Cognitive Development 48 (2018) 167–175

Contents lists available at ScienceDirect

Cognitive Development

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

Linguistic cues are privileged over non-linguistic cues in young children's categorization☆

Sarah Fairchild*, Ariel Mathis, Anna Papafragou

Department of Psychological & Brain Sciences, University of Delaware, USA

ARTICLE INFO

Keywords:
Categorization
Labels
Artifacts
Natural kinds

ABSTRACT

Language affects the way that humans build categories. When two objects share a verbal label, children and adults are encouraged to group them together. In the present study, we offer a stringent test of the potency of labels by comparing them to non-linguistic cues that have been matched in terms of critical properties. In Experiment 1, Four-year-old children were given two categorization tasks with novel natural kinds and artifacts. In both tasks, we compared the effectiveness of novel Labels like *zeg* and equally discriminable, intentionally introduced patterned Frames. In Experiment 2, we included pretest trials before each of the tasks to ensure children's awareness of the cues. We observed a pervasive advantage of Labels over Frames in both experiments. Our results offer some of the strongest evidence to date for the conclusion that young children prioritize labels over superficially equipotent non-linguistic cues when drawing category boundaries.

It is well known that language affects the way that humans build categories. At just a few months of age, the presence of verbal labels helps infants to form a category of objects where they otherwise would not, despite the fact that very little linguistic ability is present at this stage of development (Fulkerson & Waxman, 2007; Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Graham, Kilbreath, & Welder, 2004; Waxman & Markow, 1995). More sophisticated language users – preschool-aged children – continue to benefit from the presence of verbal labels: young children are likely to extend a novel property from one exemplar to another with a shared label in an induction task (Gelman & Markman, 1986, 1987; Davidson & Gelman, 1990; Gelman & Coley, 1990; Childers & Tomasello, 2003; Sloutsky & Fisher, 2004; Sloutsky, Kloos, & Fisher, 2007; Gelman & Davidson, 2013) and believe on the basis of a shared label that two objects belong together (Diesendruck & Peretz, 2013; Johanson & Papafragou, 2016; Sloutsky, Lo, & Fisher, 2001). Labels seem to facilitate categorization more so than other cues beginning at just a few months of life. In one study, three-month-old infants formed a category of dinosaurs when the exemplars were presented with labels, but not when they were presented with tones (Ferry et al., 2010). Similar results have been found with six- to twelve-month-old infants (Fulkerson & Waxman, 2007), even though younger infants are less selective (Ferry et al., 2010).

Although the strongest evidence for the unique role of language in categorization comes from these comparisons of labels with other, non-linguistic cues, such comparisons are unavailable for older children. This presents an important gap in our knowledge about how children use language to guide category formation because there are ongoing debates concerning the precise mechanisms underlying preschoolers' use of labels in category formation (Davidson & Gelman, 1990; Gopnik & Sobel, 2000; Sloutsky et al., 2001; Booth & Waxman, 2002; Sloutsky et al., 2007; Sloutsky & Fisher, 2004). Proponents of *Conceptual* accounts argue that even young children treat linguistic labels as conceptual category markers (e.g., Gelman & Markman, 1986; Diesendruck & Peretz, 2013; Gelman

☆ We thank Amelia Weiss for assistance with data collection and stimuli creation. This work was partially supported by NSF#1632849 to A.P.

* Corresponding author at: Department of Psychological & Brain Sciences, University of Delaware, 401 Wolf Hall, Newark, DE 19716, USA.
E-mail address: sarahf@udel.edu (S. Fairchild).

https://doi.org/10.1016/j.cogdev.2018.08.007
Received 28 July 2016; Received in revised form 13 August 2018; Accepted 17 August 2018

& Davidson, 2013). On these accounts, young children understand, for instance, that count nouns can be used to refer to objects, and objects that share a label also share underlying unobservable properties (i.e., belong to the same category). According to *Attentional/Perceptual* accounts, linguistic labels are external cues that serve to draw children's attention to relevant details of the stimuli, highlighting similarities and differences among them (e.g., Jones, Smith, & Landau, 1991; Sloutsky et al., 2001; Smith, 2005; Sloutsky et al., 2007). Such accounts argue that children do not treat verbal labels as signals of underlying conceptual category structure, but take language to enhance attention to relevant object features. These two accounts make different predictions about how linguistic labels should fare compared to other cues. Conceptual accounts predict an advantage of labels over non-linguistic cues, while Attentional/Perceptual accounts predict that non-linguistic cues will function just as well as linguistic labels as cues for categorization as long as such non-linguistic cues are perceptually equivalent to the labels.

