Children learn number words slowly because they don't identify number as relevant to linguistic meaning

Katharine A. Tillman[1,3], Katie Wagner[1,4], & David Barner[1,2,5]

University of California, San Diego
Departments of Psychology[1] and Linguistics[2]
[3]katillman@ucsd.edu; [4]kwagner@ucsd.edu; [5]barner@ucsd.edu

Abstract

Children learn number words slowly, acquiring exact meanings for their first words in sequence, with many months in between words. The long delays are surprising in light of evidence that infants can discriminate, e.g., sets of 2 from 3. Here, we test the hypothesis that, rather than facing a perceptual problem, children have difficulty identifying number as the dimension of meaning encoded by an adjective like "three." We trained children on an unknown number word in the context of a proper noun (a giraffe named "*Mr. 3*" with three spots), and found that 1- and 2-knowers were later better at identifying the giraffe from a lineup, relative to children who had heard the same giraffe described with an adjective ("with *three* spots"). These results support the hypothesis that identifying number as a dimension of meaning, rather than visual discriminability or salience, is a bottleneck on early number word learning.

Keywords: number cognition; number words; word learning; cognitive development; abstract concepts

Introduction

Learning number words is a first step toward engagement in formal arithmetic, and poses a difficult challenge for young children. Even after children learn to recite the count list ("*one, two, three*…") around age 2, it takes several years for them to demonstrate an understanding of the logic of counting (e.g., Wynn, 1990, 1992; Le Corre & Carey, 2007). Prior to learning how the counting procedure is used to generate precise cardinalities, children pass through a series of well-documented stages, in which they acquire exact meanings for their first three or four number words in sequence, and learn to give appropriate amounts when asked to do so in a task that has become known as "Give-a-Number" or "Give-N" (Schaeffer et al., 1974; Wynn, 1990, 1992). During this time, they transition from being "non-knowers" (who have meanings for no number words and give random amounts), to "one-knowers" (who have an exact meaning for *one*, and give one when asked for one, but not for larger numbers, to "two-knowers" (who have exact meanings for *one* and *two*), to "three-knowers", and then, sometimes, "four-knowers" before eventually learning that counting while pointing at objects can be used to identify the cardinality of sets ("CP-knowers"). Interestingly, there are long delays between these stages; for example, it takes roughly half a year between acquiring an exact meaning for "one" and an exact meaning for "two." Why is it that a child who has already figured out how to give sets of exactly 1 and 2, and can recite the count-list all the way to 10 or higher, nonetheless struggles for months before being able to produce a set of 3 in response to requests for "three"?

In the present study, we investigated the cause of these delays. In particular, we sought to test whether the bottleneck that children face in acquiring early number words is due to (1) noise in the perceptual representations of larger sets, or (2) difficulty identifying which aspect of a stimulus number words encode.[1] To do this, we introduced a linguistic intervention in which we attempted to train children on the meaning of an unknown number word - *three* - by presenting it as a proper noun rather than as an adjective. Consistent with the second hypothesis, we show that this manipulation improves learning of "three" in 1- and 2-knowers: Children use numerical features to differentiate referents of proper nouns like Mr. 3 and Mr. 2, but struggle to use the identical information to define the meanings of number words *qua* adjectives.

Despite nearly 40 years of studies using the Give-N paradigm to explore children's number knowledge, it remains unclear why children's early number word learning is sequential and so protracted. While the onset and duration of the "subset-knower" stages (i.e., 1-, 2-, 3-, and 4-knower) varies across cultures, the same stage-like progression has been observed in children around the world (e.g., Almoammer et al., 2013; Barner, Chow, & Yang, 2009; Barner, Libenson, Cheung, & Takasaki, 2009; Le Corre et al., 2016; Piantadosi, Jara-Ettinger, & Gibson, 2014; Sarnecka et al., 2007). While many studies attempting to solve this problem have focused on the critical transition from subset- to CP-knower, sparking debates about exactly what new knowledge CP-knowers acquire (e.g., Davidson, Eng, & Barner, 2012; Sarnecka & Carey, 2008), here, we focus on the puzzle of why are there such long delays between the subset knower stages.

