Measuring the development of the shape bias in early word learning across methodological variations

Anonymous CogSci submission

Abstract

Young children tend to generalize the meanings of new object categories based on their shape, a tendency known as the "shape bias." The mechanisms driving this bias and its variability remain controversial, however. Here, we explore methodological factors contributing to heterogeneity in shape bias studies, including task format, stimuli design. In two experiments, we systematically manipulated object properties—shape, material, function, and affordances—and examined their effects on 2–5 year-old children's generalization. Preliminary findings indicate a robust preference for shape, even when function is made salient, suggesting that children may prioritize shape as a proxy for category membership. However, nuanced interactions between stimuli design, task structure, and prior experiences highlight the role of methodological decisions in shaping children's behavior. Cross-linguistic and theoretical debates about the origins of shape bias underscore the need for unified methodologies to reconcile conflicting evidence.

Keywords: word learning; language acquisition; shape bias

What does "dog" mean to a 2-year-old? This old question reflects the evolving nature of our understanding of early label categorization. For young children, a dog might initially be identified by a set of general features that expands with experience. For example, The semantic features hypothesis (Clark (1973)) suggests that categorization begins with schemas of basic perceptual attributes, such as [four legs, tail, fur] for dog, hence we see overextensions of the word "dog" to beings that share these set of features like cats for example. Over time, deeper attributes, such as its internal composition, behavior and interaction with the world, or sound, may also become relevant. In contrast, the functional core hypothesis (Nelson, 1974) posits that children extract relationships among features to identify future category members without treating these features as defining. In this view, perceptual attributes like [four legs, tail, fur] act as identifiers rather than encapsulating the category's essence. Both frameworks agree, however, that shared perceptual features facilitate identification and labeling, forming the foundation for the well-studied shape bias Larissa K. Samuelson & Smith (2000).

What does "hammer" mean to a toddler? One possibility is that "hammer" refers to other hammer-shaped objects. The *shape bias* is the tendency to generalize objects names by their shape, rather than other properties. The presence of a shape bias is argued to facilitate early noun acquisition, to be an important route to vocabulary growth, and to be weaker in children with language delay (Susan S. Jones, 2003; JONES & SMITH, 2005; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Tek, Jaffery, Fein, & Naigles, 2008 ; Tek & Naigles, 2017). The shape bias is typically measured using word extension tasks, in which children are taught a novel label for an object and then tested on their ability to extend it to other objects. Yet although the shape bias is found robustly in such tasks, there is still substantial variation in when, and at what magnitude, the shape bias is detected (Abdelrahim & Frank, 2024; Kucker et al., 2019). This variation plays a key role in differentiating between theoretical accounts of the origins of this bias.

Although one possibility is that the shape bias is a universal constraint on generalization, robust evidence suggests cross-cultural variation, casting doubt on this explanation. For example, speakers of East Asian languages like Mandarin and Japanese demonstrate less reliance on shape when extending nouns compared to English speakers in the United States (Gathercole & Min, 1997; Imai & Gentner, 1997; Larissa K. Samuelson & Smith, 1999; Smith, Colunga, & Yoshida, 2003; Nancy N. Soja, Carey, & Spelke, 1991; Subrahmanyam & Chen, 2006; Yoshida & Smith, 2003b). Two key hypotheses attempt to explain these differences. First, differences in linguistic structure, such as count-mass syntax in English versus classifier systems in East Asian languages, might influence the prevalence of the shape bias (Imai & Gentner, 1997; Larissa K. Samuelson, Horst, Schutte, & Dobbertin, 2008; Nancy N. Soja et al., 1991; Nancy N. Soja, Carey, & Spelke, 1992). Second, variations in lexical and environmental statistical regularities tunes attention toward features like shape. This hypothesis emphasizes the role of existing vocabulary and environmental exposure in guiding category organization (Colunga & Smith, 2000; Gershkoff-Stowe & Smith, 2004; Jara-Ettinger, Levy, Sakel, Huanca, & Gibson, 2022; Perry, Samuelson, Malloy, & Schiffer, 2010; Larissa K. Samuelson, 2002, 2005; Larissa K. Samuelson & Smith, 1999; Yoshida & Smith, 2003a).