A related issue that bears on the debate about whether children treat labels as conceptual or perceptual/attentional cues is that existing comparisons of the potency of linguistic and non-linguistic cues for categorization have not typically equated the two types of cue for discriminability. Labels may exert a greater effect on categorization simply because two labels might be perceived as being more distinct from one another than, for instance, two tones; as a result, labels might more clearly signal a contrast between members and non-members of a particular category. Similarly, linguistic and non-linguistic cues have not typically been equated for intentionality. Verbal labels, unlike non-linguistic cues, are typically communicated by the experimenter and hence are bound to appear more intentional, and therefore more relevant for the task, than other cues. In support of the role of intentionality, infants have been found to ignore labels produced by a non-human source such as a baby monitor (Campbell & Namy, 2003; Fulkerson & Haaf, 2003) and make use of non-linguistic cues when they appear to be part of a conversation (Ferguson & Waxman, 2016). In a further recent demonstration with older learners, novel labels ('It's a zeg!') were found to be as potent as linguistically communicated facts ('It drinks milk!') in shaping four-year-old children's categorization of novel animals and plants (Johanson & Papafragou, 2016; cf. also Diesendruck & Peretz, 2013).

In sum, it is an open question whether young children prioritize linguistic labels when forming novel categories compared to perceptually comparable, intentional non-linguistic cues. The present study fills this gap in the literature by comparing the effectiveness of linguistic cues (novel verbal Labels like *zeg*) and non-linguistic cues (patterned, geometric Frames that surround each to-be-categorized object) in four-year-olds' categorization decisions. Unlike past work, we took steps to equate linguistic and non-linguistic cues for discriminability to guard against low-level differences in responses to stimuli. Additionally, we took steps to equate the two types of cues for intentionality: the experimenter drew children's attention to the Labels and Frames in a similar manner, such that both types of cues were presented as intentional. Of interest is whether language would still hold a special status in categorization (in line with Conceptual but not Perceptual/Attentional accounts). If so, children should make greater use of Labels than Frames in a categorization task despite these steps.

In a further departure from prior studies that have typically focused on a single class of stimuli (e.g., novel animals), we explore the role of Labels and Frames across two categorization tasks: one with novel man-made artifacts (mostly tools) and one with unfamiliar exemplars of natural kinds (mostly animals and plants). By replicating the task across two domains, we can determine whether the effects of labels and other cues generalize to any type of stimuli, or whether labels might hold a special status for a limited class of stimuli. For example, some theorists expect a difference in children's cue use across domains because of beliefs children might reasonably hold about category structure. Specifically, some hypothesize that the influence of labels should be greatest for artifact categories but more restricted for natural kinds (e.g., Rhodes & Gelman, 2009) because natural kinds like plants and animals are believed to have "essences," undefinable but biological internal properties that define strict category boundaries (e.g., Atran, 1990; Gelman, 2013). Man-made artifacts, on the other hand, are typically categorized according to creator intent, function, and other social information and might thus be more susceptible to influences from labels or other intentional cues (Bloom, 1996; Margolis & Laurence, 2007). Some support for this position comes from a recent study in which five-year-olds used labels more actively to categorize novel artifacts than novel animals (Diesendruck & Peretz, 2013). Nevertheless, because the labels were not compared to other cues, it is not clear whether this effect was selectively due to the presence of labels or whether similar results might have been obtained with different (e.g., non-linguistic) cues.

In line with previous developmental research, we use a strong test of the efficacy of linguistic and non-linguistic cues by manipulating whether perceptual similarity in the stimuli was consistent, inconsistent or uninformative (ambiguous) with respect to the groupings indicated by these cues. Previous research has found complex interactions between perceptual similarity and the facilitative effects of labels. A label may encourage children to group perceptually dissimilar objects together (Gelman & Markman, 1986, 1987) or draw category boundaries for perceptually ambiguous objects (Johanson & Papafragou, 2016), although the influence of labels may not completely override perceptual similarity (Sloutsky et al., 2001, 2007; Sloutsky & Fisher, 2004; Gelman & Davidson, 2013). For present purposes, children should be more likely to group two objects together as perceptual similarity between the objects increases regardless of the cue (Label, Frame) that the objects might share. However, if language is privileged in children's categorization decisions, Labels (but not Frames) should exert an influence beyond perceptual similarity alone.

1. Experiment 1

Experiment 1 investigated children's spontaneous use of linguistic and non-linguistic cues in forming novel artifact and natural kind categories.

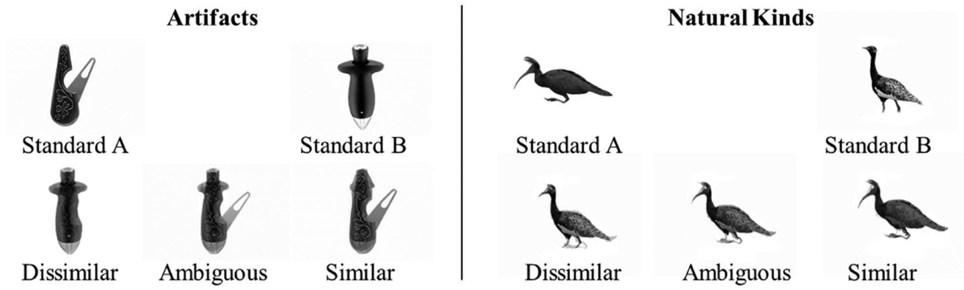


Fig. 1. Example stimuli with similarity manipulation for artifacts (left) and natural kinds (right). Dissimilar, Ambiguous, and Similar Target objects are presented below the Standard objects.

