critical trials, the cues (labels/facts) indicated that a category could include perceptual matches as well as perceptual mismatches to one of the standards (e.g., the 30% object was given the same label/fact as Standard 2 even though it looked only 30% similar to Standard 2 and 70% similar to Standard 1). We then tested participants on a novel item (i.e., the 50% object). If participants followed the cues from labels/facts and truly formed a category including perceptually mismatching exemplars, then they should also include the novel object in the same category, even though it was perceptually equally similar to Standards 1 and 2.

For this experiment, we omitted the No Cue condition and implicitly relied on the previously reported results of categorization performance for the 50% trials in Experiment 1. We also used four additional sets as fillers where cue-based groupings followed perceptual similarity. Finally, as in Experiment 2, we ensured that participants remembered the cues (labels/facts) they were provided with when making their categorization judgment.

Method

Participants

Participants were 48 children between the ages of 3;10 and 5;0 (mean age = 4;5) and 48 adults. The children were recruited from day care facilities in Newark, and the adults were recruited from psychology classes at UD. We excluded data from 1 additional child who did not pass memory controls for any target trials.

Stimuli

We used the 10%, 30%, 50%, 70%, and 90% targets and the two standards from the morphed sets in Experiments 1 and 2 to create critical sets. We also created four new sets that we used for filler trials. The new sets consisted of two pairs of animals (mouse–monkey and cheetah–wolf) and two pairs of plants (garlic–cherry and corn–pepper). The pairs of novel labels for these filler sets were *gree*/*pag* (mouse–monkey set), *weg*/*biv* (cheetah–wolf set), *zam*/*ched* (garlic–cherry set), and *meb*/*taf* (corn–pepper set). The pairs of observable facts for the filler sets were "has big/small eyes" (mouse–monkey set), "has round/pointy ears" (cheetah–wolf set), "has a thick/thin stem" (garlic–cherry set), and "is bumpy/smooth" (corn–pepper set). Finally, the pairs of unobservable facts for the filler sets were "lives on the ground/in the trees" (mouse–monkey set), "eats plants/meat" (cheetah–wolf set), "tastes sour/sweet" (garlic–cherry set), and "has small/big seeds" (corn–pepper set). The filler sets were created using FantaMorph in a way that was identical to the sets in the previous experiments.

A separate group of 12 adults completed a rating study for the new filler stimuli. Participants viewed all possible triads (Standards 1 and 2 plus a target at each 10% interval from a set; see Fig. 4) and were asked to mark on an answer sheet where the target fell on a continuum from 10% to 90%, with Standard 1 at 0% and Standard 2 at 100%. The ratings were averaged across adults for each target and were subtracted from the expected rating, resulting in a difference score (see Table 3 for all means, standard deviations, and difference scores). Participants were very accurate in identifying the correct rating. When we compared each mean rating score with the expected rating using a *t*-test, the only trials where the two differed were the 10% trial, $t(11) = 3.92$, $p = .002$, the 80% trial, $t(11) = 2.58$, $p = .03$, and the 90% trial, $t(11) = 7.86$, $p < .0001$. Overall, the ratings of the morphed objects were judged to be close enough to the FantaMorph ratings to be appropriate stimuli for the following experiment.

Procedure

The testing setup for children and adults was as in Experiment 1. Participants were randomly assigned to one of three conditions: Label, Observable Fact, or Unobservable Fact. For the critical sets, the specific cues given in each of these conditions were the same as in Experiment 1 (for the additional sets used as fillers, cues are described in the "Stimuli" section above). Each participant saw four critical and four filler sets. Each set consisted of three trials. Trials involved triads of the two standards within each set (presented at the top of the computer screen) and one of the targets presented underneath (as in Experiments 1 and 2). The first two trials in each set were training trials, and the last one was a test trial.

