

# How to Make a Proceedings Paper Submission

## Anonymous CogSci submission

### Abstract

Include no author information in the initial submission, to facilitate blind review. The abstract should be one paragraph, indented 1/8 inch on both sides, in 9~point font with single spacing. The heading ‘Abstract’ should be 10~point, bold, centered, with one line of space below it. This one-paragraph abstract section is required only for standard six page proceedings papers. Following the abstract should be a blank line, followed by the header ‘Keywords’ and a list of descriptive keywords separated by semicolons, all in 9~point font, as shown below.

**Keywords:** Add your choice of indexing terms or keywords; kindly use a semi-colon; between each term.

### Background

#### Basis of generalizations

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 perceptual features that expands with experience (CLARK, 1973). Over time, deeper attributes, such as its internal composition, behavior and interaction with the world, or sound, may also become relevant (Carey, 1985). The defining features of the lexical category manifest in the way the word is used to label other objects. Understanding how children generalize a noun learned in the presence of a few exemplars remains critical to deciphering the underlying processes of lexical concept formation. Overextensions are common in early word acquisition [Rescorla LA. Overextension in early language development], yet the nature of the underlying representations and their developmental trajectory remain debated. 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. In contrast, the functional core hypothesis (Nelson, 1974) posits that children extract probabilistic 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).

The shape bias is a word learning constraint that is argued to facilitate early noun acquisition, to be an important route to vocabulary growth, and found to be weaker in children with language delay (Jones, 2003; JONES & SMITH, 2005; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Nevertheless, the shape bias observed in word learning experiments is highly variable across ages, cultures, languages, and experimental conditions, which led to some theoretical debates.

### The cross-linguistic debate

The word extension task literature reveals cross linguistic differences in the prevalence of the shape bias that are thought to be theoretically important. For instance, East Asian languages’ speakers like Mandarin and Japanese, show less reliance on shape compared to English speakers in the US (Gathercole & Min (1997); Imai & Gentner (1997); Larissa K. Samuelson & Smith (1999); Nancy N. Soja, Carey, & Spelke (1991); Subrahmanyam & Chen (2006); Yoshida & Smith (2003b)]. Theoretical explanations of these differences can be summerized into two hypotehsis: it is either thought to reflect syntactic structure (e.g., count-mass syntax vs. classifier systems) (Larissa K. Samuelson, Horst, Schutte, & Dobbartin, 2008; Nancy N. Soja et al., 1991; Nancy N. Soja, Carey, & Spelke, 1992) or lexical and environmental statistical regularities in the language and the environment [Gershkoff-Stowe & Smith (2004); Larissa K. Samuelson (2002); Larissa K. Samuelson & Smith (1999); Perry, Samuelson, Malloy, & Schiffer (2010); Colunga & Smith (2000); Yoshida & Smith (2003a), Jara-Ettinger, Levy, Sakel, Huanca, & Gibson (2022); Abdelrahim & Frank (2024)).

### The Perception or conception debate

Another key debate centers on the mechanism underlying the shape bias. To illustarte this; We can think of the shape bias as an automatic attentional bias that tunes attention to only the relevant perceptual cues that have been probabilistically associated with naming events and learning lexical categories in the past [Smith et al. (2002); Smith, Colunga, & Yoshida (2010); Brady & Chun, 2007; Chun & Jiang, 1998). These perceptual cues are often correlated and integrated with the contextual cues of multiple sources like syntax, environment, and existing vocabulary. However, this perceptual bias are considered by some to constitute the actual representation of

the lexical category i.e. a dog is the shape of dog. On the other hand, it is considered by others to be a bias that acts as a heuristic to recognize objects, but it is controlled by general world knowledge and conceptual understanding that can be exerted upon this bias and override it when necessary. These debates highlight the complexity of the shape bias's origins and mechanisms, as well as the conflicting interpretations of existing evidence.

A meta-analytic effect size of 0.8, derived from over 300 standardized effects across 40 studies (Abdelrahim & Frank, 2024), confirms the robustness of the shape bias. However, substantial heterogeneity in the data (with over 90% of variance unexplained by age or language) suggests that cross-cultural, linguistic, and developmental differences remain masked.