1.1. Method

1.1.1. Participants

Thirty-six four-year-olds, eighteen of whom were female (M_{age} = 4;5, range = 4;1 to 5;0), were recruited from a local preschool in Newark, DE, and participated in the study. All were typically developing native English speakers without learning difficulties.

1.1.2. Materials

Materials for the Artifact categorization task consisted of twenty-four grayscale photographs of novel objects (mostly mechanical tools). These objects were selected to be unfamiliar, and a control group of adults ($N = 32$) confirmed that they could not be identified. Materials for the Natural Kinds categorization task consisted of twenty-four grayscale photographs of novel plants and animals that had been used in a set of previous studies, and were shown to be unnamable by a majority of adult participants (Johanson & Papafragou, 2016). Within the Artifacts and the Natural Kinds group, the stimuli were placed into twelve pairs that served as the Standards in the categorization task, with one stimulus in the pair being termed as Standard A, and the other as Standard B. During the categorization task, Standard A always shared a cue with the target object.

To obtain the targets for categorization, we morphed together the stimuli in each pair using FantaMorph, a professional morphing software. We manipulated perceptual similarity such that, for each pair, three morphed Target stimuli were created at 30%, 50%, and 70% similarity to Standard A, as determined by the FantaMorph program. Thus the 30% Target would be 30% visually similar to Standard A, and 70% visually similar to Standard B, the 50% Target would be equally similar to both Standards, and the 70% Target would be 70% similar to Standard A and 30% similar to Standard B. We refer to these three similarity conditions as Dissimilar (30%), Ambiguous (50%), and Similar (70%), always with respect to Standard A. Samples of stimuli are presented in Fig. 1.

To check the Similarity manipulation, two control groups of adults ($N = 10$ for Artifacts and $N = 18$ for Natural Kinds) were shown all possible triads and were asked to rate the similarity of the Target item to one of the two Standards, on a scale from 1 (completely dissimilar) to 9 (completely similar). Mean similarity ratings did not differ from expected similarity ratings (30%, 50% and 70%) for the Dissimilar, Ambiguous, or Similar Targets within each class of stimuli (all p's $> .1$).

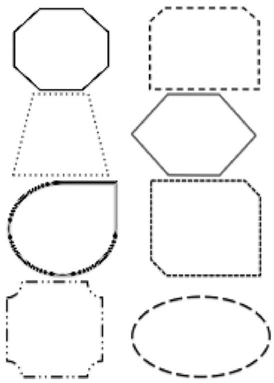
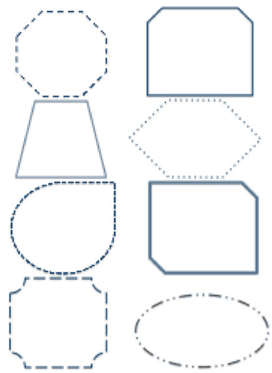
We generated two types of cues, Labels (linguistic) and Frames (non-linguistic). Sixteen pseudowords between three and six letters in length from the ARC Non-Word Database served as Labels (Rastle, Harrington, & Coltheart, 2002). All pseudowords were phonologically possible in English (e.g., *zeg*). Sixteen patterned geometric shapes meant to surround each exemplar within a triad served as Frames. For the Artifacts task, all Frames were black shapes drawn with different line patterns: solid, dashed, double line, and so on. For the Natural Kinds task, the Frames were navy blue shapes drawn with a new set of line patterns. All Labels and Frames are presented in Table 1.

We checked that Labels and Frames did not differ from each other in terms of discriminability by presenting a control group of adults ($N = 10$) with all pairs of Labels and Frames that were used in the experiment and asking them, for each pair, to rate how similar the two cues were to each other, with 1 being "completely different" and 7 being "completely identical." A *t*-test revealed that ratings for Labels and Frames did not differ significantly from one another ($p > .1$), with participants rating both paired Labels ($M = 2.5$, $SD = .89$) and Frames ($M = 2.3$, $SD = 1.25$) as very distinct from each other.

1.1.3. Procedure

Children completed both the Artifact and the Natural Kinds categorization task (order counterbalanced across participants) over the course of two sessions, approximately one week apart. Both tasks were administered using the OpenSesame presentation software (Mathôt, Schreij, & Theeuwes, 2012). Each task consisted of three blocks with twelve trials each: one Label block, one Frame block, and one block without cues (No Cue), with order counterbalanced across participants. Four unique sets of morphed objects (Standard A, Standard B, and their associated Dissimilar, Ambiguous, and Similar Targets) were presented in each block – i.e. children never associated the same set of morphed objects with both Labels and Frames. Trials were randomized within each block. On every trial, two of the cues (associated with the Target and Standard A) were the same and the other cue (associated with Standard B) was different. The position of the cue-matched Standard A was counterbalanced, with it being equally likely for Standard A to appear on the left or the right side of the screen. Additionally, object triads were fully rotated through the cues to guard against the possibility that particular cue-object pairings would influence the results.

Table 1
Frame and Label cues for Artifacts (left) and Natural Kinds (right).