In critical sets, the standards initially appeared alone on the screen and were given distinct labels, observable facts, or unobservable facts by the experimenter (as in Experiments 1 and 2). Next, two different training targets appeared in sequence at the bottom of the screen. Both were perceptually similar to the same standard (e.g., the 70% and 90% targets). In one of the training trials (*congruent* trial), the training target was given the same label/fact with the perceptually matching standard; crucially, in the other (*incongruent*) trial, the next training target was given the same label/fact as the perceptually *mismatching* standard (see Fig. 5). (The order of presentation of congruent and incongruent training trials was counterbalanced across critical sets.) Each of the training targets remained on the screen for approximately 3 seconds, during which time the label/fact was uttered by the experimenter.

For the training trials, immediately after the experimenter gave the label or fact for each standard and target, participants were asked to repeat it before the experimenter moved on to the next object on the slide (e.g., the experimenter would point to a standard and say, "This one is called a lorp! Can you say lorp?"). Afterward, the experimenter pointed to each of the standards and target again and asked, for example, if it was "a lorp or a pim" (and similarly for facts). Children were given feedback if they responded incorrectly. Adults were asked not to verbalize their response but instead to silently think about their answer. After allowing approximately 5 seconds to think about their answer, the experimenter provided adults with the correct answer (e.g., "It's a lorp!").

Immediately following the training trials for each critical set, the test trial occurred. The test trial introduced a novel test target (the 50% object) underneath the two standards (see Fig. 5). All participants were asked the same thing at test: "Which one of these [pointing to each standard] does this one [pointing to the target] go with?" If participants had not integrated the category information given in

Table 3
Mean rating (and standard deviation) and mean difference for each expected rating in the filler trials of Experiment 3, collapsing across the four sets.

Expected rating	Mean rating	Mean difference
10	12.92 (5.91)	−2.92
20	20.42 (15.67)	−0.42
30	27.92 (13.00)	2.08
40	41.04 (17.92)	−1.04
50	48.13 (11.79)	1.87
60	61.46 (13.21)	−1.46
70	69.58 (10.71)	0.42
80	75.94 (9.60)	4.06
90	81.46 (7.36)	8.54

training, then they should be at chance when choosing which standard the test target "went with" because the 50% target was perceptually ambiguous. However, if participants had integrated the category information during training and had used the labels/facts to override perceptual similarity, then the 50% target should be a novel exemplar within a broad category encompassing both perceptually similar and dissimilar exemplars (in our example, the space between the 70% bird exemplar and Standard 1; see Fig. 1). It should, therefore, be grouped with the perceptually mismatching standard (here, Standard 1). After participants made their decision, the experimenter pointed to each of the standards and asked whether it was "a lorp or a pim," for example (memory control). Unlike Experiment 2, we did not include a memory control question for the cue for the training targets because those targets were no longer visible at the end of each trial.

For the filler sets, the procedure was the same but the targets at training were always grouped with the perceptually matching standard (i.e., there were no incongruent trials). This was done to avoid having a conflict between perceptual information and labels/facts throughout the experiment. Participants were presented with the 50% target at test and were asked which standard it "went with." For the filler trials, participants should be at chance at test because the information during training provided no clue about which standard the ambiguous target belonged with.

The order of presentation of critical and filler sets was pseudorandomized, alternating between the two types of sets. For both critical and filler sets, half of the sets used 10% and 30% as the training trials and the other half used 90% and 70%. Two orders were created by counterbalancing which pair of training trials was presented per set (e.g., for the first order, participants would see the 90% target followed by the 70% target in the bird set; in the second order, participants would see the 10% target followed by the 30% target in the same set). Finally, a reverse version of each order was created, resulting in a total of four stimuli lists.

Coding

For the critical sets, the response coded as correct was the one that conformed to the broad category suggested by the experimenter's cue. Specifically, the correct response was to select the standard that matched the label or fact cue given by the experimenter in the incongruent training trial because this cue should have triggered the formation of a broad category including perceptually incongruent stimuli (e.g., 70% or 30%), which would encompass the ambiguous stimulus (50%). For the filler sets, we followed the same logic as in the coding for the ambiguous trials in the No Cue condition of Experiment 1. As a result, half of the time Standard 1 was counted as the correct response and half of the time Standard 2 was counted as the correct response.