[1] A third possibility, which we do not test here, is that children have to construct a new number concept for each new number word they learn. Interestingly, Wagner et al. (2015) presented evidence from bilingual learners suggesting that the difficulty children face is *not* one of constructing new concepts, but of appropriately mapping linguistic labels to number concepts they already possess. We focus here on the issue of why this mapping problem takes so long for children to solve.

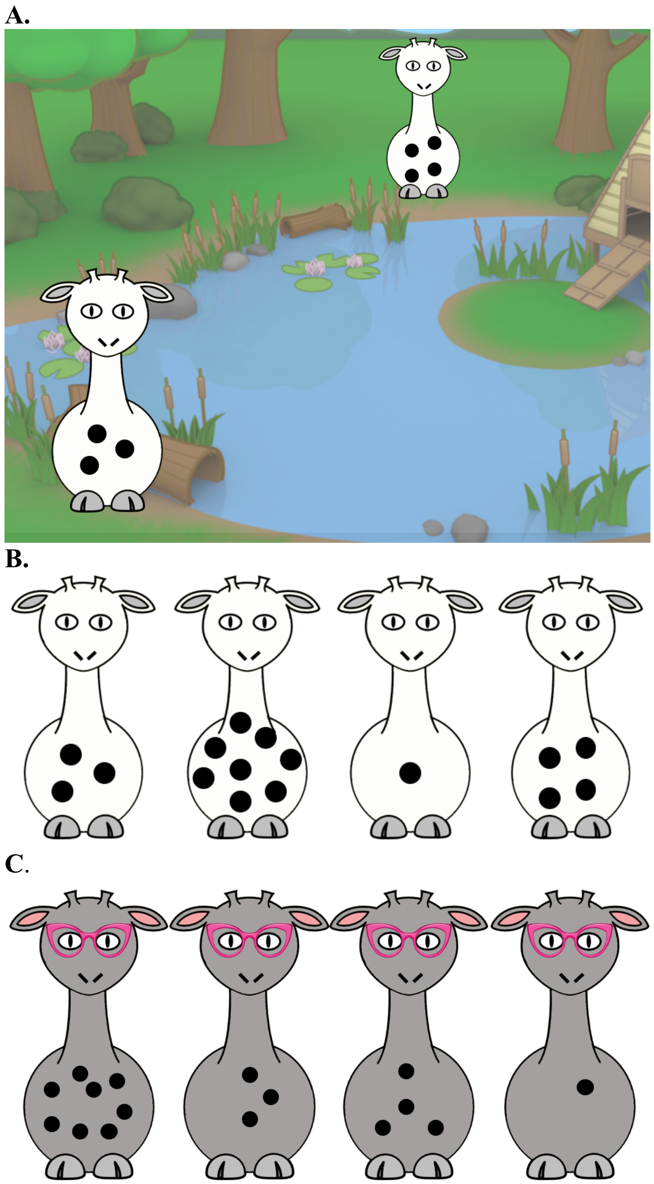

Figure 1: A. Example slide from training story, which was verbally accompanied by, in the Proper Noun condition, "Next, Mr. 3 [point] came to a bench by the road. Look at his tummy! It has spots on it! On the other side of the bench, he saw his friend under a tree. This one [point] is *not* Mr. 3; it's his friend. He has spots too, but they're different!" B. Example "line-up" of giraffes in Learning trials. C. Example "line-up" of grandmother giraffes in Transfer trials.

One possible explanation for the delays between learning number words is related to perception: Children may simply have a harder time discriminating sets of 2 vs. 3 than sets of 1 vs. 2, and a harder time yet discriminating 3 vs. 4, and so on. If so, these perceptual limitations may hinder the child's ability to form mappings between number words and exact quantities. Importantly, while previous studies indicate that infants are already able to discriminate small numbers, these studies are often limited to reporting whether groups of infants succeed or fail at small number comparisons, while remaining neutral with respect to whether some comparisons within the small number range might be more difficult than others (e.g., Feigenson & Carey, 2003). Therefore, these findings do not address whether the increased difficulty of discrimination might explain delays between the acquisition of number words.