	Artifacts	Natural Kinds
Label *"Look at the yurf!"* Frame *"Look at this!"*	*yurf, pank, croot, sten, smeel, floob, zeg, wob* 	*furp, kanf, troob, nem, leem, bloog, geb, bof* 

For each trial, a triad of stimuli was presented on a computer screen: the two Standards appeared at the top of the screen, and the Target appeared on the bottom, separated from the Standards by a black line. In the Label condition, the experimenter pointed at each of the stimuli and offered a novel Label for each one ("Look at the *lorp*!"). In the Frame condition, the experimenter pressed a button to make a patterned Frame appear around each object as she pointed and said "Look at this!" The Frame disappeared when the experimenter moved to the next stimulus in the triad, making both Labels and Frames temporarily available. In the No Cue condition, the experimenter simply said "Look at this!" when pointing to each stimulus within a triad. After introducing all stimuli within a triad, the experimenter asked children, "Where does the bottom one belong?" Children responded by pointing to one of the two Standards.

1.2. Results and discussion

We considered a response to be *Cue Compliant* when the child chose the cue-matched Standard A. For example, in the Label condition, grouping a Target named *zeg* with Standard A, also named *zeg,* would be a Cue-Compliant response. Note that, for Dissimilar trials, a Cue-Compliant response ignored the perceptual distance between the Target and Standard A. For Ambiguous trials, a Cue-compliant response indicated that cues introduced an asymmetry between Standards that were otherwise perceptually equidistant from the Target. On Similar trials, a Cue-Compliant response consisted of choosing Standard A that both shared a Label or Frame with the Target and was perceptually similar to it. Because each Standard was assigned as A or B during stimuli creation, we were able to calculate Cue Compliance for the No Cue condition in the same way (i.e., responses to Standard A). Thus in the No Cue condition mean Cue Compliance should be very low cue for Dissimilar trials, around chance (50%) for Ambiguous trials, and very high for Similar trials.

We used a mixed effects logistic regression model to analyze the data to better handle our binary dependent variable (Cue Compliance), our within-subjects independent variables, and subject-level and item-level variability. Participant and item were included as crossed random intercepts and Similarity (Dissimilar, Ambiguous, Similar), Cue Type (Label, Frame, No Cue) and Domain (Artifacts, Natural Kinds) were treated as fixed effects. Data from 11 children in the Artifact Categorization Task and 4 children in the Natural Kinds Categorization Task were excluded for either missing data or having a tendency to group together Dissimilar objects over 50% of the time and/or to split apart Similar objects over 50% of the time in the absence of any cue (No Cue condition).

Results are shown in Fig. 2. Cue Compliance varied significantly across levels of Similarity, $\chi^2(2) = 262.699$, $p < .001$, such that Cue Compliance increased as perceptual similarity increased, as expected. Cue Compliance also varied significantly by Cue Type, $\chi^2(2) = 101.311$, $p < .001$. Bonferroni-corrected post-hoc tests indicated that there was significantly higher Cue Compliance for Labels as compared to Frames, $p < .001$ and for Labels as compared to the No Cue condition, $p < .001$. Cue Compliance in the Frame condition did not differ significantly from the No Cue condition, $p > .1$. Finally, Cue Compliance differed significantly across Domains, $\chi^2(2) = 8.324$, $p = .004$, such that there were more Cue Compliant responses in the Artifact Task than the Natural Kinds Task.

There was a significant interaction between Similarity and Cue Type, $\chi^2(4) = 14.105$, $p = .007$. Post-hoc tests indicated that, on Dissimilar trials, children gave significantly more Cue-Compliant responses in the Label condition than in the Frame and No Cue conditions, both p's $< .001$; Cue Compliance did not differ between the Frame and No Cue conditions ($p > 0.1$). On Ambiguous trials, there was significantly higher Cue Compliance in the Label condition than either the Frame condition ($p < .001$) or the No Cue condition ($p = .006$); Cue Compliance did not differ between the Frame and No Cue conditions ($p > .1$). On Similar trials, Cue Compliance did not differ between Labels and Frames ($p > .1$) and neither Labels nor Frames differed from the No Cue condition

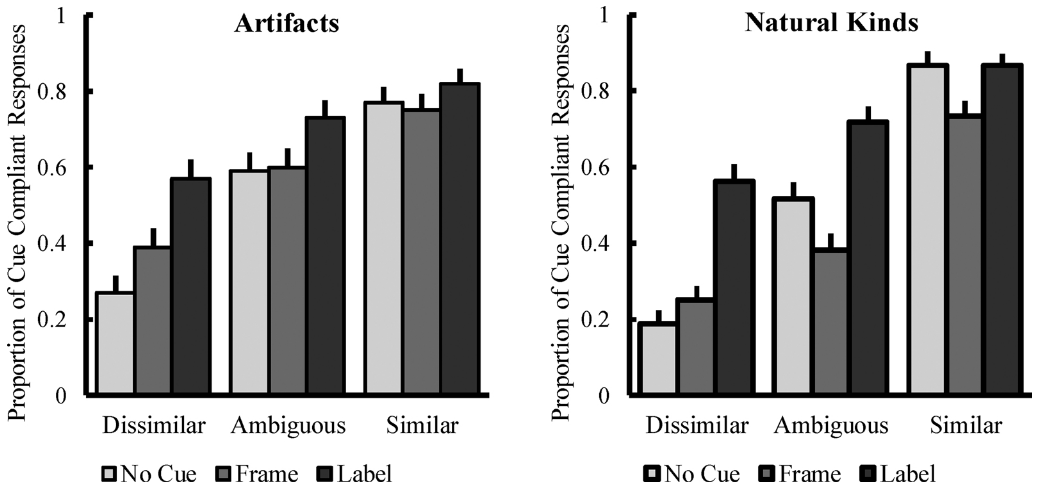


Fig. 2. Cue compliance for the Artifact and Natural Kinds Categorization Tasks in Experiment 1. Error bars represent ± 1 S.E.M.