Results

We excluded any test trials on which participants did not correctly pass the memory control questions. All excluded trials belonged to the child data and appeared somewhat higher in the Label condition. For critical sets, we excluded 18 Label trials, 8 Observable Fact trials, and 11 Unobservable Fact

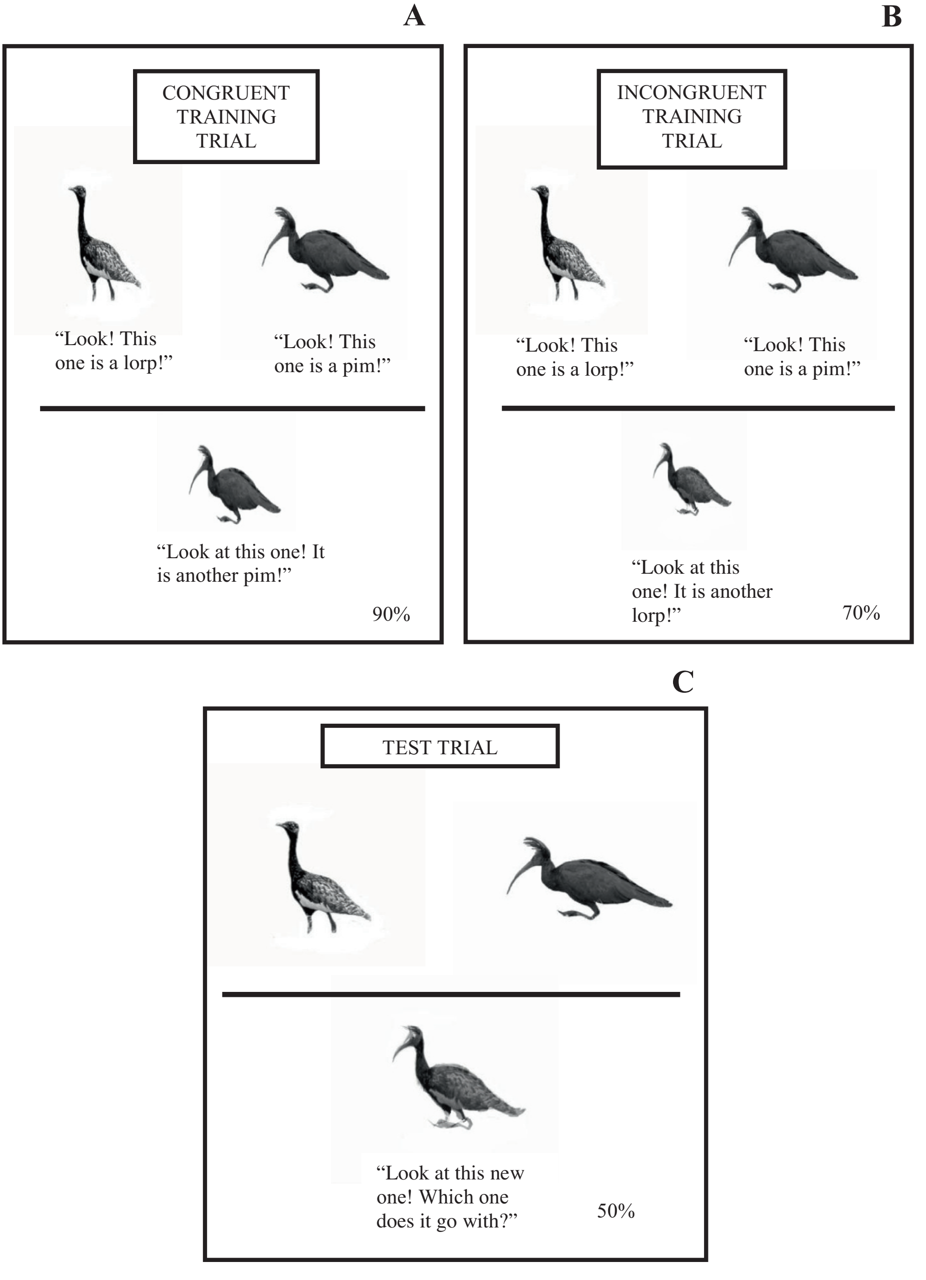

Fig. 5. Set of items in the Label condition of Experiment 3. (A) Congruent training trial. (B) Incongruent training trial. (C) Test trial.

trials (out of a total of 64 trials per condition). For filler trials, we excluded 14 Label trials, 10 Observable Fact trials, and 10 Unobservable Fact trials (out of a total of 64 trials per condition). In total, we excluded 18% of the child data.