## Sources of Heterogeneity

### Task format

Generalization in word learning is often assessed using the word extension task. Here, children are taught a novel label for a novel object and tested on their ability to extend it to other objects that share features like shape, or material, etc., with the exemplar object. Word extensions are often measured via Forced-choice tasks, which require restrictive generalizations, yes/no endorsement tasks, allowing broader acceptance of category membership which allows for different levels of similarity and difference (cite), or Open-choice tasks, enabling children to reject all options, assumed to indicate an understanding of category membership that goes beyond shared perceptual features (Cimpian et al 2005). Forced-choice tasks may yield different results compared to yes/no endorsement tasks, and allowing children to select “none of those” reduces shape bias, especially with complex objects (Cimpian, 2005). The choice of the task format is often guided by the theoretical framework of the researchers, leading to what seems like a circular stream of events in which theory informs task selection, and task selection confirms theory.

### Stimuli

A significant source of variation in the word extension findings comes from differences in stimuli. Just like task format, these methodological decisions are often driven by theoretical frameworks as well. For instance, studies focusing on low-level attentional biases typically emphasize contrasts between 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 to explore broader conceptual frameworks (cite). When functionality is emphasized, stimuli are often paired with demonstrations of an affordance, stories, or narratives to contrast shape with function. Children aged [X–Y] are frequently found to pri-

oritize shape, even when provided with functional information (Gentner & Rattermann, 1991; Woodward & Markman, 1998; etc.). However, conflicting evidence shows children sometimes prioritize function or other cues (Kemler Nelson, 1995; Gelman & Medin, 1993; etc.), with variation linked to factors like whether test objects were handled or how stimuli were designed (e.g., functional bases vs. appended parts). In addition, some studies use pictures or drawings, while others use physical objects (cite).

Lastly, most studies employ between-subjects designs, which do not control for individual differences, further amplifying heterogeneity. These procedural and stimuli variations reflect broader theoretical questions about the origins of the shape bias (Smith & Medin, 1981). Is it a Low-level attentional mechanisms driven by attentional processes that guide children to perceptual features associated with category labels? Or a Top-down conceptual processes in which the perceptual features act as identifiers rather than defining properties? Where should the line be drawn between perceptual feature identification and the core representation of conceptual labels? Are these separate processes, or do they exist on a continuum that develops as children acquire more information? How do attention to perceptual attributes and conceptual understanding interact during development (Madole & Oakes, 1999)? These foundational questions influence procedural decisions and should be kept in mind when investigating label categorization and concept formation.

## The nature of Knowledge

Given that task design is influenced by theoretical assumptions, this raises another important question: What type of knowledge do these tasks measure? Two key assumptions can stem out of this: Knowledge as Stable and Fixed: If knowledge is stable, tasks merely elicit pre-existing constructs. Investigating heterogeneity would then focus on ensuring task validity and reliability in capturing the theoretical construct. Knowledge as Dynamic and Task-Dependent: If knowledge is dynamic, categorization depends on the interaction between task specifics and children's behavior. This view suggests that children dynamically select information sources to organize categories, influenced by the context of the task and the nature of the stimuli (Smith et al., 2010). If the second assumption holds, achieving consistency in measures across studies is crucial to isolating the relevant contextual cues that tasks provide specially for cross-cultural studies that aim to adjudicate theoretical debates. Regardless of which assumption holds, it is very important to evaluate the heterogeneity in the word categorization studies which highlights the importance of methodological consistency and the need to consider the theoretical premises underlying procedural decisions. Understanding how these premises shape task designs and influence results is key to advancing our knowledge of label categorization and concept formation.

## Current Study

This project attempts to provide a first step into a larger scale assessment of word generalization across age groups and languages. The main goal of this study is to evaluate the effect of procedural design and stimuli design on the heterogeneity seen in word generalization tasks. We utilize a within-subject design, controlling for individual differences which we believe is important for a proper comparison between conditions that require giving different instructions and cues to the participants. We aim to recruit a larger sample size of a wider age group, with a variety of items and test trials. Given our focus on early language acquisition and the noun bias dominating early vocabulary (cite), we prioritize studies examining functional information over other types of conceptual knowledge.

### Stimuli design

Seven stimuli sets were hand crafted such that each set contains one exemplar, a material match, a shape match, a function match, and a distractor. In experiment 1, the shape match is contrasted with a material match, it served as a basic check and a replication of previous findings, while in experiment 2, the shape match is contrasted with a function match. The same exemplar is used in both experiments, and color is not shared across any of the objects. For the function test object, it is modified in a way that doesn't change the shape at all, but allows for the function (i.e. an object that is wrapped all over and a similar one that can clearly open). Objects are designed in a way that can answer to questions of how much children reason about the whole vs parts similarity, how they reason about material, and the influence of previous experiences with objects (see supplementary) (cite Neslon and graham and ).