(both p's > .1).

The interaction between Cue Type and Domain was also significant, $\chi^2(4) = 6.860$, $p = .032$. Post-hoc tests indicated that this interaction was due to significantly greater use of Frames in the Artifact Task as compared to the Natural Kinds Task, $p = .013$. Cue Compliance in the Label and No Cue conditions did not differ across Domains, both p's > .1. Interactions between Similarity and Domain, $\chi^2(4) = 4.100$, $p = .129$; and among Similarity, Cue Type, and Domain, $\chi^2(6) = 1.280$, $p = .865$; did not reach significance.

In sum, we observed an advantage of linguistic Labels over non-linguistic Frames in four-year-olds' formation of novel artifact and natural kind categories. Categorization was also sensitive to perceptual similarity, with visually similar stimuli being more likely to be grouped together. The advantage of Labels over Frames was particularly strong when the objects being categorized were visually dissimilar or ambiguous with respect to perceptual similarity. Furthermore, the use of Labels for categorization remained constant across domains, with the use of non-linguistic Frames slightly higher for Artifacts than for Natural Kinds, potentially because of the more malleable boundaries of Artifact categories (cf. General Discussion below).

2. Experiment 2

The results of Experiment 1 suggest that, by age four, children already understand that labels serve as category markers but arbitrary non-linguistic cues (patterned frames) do not. One possible concern is that, despite our efforts, frames may have appeared less intentional than the labels in our paradigm (for instance, they could have been interpreted as having no particular meaning for the experimenter). As a result, children may not have paid attention to frames, noticed that frames varied, or connected them to category membership. In Experiment 2 we replicated Experiment 1 but inserted pretest trials at the beginning of each block to address this concern. If the label advantage persisted, it would strongly support the conclusion that children have a sophisticated understanding of the role of language in categorization.

2.1. Method

2.1.1. Participants

Thirty-one four-year-olds, twenty of whom were female (M_{age} = 4;5, range = 3;11–5;4), were recruited from a local preschool in Newark, DE, and participated in the study. All were typically developing native English speakers without learning difficulties. None of them had participated in the earlier study.

2.1.2. Materials

The stimuli were identical to Experiment 1.

2.1.3. Procedure

The categorization tasks proceeded in the same fashion as in Experiment 1, with one key change. At the beginning of each block of the experiment (Label/Frame/No Cue), a pretest trial was administered in which three familiar objects (two distinct chairs and a table for the Artifact Categorization Task, two different dogs and a cat for the Natural Kinds Categorization Task) were presented in the same format as the experimental trials (such that, e.g., the table and one chair would appear at the top of the screen as Standards and the other chair would appear at the bottom as the Target). With the exception of the No Cue condition, the experimenter introduced cues appropriate for that particular block and explicitly addressed the fact that two of the cues were the same. For example, for the Label pretest the experimenter pointed at each object and said, "Look, this one is a table! Look, this one is a chair! Look, this one is *also* a chair!" For the Frame pretest, the experimenter pointed at each object, pressed a button to make a Frame

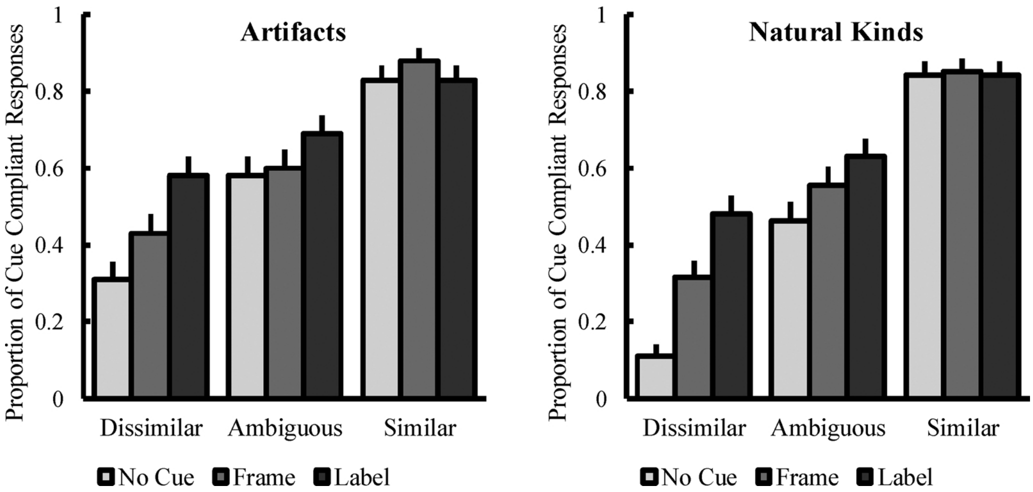


Fig. 3. Cue compliance for the Artifact and Natural Kinds Categorization Tasks in Experiment 2. Error bars represent ± 1 S.E.M.