Ambiguous trials in filler sets

We conducted an ANOVA using age group (adults or children) and condition (Label, Observable Fact, or Unobservable Fact) as the independent variables and the percentage of correct responses on the filler trials as the dependent variable. The results showed no effect of age group or condition and no age group by condition interaction. Furthermore, we compared performance in each age group and condition with chance ($M = .50$). As expected, children's performance was generally no different from chance ($M_{Label} = .45$ and $M_{Unobs\_Fact} = .52$), with the exception of the Observable Fact condition ($M_{Obs\_Fact} = .33$), $t(15) = -2.42$, $p = .03$. This effect was due to the fact that, for reasons that are not clear, performance in one set for this condition differed from chance ($M_{garlic-cherry} = .17$), $t(11) = -2.97$, $p = .01$, but in the other three sets it did not ($M_{corn-pepper} = .43$, $M_{mouse-monkey} = .50$, and $M_{cheetah-wolf} = .29$, all $ps > .05$). Adult performance was also generally no different from chance ($M_{Label} = .55$ and $M_{Obs\_Fact} = .42$) except for a marginal difference in the Unobservable Fact condition ($M_{Unobs\_Fact} = .38$), $t(15) = -2.07$, $p = .056$.

Ambiguous trials in critical sets

We conducted an ANOVA using age group (adults or children) and condition (Label, Observable Fact, or Unobservable Fact) as the independent variables and the percentage of correct responses on the critical trials as the dependent variable (see Fig. 6). The results showed no effect of age group or condition and no age group by condition interaction ($Fs < .64$, $ps > .05$, *ns*). Children differed from chance (.50) in all conditions (Label: $t(15) = 3.01$, $p = .008$; Observable Fact: $t(15) = 3.45$, $p = .004$; and Unobservable Fact: $t(15) = 3.14$, $p = .007$). Similarly, adult performance was systematically different from chance (Label: $t(15) = 2.66$, $p = .02$; Observable Fact: $t(15) = 7.46$, $p < .0001$; Unobservable Fact: $t(15) = 2.41$, $p = .03$).

Discussion

The results revealed that both children and adults could use labels, observable facts, and unobservable facts to establish a novel perceptually heterogeneous category. Children and adults were equally good at using these different types of cues to create a perceptually incongruent category. As in Experiment 2, after the proper memory controls were applied, labels were neither more nor less influential than facts. These data go beyond previous work to show that both labels and factual cues can override perceptual congruency for 4-year-olds and adults. Our data also show that when labels and facts conflict with perceptual information, neither 4-year-olds nor adults fully rely on them to determine categorization given that performance is not at ceiling (cf. Davidson & Gelman, 1990; Plunkett et al., 2008; Sloutsky et al., 2001). We expand on the theoretical significance of these findings in the General Discussion below.

General discussion

Role of labels in categorization

A first major goal of the current studies was to examine the influence of labels on 4-year-old children's and adults' categorization when perceptual information provided no category information (Experiments 1 and 2) or provided category information that conflicted with labels (Experiment 3). The results of our studies show that, in the absence of determinate perceptual evidence (i.e., when stimuli are perceptually ambiguous), labels can influence how children and adults place category boundaries for unfamiliar natural kinds (Experiments 1 and 2). Moreover, when labels point to perceptually incongruent categories, both adults and children follow the labels in forming novel categories (Experiment 3). Overall, our studies revealed no differences between the role of labels and facts in categorization or between the use of different cues in categorization in 4-year-olds and adults (with the exception of a marginal effect in Experiment 1 that disappeared in later studies that included memory controls).