### Adults ratings

#### Experiment 1: a preliminary

**Participants** Twenty four typically developing English speaking participants (2-5 years old, mean=44.05, SD=15.29) were recruited from a local nursery school and children's museum in the US.

**Procedure** Seven trials were conducted in which each participant sees an object being labeled "this is a dax", the object is taken away but still in view, both test objects and the distractor are displayed simultaneously while asking the child "can you find another dax by pointing to it?". The child gets to hear the label 3 times while viewing it without touching it.

**Results** Participants showed an overall shape bias across all trials (shape:61%, material:30%, distractor: 9%). Figure 1 shows a developmental shift to choose by shape by age 3, replicating what is seen previously in the literature. A generalized logistic mixed-effects model (GLMM) reveals an average intercept odds of 0.11 (odds of 0.11:1 at the mean age,  $p < .001$ ), with a significant increase in odds of 1.06 per unit 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). After

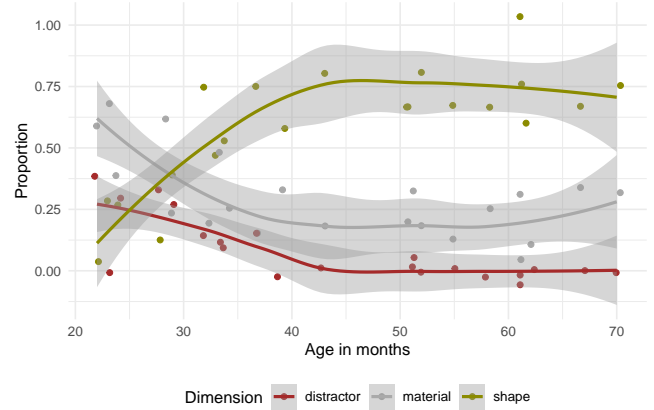


Figure 1: Effect size plotted by average participant age. Points indicate individual effect, and error bars show variance. Points are colored based on language, and model fits for Indo-European and non-Indo-European languages are shown in black.

replicating the shape bias effect using the set of stimuli we created in a simple set up, our next experiment explores an design that tests for both conditions when shape is only contrasted with material without any additional information, and a condition in which shape is contrasted with function after demonstrating the function for the exemplar, while controlling for individual differences with a bigger sample size to capture variability at the item level.

### Experiment 2

**Participants** 31 (target n=96, 24 per each age group) participants between 2-5 years old (mean=48.22, SD=5.54, n per age group) were recruited from a local nursery school in the US.

**Procedure** A within subject manipulation with two conditions: material or function. The material condition is identical to the first experiment. In the function condition, the experimenter introduce the exemplar object "this is a dax", gives the child 15 seconds to play with it, provides functional information "the dax grapes toys", gives another 15 seconds to play with it, and puts the toy away but within view, before introducing the test objects and asks for a response.

**Preliminary results** Similar to what is conveyed in Figure 2, A generalized logistic mixed-effects model (GLMM) showed baseline odds of 2.25 (odds of 2.25:1 for the function condition  $p = .693$ ). The odds increase with age by a factor of 1.01 ( $p = .742$ ). While The odds for the "material" condition are 0.07 times lower than for the "function" condition ( $p = .338$ ). Additionally, the interaction between age and condition shows that for the "material" condition, the odds decrease by a factor of 0.94 per unit increase in age ( $p = .294$ ). Random effects indicate variability in intercepts across participants (SD = 0.22) and across items (SD = 1.38) for 31