A second key question is the mechanism or representation underlying the shape bias. While the tendency to extend nouns based on shape may be influenced by syntax or statistical regularities, its precise cognitive basis has been controversial [Smith et al. (2002); Smith, Colunga, & Yoshida (2010); Smith, Jones, & Landau (1996); Brady & Chun, 2007; Chun & Jiang, 1998; Larissa K. Samuelson & Perone (2010); Ware & Booth (2010); A. Booth & Waxman (2006) ; Susan S. Jones & Smith (1993); L. Samuelson & Horst (2008);

Bruner (1964)]. Two competing perspectives attempt to explain this mechanism. One possibility is that the shape bias is an associative (non-strategic) mechanism, driven by perceptual features that guide children to perceptual attributes associated with category labels [Smith et al. (2002); Smith et al. (2010); Smith et al. (1996); Brady & Chun, 2007; Chun & Jiang, 1998]. Alternatively, the shape bias may be a strategic and conceptually-controlled mechanism, governed by general world knowledge and conceptual understanding; on this (A. E. Booth & Waxman, 2002; A. Booth, Waxman, & Huang, 2005). In this view, the bias is a heuristic whose magnitude might vary when the context of the generalization requires the child to attend to other features, such as the object's function or material.

Thus, variability in the shape bias is a key source of constraint on theories of its origins. However, the sources of this variability remain unclear. A recent study used statistical meta-analysis to estimate the effect size of the shape bias across 40 studies, finding a large average effect (0.8 standard deviations) (Abdelrahim & Frank, 2024). However, effects varied widely across seemingly similar studies, and the vast majority of the variance in the data remained unexplained by moderators such as age or language. One possibility is that methodological differences across studies contribute to the observed variation. For example, task format, stimuli design, and participant characteristics might influence the magnitude of the shape bias. In this paper, we explore the role of methodological factors in shaping the shape bias, using a series of experiments that systematically manipulate object properties and examine their effects on children's generalization.

Word extension studies typically teach children a novel label for a novel object, and then test them on the ability to extend it to other objects that share features like shape, or material with the target object. But these tasks vary in their dependent measure and their stimuli. With respect to dependent measures, children are typically either asked to make a forced choice between possible extensions, requiring a choice of the best generalization, or are asked to perform yes/no endorsement of different possible extensions, allowing multiple exemplars to be judged similar to the target (Landau et al., 1988). Some paradigms also allow children to reject all options. It is not yet clear how different tasks affect performance. For example, allowing children to select "none of those" reduces shape bias, especially with complex objects (Cimpian & Markman, 2005).

The second major methodological variation in the word extension findings comes from differences in stimuli. The key dimension of variation is the alternative possible dimensions of generalization beyond shape. Some studies contrast shape and other perceptual features, such as color or material, using stimuli designed to highlight these attributes. On the other hand, studies investigating conceptual understanding often include cues related to animacy, such as eyes, shoes, or other salient features, and use test objects that share multiple dimensions with the exemplar instead of only one (Susan S. Jones & Smith, 1998; Yoshida & Smith, 2003b). When functionality is emphasized, stimuli are often paired with demonstrations of an affordance, stories, or narratives to contrast shape with function. Children aged 2 to 5 years are found to prioritize shape, even when provided with functional information (Centner, 2003; Graham, Williams, & Huber, 1999; Landau & Jones, 1998; Merriman, Scott, & Marazita, 1993). However, conflicting evidence shows children sometimes prioritize function or other cues (Kemler Nelson, 1995; Gelman & Medin, 1993; etc.). In addition, some studies use pictures or drawings, while others use physical objects (cite).

Our goal in the current study is to evaluate procedural sources of variability as part of a larger scale assessment of word generalization across age groups and languages. We used a within-subject design to maximize precision in estimating condition differences. We also aim to measure developmental change by recruiting a sample size large enough to estimate age-related change. In Experiment 1, we began by measuring developmental change in the shpae bias. In Experiment 2, we manipulated the presence of alternative generalization items that shared function but not shape with the target object. Across both studies, we observed a robust shape bias that increased with age; preliminary data from Experiment 2 yielded no evidence that function information reduced the shape bias.

Experiment 1

Methods

Participants Twenty four typically developing English speaking participants (2-5 years old, mean=3.7, SD=1.29) were recruited from a local nursery school and children's museum in the US (preregistration link).

Materials To investigate children's reasoning about objects' properties and functions, a series of object sets were designed, each containing one exemplar, a material match, a shape match, a function match (used in Experiment 2), and a distractor (e.g., dax, fep, blicket, gorp, zimbo, wap, blint). Example objects are shown in Figure 1.

To investigate children's reasoning about objects' properties and functions, a series of object sets were designed, each containing one exemplar, a material match, a shape match, a function match, and a distractor (e.g., dax, fep, blicket, gorp, zimbo, wap, blint). These sets allowed for systematic manipulation of object features to assess various cognitive processes related to word learning and category generalization. In Experiment 1, the shape match was contrasted with a material match. This served as a baseline check and replication of prior findings regarding shape bias in categorization tasks. In Experiment 2, the same exemplar was used, but the shape match was contrasted with a function match.