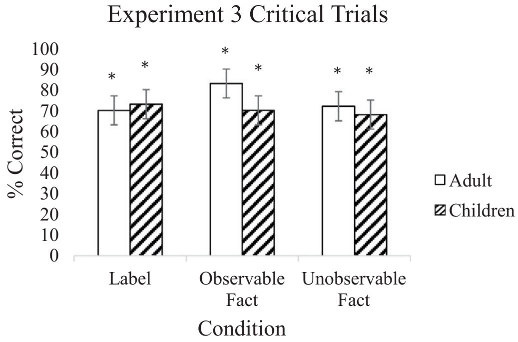

Fig. 6. Percentages of correct responses for the critical trials in Experiment 3. Asterisks (*) indicate means that are significantly different from chance ($p < .05$). Error bars represent standard errors of the means.

These results break new ground by expanding the empirical evidence about the role of labels in categorization. Prior work had shown that labels help children to establish a broad category including two narrower perceptual categories (Landau & Shipley, 2001; Plunkett et al., 2008). The current data demonstrate an additional striking effect, namely that labels motivate learners to divide items into discrete categories even when (a) there is no evidence for categorical structure in the visual stimuli (i.e., when the stimuli are perceptually ambiguous) or (b) the categorical structure of the visual stimuli is opposite to that suggested by the labels. In other words, our data show that labels trigger category formation (Waxman & Markow, 1995) *even when labeling is not obviously supported by the perceptual structure in the stimuli.*

Our findings offer novel evidence with respect to the deeply debated question of whether there is a developmental shift in the use of labels to draw category boundaries. Prior work that evaluated the role of labels in the way in which preschoolers categorize natural kinds had produced mixed results, especially when labels were used to refer to perceptually incongruent categories (Davidson & Gelman, 1990; Deng & Sloutsky, 2012; Gelman & Markman, 1986, 1987; Sloutsky et al., 2001; see also below). The same studies showed that adults consistently prioritized labels when presented with novel natural kinds whose appearance conflicted with the label cues (Deng & Sloutsky, 2012; Sloutsky et al., 2001). Our results show that both adults and 4-year-olds use labels to infer category structure for newly presented exemplars of natural kinds, even in the absence of supporting perceptual information, and do so to the same degree. Thus, our data support the presence of strong continuities in the basic categorization mechanism used by adults and children as young as 4 years (even though they do not preclude other types of developmental change in children's categorization strategies, especially at younger ages; see, e.g., Landau et al., 1988; Subrahmanyam, Landau, & Gelman, 1999).

The strong effects of labels in our data are particularly noteworthy because, in some sense, our categorization task offers a more stringent test for the role of labels in category formation in the absence of supporting perceptual evidence compared with some past work. Recall that, in prior induction tasks, preschoolers were typically presented with objects that were assigned distinct labels and distinct facts. Then, an additional object was introduced that was perceptually more similar to one of the original objects but shared a label with the other perceptually distant foil. Of interest was determining which of the two facts preschoolers would extend to the third object and whether in doing so they would prioritize labels or appearances (see Davidson & Gelman, 1990; Gelman & Markman, 1986, 1987; Sloutsky et al., 2001). A common feature of these studies is that label assignment in the original objects was always accompanied by fact assignment; that is, each label was assigned to an object with a distinct appearance and a distinct characteristic (given by the fact). Thus, in establishing category membership, labels did not function only in combination with the perceptual appearance of the stimuli but were also combined with a fact about the object. By contrast, in our task, the cues given to children were sparser; throughout our studies, in the Label condition, children had access only to the perceptual appearance of the standards and the label for each one (at test, children were simply asked which of the standards the target "went with"). As a result, our method provides a purer contrast between the role of labels and appearances in the formation of categories.