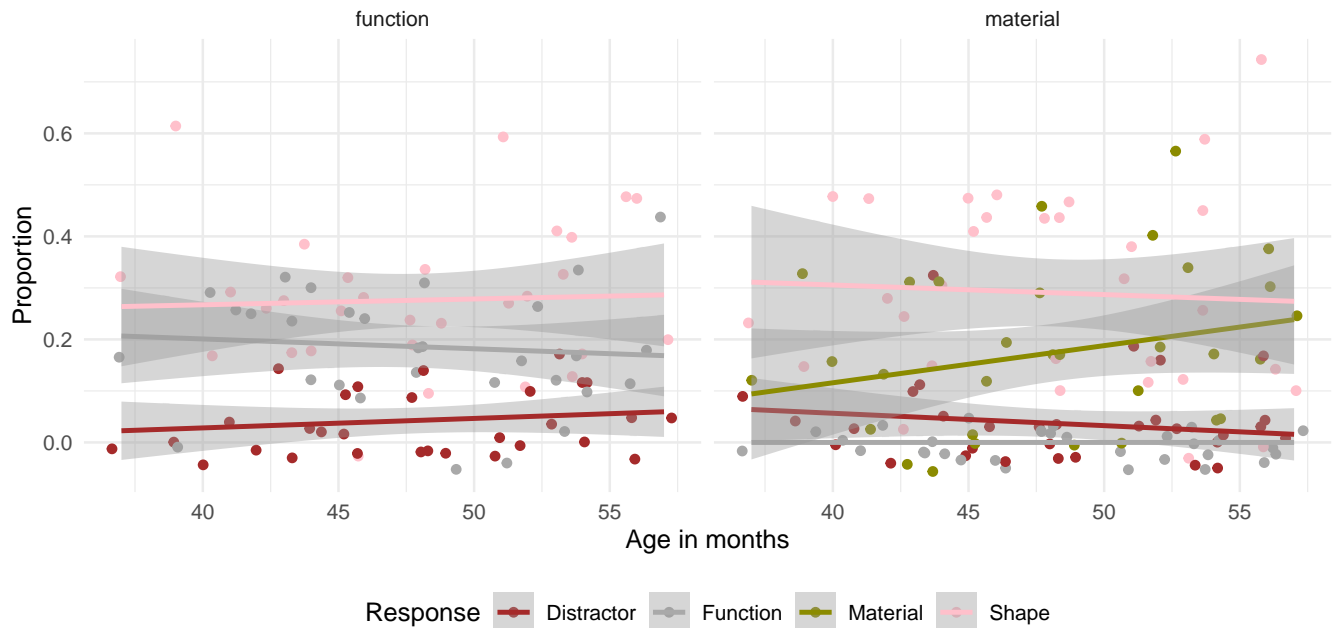


Figure 2: Effect size plotted by average participant age. Points indicate individual effect, and error bars show variance. Points are colored based on language, and model fits for Indo-European and non-Indo-European languages are shown in black.

participants and 7 items. Notably, the confidence intervals show uncertainty however data collection is still ongoing.

## Discussion

The word extension and category organization literature is highly heterogeneous. Studies in this domain lack an integrative and commensurable design, which hinders our ability to draw consistent conclusions. To achieve a more accurate measurement of category organization and concept learning, we need a reliable and valid wide range set of stimuli objects, consistent task formats and test designs, as well as multi-site cross-cultural experiments unified across laboratories to maximally account for the variability. Our evaluation of the word extension literature reveals that making procedural decisions, which we think are likely a primary source of unexplained variability, is unattainable without running a series of controlled experiments that would allow us to systematically assess how different designs and stimuli covary with response patterns. In our initial results, we see a strong tendency to generalize by shape, even in conditions designed to make function salient. This suggests that, even with a potential saliency effect, where the trials highlighted functional information, it failed to override the preference for shape-based choices. In addition, many children explored whether their chosen test object could perform the intended function after selecting it based on shape. This behavior implies that the shape-based selection might not reflect a disregard for functional information but rather a hypothesis that objects sharing shape might also share functionality.

What does it mean to talk about validity and reliability in

a task like the word extension one? What is the construct we are talking about here. It is a measure at the group level rather than at the individual level >> this kid had this level of shape bias. >> is it an issue of the signal is not even there, or is it a matter of the level/degree of the signal when it comes to cross-cultural work. (Madole, K. L., & Oakes, L. M. (1999)): Although the distinction between perceptual and conceptual categories makes a certain intuitive sense, it may only confuse our attempts to understand psychological reality. Establishing a reliable metric for this distinction is extremely difficult and attempts to operationalize the terms perceptual and conceptual are invariably highly ambiguous and task-specific. How the perceptual vs conceptual debate maps into the cross cultural differences debate? As we mentioned in the introduction, procedural variation observed in the literature have followed theoretical debates and in fact reinforced by them. Thus, evaluating the theories given the state of evidence is not feasible and it will end up crashing with this heterogeneity and with the broader question on the type of knowledge being highlighted by each one of the evidence.