appear around it and said, "Look, this one gets this shape! Look, this one gets this shape! Look, this one *also* gets this shape!" For the No Cue pretest, the experimenter simply said, "Look at this one!" while pointing to each object. For each pretest, children were then asked, "Where does the bottom one belong?", and had to choose one of the two Standard objects. Because we used familiar objects, there was a clear correct answer to the pretest trials. Children were given feedback on their answer that included an additional mention of the cue: "(That's right!) It belongs with *this* one, and they both have the same name/have the same shape/look alike!" We reasoned that pretests would encourage children to (a) interpret both labels and frames as intentional, (b) notice that two of them were the same within a trial, (c) connect them with category membership.

2.2. Results and discussion

Again we used a mixed effects logistic regression model to analyze the data to better handle our binary dependent variable (Cue Compliance), our within-subjects independent variables, and subject-level and item-level variability. Participant and item were included as crossed random intercepts and Similarity (Dissimilar, Ambiguous, Similar), Cue Type (Label, Frame, No Cue) and Domain (Artifacts, Natural Kinds) were treated as fixed effects. Data from two children in the Artifact Categorization Task and four children in the Natural Kinds Categorization Task were excluded for either missing data or having a tendency to group together Dissimilar objects over 50% of the time and/or to split apart Similar objects over 50% of the time in the absence of any cue (No Cue condition).

Results are shown in Fig. 3. Cue Compliance varied significantly across levels of Similarity, $\chi^2(2) = 356.171$, $p < .001$, such that Cue Compliance increased with Perceptual Similarity. Cue Compliance also significantly varied by Cue Type, $\chi^2(2) = 41.418$, $p < .001$. Post-hoc tests indicated that there was significantly higher Cue Compliance for Labels as compared to Frames, $p = .012$. Cue Compliance was significantly higher in the Label, $p < .001$, and Frame, $p = .001$, conditions as compared to the No Cue condition. Finally, Cue Compliance differed significantly across Domains, $\chi^2(2) = 15.129$, $p < .001$, such that there were more Cue Compliant responses in the Artifact Task than the Natural Kinds Task.

There was a significant interaction between Similarity and Cue Type, $\chi^2(4) = 19.283$, $p < .001$. Post-hoc tests indicated that on Dissimilar trials, children gave significantly more Cue-Compliant responses in the Label condition than in the Frame, $p = .016$, and No Cue, $p < .001$, conditions. Cue Compliance was higher in the Frame condition as compared to the No Cue condition, $p = .019$. On Ambiguous and Similar trials, the three Cue Types did not differ from one another, all p's $> .1$.

The interaction between Similarity and Domain was also significant, $\chi^2(4) = 6.392$, $p = .041$. Specifically, Cue Compliance was greater for Artifacts as compared to Natural Kinds for Dissimilar trials, $p = .001$. Cue Compliance did not differ significantly across Domains on either Ambiguous or Similar trials, both p's $> .1$. There were no interactions between Cue Type and Domain, $\chi^2(4) = 1.640$, $p = .441$, or among Similarity, Cue Type, and Domain, $\chi^2(6) = 5.954$, $p = .203$.

These results align with those of Experiment 1 in that we observed a preference for Labels over Frames while forming novel categories. Additionally, as in Experiment 1, Cue Compliance was higher overall for Artifacts than for Natural kinds (particularly on Dissimilar trials), potentially pointing to the more malleable nature of artifact category boundaries. Unlike Experiment 1, the use of cues did not differ across domains (children used Frames more often for Artifacts than for Natural Kinds in Experiment 1). This difference is potentially due to a greater overall use of Frames for natural kinds in Experiment 2, a possibility that we test directly in the next section.

2.2.1. Comparisons to experiment 1

To directly assess the influence of the pretest trials included in Experiment 2, we conducted two analyses – one for each Domain – to compare patterns of performance between Experiments 1 and 2. First, we conducted a mixed effects logistic regression predicting Cue Compliant responses on the Artifacts Task that included Similarity (Dissimilar, Ambiguous, Similar), Cue Type (Label, Frame, No

Cue) and Experiment (Experiment 1, Experiment 2) as fixed effects. Cue Compliance varied significantly across levels of Similarity, $\chi^2(2) = 222.113$, $p < .001$, and Cue Type, $\chi^2(2) = 38.842$, $p < .001$. The difference in Cue Compliance between Experiments 1 and 2 was not significant, $\chi^2(1) = 0.627$, $p = .428$. Unsurprisingly, the interaction between Similarity and Cue Type was significant, $\chi^2(4) = 11.458$, $p = .022$. There were no other significant interactions (all p's > .1).