Furthermore, even if children are able to make the relevant visual discriminations between set sizes, they may nonetheless struggle to encode number because they do not spontaneously *attend* to it when processing visual scenes; they may be biased to attend to the shape or function of objects rather than their membership in sets, and the properties of these sets like cardinality. In other words, even if the visual difference between a set of 2 objects and a set of 3 objects is detectable, this property of the visual stimulus ("threeness") may not be not particularly salient to the child, who may be more likely to attend to properties such as shape (e.g., Landau, Smith, & Jones, 1988). We call this hypothesis, that attention or perception limits number word learning, the *Perceptual Bottleneck Hypothesis*.

A second possibility is that children can perceive and attend to the relative quantities well enough, but nonetheless fail to identify the property of number as the domain of meaning for number-word learning. In other words, even if numerical quantities present in a scene are perfectly salient, children may not identify number as relevant to the problem of learning new words, and therefore to learning a number word like "three." We call the hypothesis that children fail to abstract the property of number as the dimension of meaning for a new word the *Meaning Abstraction Hypothesis*.

In order to test these possibilities, we leveraged a previously-attested difference in children's interpretation of proper nouns and adjectives. Previous studies of color-word learning have shown that children are sensitive to the fact that adjectives label properties while proper nouns label unique individuals that can be differentiated according to these same properties. In one study by Soja (1994), 2-year-olds who had not yet learned any color words were nevertheless able to use the perceptual property of color to identify the referent of a proper noun (Soja, 1994). For example, when shown two dinosaurs that differed only in color, children learned that the red one (and not, e.g., the blue one) was named Emily, even if they couldn't identify the referent of the word *red*. These results indicate that children who can perceptually discriminate, e.g., red from other colors and use this to learn proper names, nonetheless struggle to identify color as the stimulus dimension being labeled, and to form the intended association between the word *red* and its target hue.

Can children who have not yet learned the meanings for small number words nonetheless use numerical properties to learn proper nouns? If so, this might suggest that the bottleneck children face when moving from one knower-level to the next is not a failure of attention or perception, but instead one of identifying numerosity as relevant to the word's lexical meaning. On the Perceptual Bottleneck Hypothesis, if children's difficulty with learning number words is due to a failure to discriminate or attend to numerical quantities, then they should struggle equally when learning adjectives and proper nouns. However, on the Meaning Abstraction Hypothesis, if children struggle to identify number as relevant to word-learning, then they may successfully learn proper nouns that depend on discrimination and attention to specific quantities, while failing to learn adjectives that directly encode these same properties. To test this, we attempted to teach children the word "three" as either a proper noun, "*Mr. 3,*" or an adjective, the "giraffe with *three* spots." While no prior studies have taught children number words as proper nouns, one trained 2-knowers to identify a set of "*three* dogs" (Huang, Spelke, & Snedeker, 2010). However, they found no evidence of transfer to other sets of 3, and did not attempt to teach number words more than one unit higher than a child's knower-level. Here, to also explore whether proper-noun training could allow children to "skip" a knower-level, we trained both 1- and 2-knowers on "three."

Method

Participants

We conducted two experiments to test our hypotheses. However, as their methods and the pattern of results they generated were extremely similar, we present combined data from both experiments here. In total, 132 2- to 3-year-old children, identified as 1-knowers or 2-knowers, were included in this sample. Sixty-one children participated in the Adjective condition (*M* age = 3.1 years, range = 2.1-4.0 years), and 71 participated in the Proper Noun condition (*M* age = 2.9 years, range = 2.2-3.8 years).[2] Children in the Adjective condition included 31 1-knowers and 30 2-knowers. Children in the Proper Noun condition included 39 1-knowers and 32 2-knowers.

An additional 254 children were tested but not included in these analyses due to not being 1- or 2-knowers (n = 179), failure to complete the task (n = 30), experimenter error (n = 25), failing the "giraffe pre-test" described below (n = 8), being outside the target age range (n = 5), parent interference (n = 3), speaking a primary language other than English (n = 3), and developmental delay (n = 1).