The function test object was modified in a way that preserved its shape but altered its functionality (e.g., an object wrapped entirely vs. one that could clearly open). Color was excluded across all objects to ensure that visual similarity was driven solely by shape, material, and functional cues.

Exemplar	Shape match	Function (fx)	fx involves:
		holds small toys	whole structure
		scoops sand	part of the structure
		makes a heart	part of the structure
		holds water	reasoning about material
		grabs toys	reasoning about material

Figure 1: Examples of stimuli objects used in both experiments

Objects were crafted to explore how children reason about similarity based on whole-object vs. part-based features (e.g., whether specific parts afford a function). Some objects, such as the “fep,” “blint,” and “wap,” were designed with material-critical functions (e.g., holding water while made of a paper towel). This design tested whether children could prioritize material when reasoning about function and to capture the developmental changes. The degree to which object affordances were visually apparent varied across designs. For example,the “zimbo” was designed to afford functionality only through a specific part, while the overall structure was irrelevant. The “gorp” was modeled to resemble objects familiar to slightly older children, like scissors, allowing exploration of prior experiences’ influence on categorization.

Procedure On each trial, the child was shown an object, which was labeled with the phrase “this is a [X]”. The object was moved away from the child remained in view, and both test objects and the distractor were displayed simultaneously while asking the child “can you find another [X] by pointing to it?”. The child gets to hear the label 3 times while viewing it without touching it. Each child saw seven trials.

Results

Participants showed an overall shape bias across all trials (shape: 61%, material: 30%, distractor: 9%). Figure **??** shows a developmental shift to choose by shape by age 3, replicating what is seen previously in the literature.

To quantify these trends, we fit a generalized logistic mixed

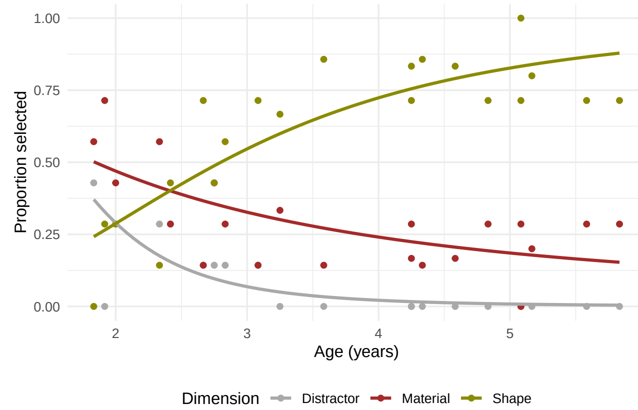

Figure 2: Developmental trend of choosing by each dimension. Smoothed lines are standard error

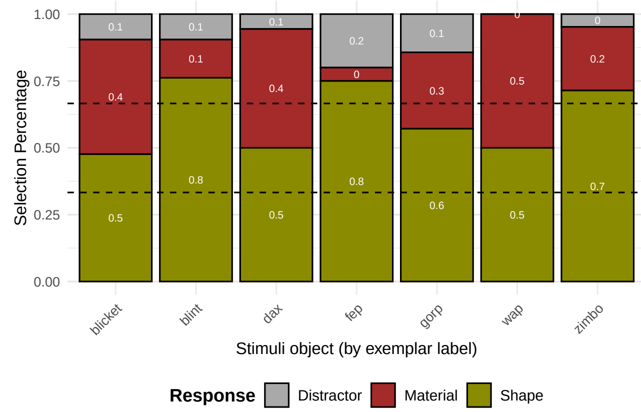

Figure 3: Percentage of choosing by dimension per stimuli item, indicated by its novel label (aggregate across participants). Dashed line is chance level = 33.3%

effects model predicting the binary outcome of choosing by shape for each trial by each child. We included random intercepts at the level of both participants and stimuli objects. The average intercept odds of choosing by shape were 0.11 (odds of 0.11:1 at the mean age, $p = < .001$), with a significant increase in odds of 1.06 per month increase in age ($p = < .001$). The model also shows variability at the item-level intercept (variance = 0.11, SD = 0.32) across 7 unique items (standardlabel groups).

Discussion

Even in a small sample of children between 2 and 5 years of age, data from this experiment confirmed the robustness of the shape bias. Although it is not clear why children below age 3 chose more by material, there was variability at the item level such that for three objects we saw above chance material choices when collapsing across ages.