We next conducted a complementary analysis predicting Cue Compliance on the Natural Kinds Task. Cue Compliance varied significantly across levels of Similarity, $\chi^2(2) = 411.915$, $p < .001$, and Cue Type, $\chi^2(2) = 104.816$, $p < .001$, and there was an interaction between Similarity and Cue Type, $\chi^2(4) = 22.650$, $p < .001$. There was no difference in Cue Compliance between Experiments 1 and 2, $\chi^2(1) = 0.438$, $p = .508$. There was a significant interaction between Similarity and Experiment, $\chi^2(3) = 6.924$, $p = .031$. Planned contrasts indicated that the difference in Cue Compliance between Dissimilar and other types of trials (Ambiguous and Similar combined) was greater in Experiment 2 as compared to Experiment 1. Finally, there was a significant interaction between Cue Type and Experiment, $\chi^2(3) = 14.394$, $p < .001$, such that there was greater Cue Compliance in the Frame condition in Experiment 2 as compared to Experiment 1, $p = .030$. Cue Compliance did not differ between Experiments for the Label and No Cue conditions, both p's > .1.

Thus children performed similarly on the Artifact Tasks of Experiment 1 and 2. On the Natural Kinds Tasks, Cue Compliance increased for Frames from Experiment 1 to Experiment 2. This suggests that the pretest trials affected children's categorization behavior by emphasizing the intentionality and potential significance of the frames. However, the advantage of Labels over Frames remained.

3. General discussion

In the present study, we set out to assess the role of language in young children's category formation. Previous studies suggested that infants make use of linguistic cues to a greater extent than non-linguistic cues when forming novel categories but the properties of the two types of cues (discriminability, intentional status) were not always matched; furthermore, comparisons of labels to other cues in preschool-aged children were limited. We addressed these issues by comparing the effects of two salient cues – Labels and patterned Frames – on the categorization decisions of four-year-olds after taking steps to equate these cues in terms of discriminability and intentionality. We observed a pervasive advantage of Labels over Frames across two tasks involving Artifacts and Natural kinds (Experiment 1). Although Frame use increased when the intentional nature of both the linguistic and non-linguistic cues and the fact that they co-varied with category membership was pointed out during pretests (particularly for natural kinds), the label advantage persisted (Experiment 2). Our data thus suggest that language holds a special status in four-year-olds' categorization. A shared label encourages four-year-olds to group together objects and animals, but a shared non-linguistic frame is not treated in the same way. Because label/frame compliance overall increased with perceptual similarity, our results indicate that categorization is a nuanced process, involving integration of both perceptual and conceptual/linguistic information (see also Gelman & Davidson, 2013).

Could there be alternative explanations for the present findings that do not rely on a special status for labels? One possibility is that the advantage of labels over frames in the present study was due to children's preference for auditory over visual information. Some researchers have argued that auditory information is preferred over visual information in infancy when the visual system is underdeveloped, reversing to a visual preference in adulthood (Robinson & Sloutsky, 2004). However, these researchers report that children of preschool age shift attention between visual and auditory stimuli and thus have no preference for auditory information. We also know that four- to five-year-old children show an overwhelming tendency to match geometric shapes to each other in a triad-matching task (Abravanel, 1970) and massively attend to shapes of objects in categorizing and applying novel labels to objects (Landau, Smith, & Jones, 1992; Landau, Smith, & Jones, 1988; Landau, Smith, & Jones, 1998). Moreover, in a naturalistic setting (viewing a communicative exchange on television), visual information was found to be more salient and memorable for five-year-old children than auditory information (Pezdek & Stevens, 1984). Thus we doubt that this is a viable explanation of our findings.

A second possibility is that the advantage of labels over frames, especially in Experiment 2, may have been due to an asymmetry in how valid the two types of cues were assumed to be. Recall that, in the pretest trials of Experiment 2 that introduced the logic of the task, familiar labels were used to refer to the stimuli, thereby presumably reinforcing the assumption that labels validly capture category membership (an assumption that children could have later generalized to novel labels during the main phase of the experiment). On this reasoning, the pretest manipulation might have provided an unfair advantage to labels over frames. We consider this an unlikely explanation. Notice that the label pretest trials did not provide any new information to children who already had a concept and a label for the stimuli. Furthermore, and in support of this possibility, Frame use (but not Label use) increased in Experiment 2 following the pretest trials.

One aspect of our data that merits further comment concerns the role of Labels across different domains. Recall that a previous study (Diesendruck & Peretz, 2013) found that five-year-olds relied on labels more when categorizing artifacts than when categorizing natural kinds. The authors took this difference as evidence that five-year-olds draw different conclusions about the role of labels in categorization depending on the nature of the stimuli to be categorized (with man-made Artifacts having relatively flexible category boundaries and Natural Kinds being defined by "essences" – biological properties that mark rigid category boundaries; cf. also Atran, 1990; Bloom, 1996; Margolis & Laurence, 2007; Gelman, 2013). In our own data, Frame use pointed in the direction documented by prior work – with Frames being used more actively to shape categories for artifacts than for natural kinds. Label use, however, was equally potent across domains. There are several differences between our own and the earlier study that might explain this small discrepancy. One difference is that we tested a slightly younger age group (note, however, that in a somewhat different paradigm, sensitivity to the differences in category structure between artifacts and natural kinds has been observed even at the age of three; Rhodes & Gelman, 2009). Another difference is that Diesendruck and Peretz (2013) repeatedly instructed children that they would be sorting “animals” or “objects”. In our own study, the difference between artifacts and natural kinds was not as strongly highlighted.