The strong effects of labels are also noteworthy because our stimuli involved both novel natural kinds and novel labels. Recall that in earlier work that pitted labels against perceptual cues in the categorization of natural kinds, labels had a strong role in shaping preschoolers' categorization of familiar exemplars (Gelman & Markman, 1986, 1987) but did not consistently drive their categorization of novel exemplars (e.g., Davidson & Gelman, 1990; Deng & Sloutsky, 2012; Sloutsky et al., 2001).[4] One might argue that our stimuli, even though exemplars of unfamiliar subordinate-level categories, nevertheless belonged to known basic-level categories (e.g., bird, fish).[5] However, none of the past studies that pitted labels against perceptual similarity in categorization used subordinate-level distinctions, and the general consensus from the literature is that children have trouble in reasoning about such distinctions—even for familiar items. For instance, children have difficulty in understanding a novel noun as referring to a subordinate-level item (Waxman, Shipley, & Shepperson, 1991). Similarly, 4-year-olds do not readily distinguish between objects at the subordinate level when asked to label a wide variety of objects (Waxman, 1990), nor do they choose the subordinate-level label as the best label when given the basic-level alternative (Blewitt, Golinkoff, & Alioto, 2000; cf. Graham et al., 2003). Most relevant for current purposes, preschoolers find novel nouns at the basic and superordinate levels to be more helpful in categorization compared with nouns at the subordinate level (Waxman, 1990). Against the backdrop of these studies, the fact that labels in our experiments led 4-year-olds to draw boundaries for unfamiliar *subordinate-level* categories is all the more striking.

Finally, our results show that effects of labels on adults' and young children's categorization are strongly mediated by the degree of perceptual similarity between the stimuli sharing a common label; for perceptually neutral stimuli both adults and children overwhelmingly followed the label in placing category boundaries (Experiment 2), but for perceptually incongruent stimuli they did so only approximately 70% of the time (Experiment 3) (cf. Davidson & Gelman, 1990; Plunkett et al., 2008; Sloutsky et al., 2001). We propose that labels suggest—but do not directly supply—the presence of a causal relation that grounds the named category in a principled way; furthermore, the causal relation suggested by the label is typically supported by other available cues (e.g., perceptual cues in the case of natural kinds). For instance, children expect that creatures that are called "tiger" have some shared (e.g., biological) features linking them together into a category. As mentioned earlier, our experimental setup gave no additional cues to support the causal role of labels (especially when labels conflicted with appearances as in Experiment 3). We predict that if novel labels in Experiment 3 were accompanied by more information about the causal structure of the labeled category ("This is a dax. Daxes smell nice"), children's (and adults') willingness to rely on labels in categorization would increase even for perceptually incongruent stimuli.

Role of labels versus facts in categorization

A second major goal of our studies was to compare the role of labels with the potency of different types of facts when perceptual information provided no category information (Experiment 1) or provided category information that conflicted with labels (Experiment 3). We found that, just like labels, facts maintain a strong influence over categorization for both children and adults. Specifically, both perceptually grounded/observable facts ("This one has a long beak") and nonperceptual/unobservable facts ("This one drinks water") help children and adults to place category boundaries for perceptually ambiguous novel stimuli (Experiments 1 and 2); this pattern occurs even when the facts group together perceptually dissimilar stimuli (Experiment 3).

[4] Our findings also suggest that a potential (and as of now unidentified) factor contributing to children's apparent challenge in using labels in existing paradigms might be the difficulty of remembering novel labels and binding them with novel exemplars in the context of the categorization task (see especially Footnote 3 in Experiment 2). Because categorization tasks in the literature did not generally include memory controls for novel labels, it remains an open possibility that memory factors affected (at least in part) children's performance in those tasks.

[5] We bypass here the complexities around the issue of how to specify the taxonomic level of certain natural kinds (see Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

The current studies are among the first to address the use of facts in categorization by both young and more experienced (i.e., adult) cognizers. These results make two significant contributions. First, performance with facts offers clear empirical support to the conclusion that young children can go beyond appearances in placing category boundaries (see also Keil, 1986). Second, the finding that both observable and unobservable facts offered a powerful tool in forming categories suggests that both young children and adults were sensitive to the nature of the information conveyed; specifically, they recognized that these facts were biologically inspired and, hence, were stable reliable indicators of category structure (see also Diesendruck & Peretz, 2013; Gelman & Markman, 1986; Graham et al., 2003). Just as with labels, the potency of both observable and unobservable facts was subject to constraints placed by the degree of perceptual similarity among the exemplars sharing the same facts; both adults and 4-year-olds consistently followed the facts in placing category boundaries for perceptually ambiguous stimuli (Experiment 2) but showed reduced commitment to the facts for perceptually incongruent stimuli (Experiment 3).