## References

- 10 Abdelrahim, S., & Frank, M. C. (2024, September). Examining the robustness and generalizability of the shape bias: A meta-analysis. *PsyArXiv*. <http://doi.org/10.31234/osf.io/3by54>
- Baldwin, D. A. (1992). Clarifying the role of shape in children's taxonomic assumption. *Journal of Experimental Child Psychology*, 54 3, 392–416. Retrieved from <https://api.semanticscholar.org/CorpusID:>

- CLARK, E. V. (1973). WHAT'S IN a WORD? ON THE CHILD'S ACQUISITION OF SEMANTICS IN HIS FIRST LANGUAGE. This research was supported in part by NSF grant GS-1880 to the language universals project, Stanford University, and in part by NSF grant GS-30040 to the author. In T. E. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 65–110). San Diego: Academic Press. <http://doi.org/https://doi.org/10.1016/B978-0-12-505850-6.50009-8>
- Colunga, E., & Smith, L. B. (2000). Learning to learn words: A cross-linguistic study of the shape and material biases. In *Proceedings of the 24th annual Boston University conference on language development* (Vol. 1, pp. 197–207).
- Gathercole, V. C. M., & Min, H. (1997). Word meaning biases or language-specific effects? Evidence from English, Spanish and Korean. *First Language*, 17(51), 031–56. <http://doi.org/10.1177/014272379701705102>
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098–1114.
- Graham, S. A., & Diesendruck, G. (2010). Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. *Cognitive Development*, 25(2), 111–123. <http://doi.org/10.1016/j.cogdev.2009.06.002>
- Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language*, 26, 295–320. Retrieved from <https://api.semanticscholar.org/CorpusID:43424185>
- Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2), 169–200.
- Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development*, 9(1), 45–75. [http://doi.org/10.1016/0885-2014\(94\)90019-1](http://doi.org/10.1016/0885-2014(94)90019-1)
- Jara-Ettinger, J., Levy, R., Sakel, J., Huanca, T., & Gibson, E. (2022). The origins of the shape bias: Evidence from the tsimane'. *Journal of Experimental Psychology: General*.
- Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Developmental Science*, 6(5), 477–483. <http://doi.org/https://doi.org/10.1111/1467-7687.00304>
- JONES, S. S., & SMITH, L. B. (2005). Object name learning and object perception: A deficit in late talkers. *Journal of Child Language*, 32(1), 223–240. <http://doi.org/10.1017/S0305000904006646>
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299–321. [http://doi.org/https://doi.org/10.1016/0885-2014\(88\)90014-7](http://doi.org/https://doi.org/10.1016/0885-2014(88)90014-7)
- Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, 81, 267–285. Retrieved from <https://api.semanticscholar.org/CorpusID:143965074>
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, 21(12), 1894–1902.
- Samuelson, Larissa K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, 38(6), 1016–1037. <http://doi.org/10.1037/0012-1649.38.6.1016>
- Samuelson, Larissa K., Horst, J. S., Schutte, A. R., & Dobbertin, B. N. (2008). Rigid thinking about deformables: Do children sometimes overgeneralize the shape bias? *Journal of Child Language*, 35(3), 559–589.
- Samuelson, Larissa K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1–33.
- Samuelson, Larissa K., & Smith, L. B. (2000). Children's attention to rigid and deformable shape in naming and non-naming tasks. *Child Development*, 71(6), 1555–1570.
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, 34(7), 1287–1314. <http://doi.org/https://doi.org/10.1111/j.1551-6709.2010.01130.x>
- 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(1), 13–19. <http://doi.org/10.1111/1467-9280.00403>
- Soja, Nancy N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38(2), 179–211.
- Soja, Nancy N., Carey, S., & Spelke, E. S. (1992). Perception, ontology, and word meaning. *Cognition*, 45(1), 101–107. [http://doi.org/10.1016/0010-0277\(92\)90025-D](http://doi.org/10.1016/0010-0277(92)90025-D)
- Subrahmanyam, K., & Chen, H.-H. N. (2006). A crosslinguistic study of children's noun learning: The case of object and substance words. *First Language*, 26(2), 141–160. <http://doi.org/10.1177/0142723706060744>
- Yoshida, H., & Smith, L. B. (2003a). Known and novel noun extensions: Attention at two levels of abstraction. *Child Development*, 74(2), 564–577.
- Yoshida, H., & Smith, L. B. (2003b). Shifting ontological boundaries: How Japanese and English-speaking children generalize names for animals and artifacts. *Developmental Science*, 6(1), 1–17.