As mentioned in the Introduction, there is currently a debate in the literature about the mechanisms underlying the role of language in category formation. The pervasive advantage of Labels over Frames observed in the present study is consistent with (and anticipated by) Conceptual accounts, since these accounts assume that labels are privileged because of the conceptual weight that they carry. The Label advantage in our data seems less consistent with a simple version of a Perceptual/Attentional account, on which Labels and Frames (arguably similar in terms of discriminability and perceived intent) might be expected to draw children’s attention to similarities and differences between stimuli to a comparable degree. Additional assumptions leading to richer versions of such accounts might be able to accommodate our findings. Furthermore, the fact that intentional non-linguistic cues (Frames) were more likely to affect artifact compared to natural kind categories in our data strengthens the conclusion that preschoolers’ categorization is driven by rich knowledge of the stimuli to be categorized (as predicted by Conceptual accounts) as opposed to simple perceptual/attentional factors that should apply equally to any type of stimulus.

The present method has a number of limitations. For instance, we did not directly evaluate the relative perceptual discriminability of Labels and Frames as cue types for children (recall that discriminability ratings came from adults), nor did we directly compare whether children perceived the two cue kinds as equally intentional. Future efforts to distinguish between Perceptual/Attentional and Conceptual accounts of children’s categorization could benefit from additional measures of how children perceive these properties of linguistic and non-linguistic cues and integrate such cues with to-be-categorized stimuli.

To summarize, our work provides evidence for an advantage of linguistic over non-linguistic cues during young children’s novel category formation. In line with developmental research in numerical cognition (e.g., Gallistel & Gelman, 1992; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999), spatial reasoning (e.g., Loewenstein & Gentner, 2005), and spatial memory (e.g., Dessalegn & Landau, 2008, 2013), the present data support the conclusion that, in addition to being used for communication, language can be a unique and efficient tool for solving cognitively demanding tasks (see Wolff & Holmes, 2011; Ünal & Papafragou, 2016).

References

Abravanel, E. (1970). Choice for shape vs textural matching by young children. *Perceptual and Motor Skills, 31*, 527–533.
Atran, S. (1990). *Cognitive foundations of natural history*. Cambridge, England: Cambridge University Press.
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.
Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60*, 1–29.
Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology, 38*(6), 948.
Campbell, A. L., & Namy, L. L. (2003). The role of social-referential context in verbal and nonverbal symbol learning. *Child Development, 74*, 549–563.
Childers, J. B., & Tomasello, M. (2003). Children extend both words and non-verbal actions to novel exemplars. *Developmental Science, 2*, 185–190.
Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development, 5*, 151–176.
Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970–974.
Dessalegn, B., & Landau, B. (2008). More than meets the eye the role of language in binding and maintaining feature conjunctions. *Psychological Science, 19*(2), 189–195.
Dessalegn, B., & Landau, B. (2013). Interaction between language and vision: It’s momentary, abstract, and it develops. *Cognition, 127*(3), 331–344.
Diesendruck, G., & Peretz, S. (2013). Domain differences in the weights of perceptual and conceptual information in children’s categorization. *Developmental Psychology, 49*, 2383–2395.
Ferguson, B., & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition, 146*, 185–189.
Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development, 81*, 472–479.
Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9-and 15-month-olds’ object categorization. *Infancy, 4*, 349–369.
Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 105*, 218–228.
Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1), 43–74.
Gelman, S. A. (2013). Artifacts and essentialism. *Review of Philosophy and Psychology, 4*, 449–463.
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., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology, 66*, 327–353.
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.
Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*(5), 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.
Johanson, M., & Papafragou, A. (2016). The role of labels and facts in children’s and adults’ categorization. *Journal of Experimental Child Psychology, 144*, 130–151.
Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development, 62*, 499–516.
Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*(3), 299–321.
Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children’s and adults’ lexical learning. *Journal of Memory and Language, 31*(6), 807–825.
Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language, 38*(1), 1–27.
Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315–353.
Margolis, E., & Laurence, S. (Eds.). (2007). *Creations of the mind: Theories of artifacts and their representation*. Oxford: Clarendon Press.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324.
Pezdek, K., & Stevens, E. (1984). Children’s memory for auditory and visual information on television. *Developmental Psychology, 20*, 212.
Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A, 55*, 1339–1362.
Rhodes, M., & Gelman, S. A. (2009). Five-year-olds’ beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review, 16*, 920–924.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401.
Sloutsky, V. M., & Fisher, A. V. (2004). When development and learning decrease memory evidence against category-based induction in children. *Psychological Science, 15*, 553–558.
Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*, 179–185.
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. (2005). Emerging ideas about categories. In L. Gershkoff- Stowe, & D. H. Rakison (Eds.). *Building object categories in developmental time* (pp. 159–173). Mahwah, NJ: Erlbaum.
Ünal, E., & Papafragou, A. (2016). Interactions between language and mental representations. *Language Learning, 66*, 554–580.
Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology, 29*, 257–302.
Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews Cognitive Science, 2*, 253–265.