A particularly striking—and novel—feature of our data is the fact that labels and facts had very similar effects on children's and adults' categorization. The equal potency of labels and facts might at first appear to be surprising from the position that labels have a privileged status in categorization for both children (Gelman, 2003; Gelman & Markman, 1986, 1987; Markman, 1989; Waxman, 2003) and adults (Yamauchi & Markman, 2000). We see no inconsistency here, however. First, prior literature arguing for the special status of labels did not compare labels with stable biologically inspired facts of the sort used in the current studies. Second, if one assumes (together with the researchers above) that children and adults share richly structured conceptual representations of categories, it should come as no surprise that both labels and biologically based conceptual information are strongly predictive of category structure (and continue to be so even in the face of indeterminate or inconsistent perceptual information). As mentioned earlier, labels suggest the presence of a causal relation that grounds the named category in a principled way—but do not directly provide the causal relation themselves. Biologically inspired facts, on the other hand, supply a specific piece of the causal nexus that structures human categories. In the current studies, the unobservable facts that we used captured signature features of natural kinds such as their diet ("drinks milk"), habitat ("lives in the river"), and other biological properties ("grows in the dark"); similarly, the observable facts captured distinctive biological attributes of the stimuli ("has a long beak" or "has thin petals"). Thus, both labels and facts in our studies established dependable pathways into natural kind categories. More generally, our data suggest that rich conceptual information (e.g., biological facts for natural kinds) can be as effective as labels in establishing membership in a category.

Several further questions remain open and need to be addressed by additional experimentation and theorizing. First, our research revolved around a specific type of label (count names for objects) and specific types of facts (biological facts). It is important to extend this work to other types of linguistic stimuli (e.g., labels for relations, properties, and events) provided that perceptual similarity can be manipulated with the same precision as it was with objects. Similarly, it would be interesting to test the effectiveness of other types of facts both linguistically and nonlinguistically presented. A straightforward prediction from our approach is that the potency of facts should disappear for transient facts that have little predictive value in terms of deeper categorical structure ("This one is sad" or "This one has a sticker on") (cf. Gelman & Coley, 1990).

Second, our stimuli depicted natural kinds that are generally tightly clustered by similarities, unlike artifact categories that tend to be more flexible and spread out in similarity space (Malt et al., 1999). How would the current results concerning the role of labels and facts generalize beyond natural kinds to artifact categorization? We know that 3- and 4-year-olds rely on unobservable facts such as the creator's intent more than perceptual similarity to categorize artifacts such as drawings (Bloom & Markson, 1998). Nevertheless, recent work suggests that natural kinds and artifacts behave differently in terms of how conceptual and perceptual cues are weighed in categorization (Diesendruck & Peretz, 2013). Further research is required to understand the full extent to which labels and facts affect aspects of early categorization.

Acknowledgments

This article is based on the first author's doctoral dissertation. This research was supported in part by National Science Foundation (NSF) Grant BCS-0641105 to Anna Papafragou. We thank the children, parents, and staff at the Early Learning Center (Newark, DE) and the other Newark day care facilities that participated. We thank Paul Quinn, Roberta Golinkoff, Helene Intraub, and Brian Ackerman for their input on this project. We also thank Madeleine Page, Felix Gervits, and Jennifer Damonte for their assistance in designing the stimuli and running the study. Megan Johanson is now at the Crane Center for Early Childhood Research and Policy at The Ohio State University and is funded via Grant R305F05124 from the U.S. Department of Education, Institute of Education Sciences.

References

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64*, 3–26.
Blewitt, P., Golinkoff, R. M., & Alioto, A. (2000). Do toddlers have label preferences? A possible explanation for word refusals. *First Language, 20*, 253–272.
Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science, 9*, 200–204.
Bloom, P., & Markson, L. (2001). Are there principles that apply only to the acquisition of words? A reply to Waxman and Booth. *Cognition, 78*, 89–90.
Booth, A., & Waxman, S. R. (2002a). Object names and object functions serve as cues to categories for infants. *Developmental Psychology, 38*, 948–957.
Booth, A., & Waxman, S. R. (2002b). Word learning is "smart": Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition, 84*, B11–B22.
Booth, A., Waxman, S. R., & Huang, Y. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology, 41*, 491–505.
Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition, 104*, 287–314.
Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15*, 73–86.
Casasola, M., & Bhagwat, J. (2007). Do novel words facilitate 18-month-olds' spatial categorization? *Child Development, 78*, 1818–1829.
Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development, 74*, 679–693.
Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development, 5*, 151–176.
Deng, W., & Sloutsky, V. M. (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science, 23*, 178–186.
Diesendruck, G., & Peretz, S. (2013). Domain differences in the weights of perceptual and conceptual information in children's categorization. *Developmental Psychology, 49*, 2383–2395.
Feigenson, L., Carey, S., & Spelke, E. S. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33–66.
Ferry, A., Hespos, S., & Waxman, S. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development, 81*, 472–479.
Fulkerson, A., & Waxman, S. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 105*, 218–228.
Gelman, S. A. (2003). *The essential child*. 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*, 796–840.
Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.
Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532–1541.
Gelman, S. A., & O'Reilly, A. W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development, 59*, 876–887.
Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development, 75*, 1098–1114.
Graham, S. A., Welder, A. N., & McCrimmon, A. W. (2003). Hot dogs and zavy cats: Preschoolers' and adults' expectations about familiar and novel adjectives. *Brain and Language, 84*, 16–37.
Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science, 12*, 141–147.
Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45–75.
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., & Landau, B. (1991). Object properties and knowledge and knowledge in early lexical learning. *Child Development, 62*, 499–516.

Keil, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), *Language learning and concept acquisition* (pp. 133–153). Norwood, NJ: Ablex.

Landau, B., & Shipley, E. (2001). Labeling patterns and object naming. *Developmental Science, 4*, 109–118.

Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language, 38*, 1–27.

Lupyan, G., Rakison, D., & McClelland, J. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18*, 1077–1083.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*, 230–262.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Markman, E. M., & Hutchinson, J. E. (1988). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. In M. B. Franklin & S. S. Barten (Eds.), *Child language: A reader* (pp. 137–157). New York: Oxford University Press.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385*, 813–815.

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition, 80*, B11–B20.

Oakes, L. M., & Rakison, D. H. (2003). *Early category and concept development*. New York: Oxford University Press.

Plunkett, K., Hu, J.-F., & Cohen, L. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*, 665–681.

Quinn, P. C. (2004). Spatial representation by young infants: Categorization of spatial relations or sensitivity to a crossing primitive? *Memory & Cognition, 32*, 852–861.

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.

Quinn, P. C., Norris, C., Pasko, R., Schmader, T., & Mash, C. (1999). Formation of a categorical representation for the spatial relation of between by 6- to 7-month-old infants. *Visual Cognition, 6*, 569–585.

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.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72*, 1695–1709.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*, 13–19.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 29–56.

Subrahmanyam, K., Landau, B., & Gelman, R. (1999). Shape, material, and syntax: Interacting forces in children's learning in novel words for objects and substances. *Language and Cognitive Processes, 14*, 249–281.

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development, 5*, 123–150.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition, 70*, B35–B50.

Waxman, S. R., & Booth, A. (2000). Principles that are invoked in the acquisition of words, but not facts. *Cognition, 77*, B33–B43.

Waxman, S. R., Lynch, E. B., Casey, K. L., & Baer, L. (1997). Setters and samoyeds: The emergence of subordinate level categories as a basis for inductive inference in preschool-aged children. *Developmental Psychology, 33*, 1074–1090.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-month-old infants. *Cognitive Psychology, 29*, 257–302.

Waxman, S. R., & Namy, L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology, 33*, 555–567.

Waxman, S. R. (2003). Links between object categorization and naming: Origins and emergence in human infants. In D. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the booming, buzzing confusion* (pp. 213–241). New York: Oxford University Press.

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

Welder, A. N., & Graham, S. A. (2006). Infants' categorization of novel objects with more or less obvious features. *Cognitive Psychology, 52*, 57–91.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 776–795.

Yoshida, H., & Smith, L. (2003). Known and novel noun extensions: Attention at two levels of abstraction. *Child Development, 74*, 564–577.