Procedure

Give-N pre-test. To determine their knower levels prior to training, children first performed a titrated version of the Give-a-Number task (see Wynn, 1990, 1992) in which they were asked to put different numbers of toys on a plate on each trial. Only 1-knowers or 2-knowers continued.[3]

Giraffe pre-test. After the Give-N pre-test, a subset of children (n = 89) were also presented with two "giraffe pre-test" trials, in order to test whether they were able to correctly identify the target 3-spotted giraffe prior to training. On each trial, similar to the Learning test trials described below, the child was presented with a 4-alternative forced choice among cartoon giraffes with 1, 3, 4, or 8 spots (see Fig 1B). Depending on condition, children were asked either to point to the giraffe they thought was "Mr. 3" or "the giraffe with three spots" No feedback was given. After both trials were complete, the child was shown the target alone and told that this one was [Mr. 3/the giraffe with three spots]. Children who answered both giraffe pre-test questions correctly (n = 3 in the Proper Noun condition; n = 5 in the Adjective condition) were excluded from all subsequent analyses, as we could not be sure their performance in the test phase was influenced by training.

Training. In training phase, children viewed a slideshow on a computer while the experimenter told a story featuring the 3-spotted giraffe (see Fig. 1A for example screen). Seven times over the course of the story, children in the Proper Noun condition heard the experimenter label the giraffe as "Mr. 3" while their attention was drawn to the dots on his stomach ("This is Mr. 3! Look at his belly! It has spots on it."). Children in the Adjective condition heard an identical story in which the giraffe was instead labelled as "The giraffe with three spots." In the story, during a walk to his grandmother's house, the target giraffe encountered three giraffe "friends," with 1, 4, and 8 spots, once each. These distractors were labelled contrastively with the target ("This is not Mr. 3; it's his friend"). On the final slide, the giraffe arrived at his grandmother's house, and her face (but not body) was shown in the window.

Test. *Learning.* Immediately following training, the experimenter said, "That's the end of the story. Now I need your help. Can you find [Mr. 3/the giraffe with three spots]?" On each of four "Learning" trials, the child was presented with a line-up of giraffes, including the target and the three distractors from the story (Fig 1B). On each trial, the child was asked to point to [Mr. 3/the giraffe with three spots]. *Transfer.* Next, the child completed four "Transfer trials" in which they were asked to point to either "Mr. 3's grandma" in the Proper Noun condition or "the grandma with three spots" in the Adjective condition. Again, a lineup of giraffes with 1, 3, 4, and 8 spots was presented (Fig 1C). The configurations of dots differed from those used on Mr. 3 and his friends, though the spots were the same size and color. Recall that the child never saw the spots on the grandmother giraffe during training.

All children completed the Learning trials prior to the Transfer trials, but the order of the 4 trials within each test phase, and the positioning of the 4 giraffes within each trial, were counterbalanced across subjects.

Give-N post-test. After completing the Transfer trials, most children (n = 118) also completed a second iteration of the Give-N task. The procedure was identical to the Give-N pre-test, except that a different type of toy object was used.

[2] The unequal sample sizes were due to a clerical error resulting in additional children being tested, exceeding our target N.

[3] In one experiment, non-knowers were also tested, but these data are not reported here.

Results

Learning

First, we asked whether children in the Proper Noun condition were more likely to learn the identity of the three-spotted giraffe than those in the Adjective condition, as predicted by the Meaning Abstraction Hypothesis. As shown in Figure 2A, in the Proper Noun condition, 1-knowers correctly identified the target on 44% (s.e.m. = 6%) of trials, while 2-knowers did so on 68% (s.e.m = 6%) of trials. Meanwhile, in the Adjective condition, 1-knowers correctly identified the target on 33% of trials (s.e.m = 6%), while 2-knowers did so on 52% of trials (s.e.m. = 7%). To test whether Proper Noun training improved children's learning, we conducted a mixed-effects logistic regression predicting the likelihood of choosing the correct target, using Knower Level (1-knowers vs. 2-knowers) and Condition (Proper Noun vs. Adjective) as predictors, and including both an interaction term and a random effect of Subjects. We found that the effect of Knower Level significantly improved the fit of this model. $X^2(1) = 11.2$, $p < 0.001$, as did the effect of Condition, $X^2(1) = 4.6$, $p = 0.03$, but that their interaction did not, $X^2(1) = 0.16$, $p = 0.7$. In other words, 2-knowers were better learners than were 1-knowers, and children in the Proper Noun condition were better learners than those in the Adjective condition.

Next, we asked whether children's performance was significantly better than would be predicted by random guessing. Because each trial was a 4-alternative forced-choice, chance was defined as 25% correct. In the Proper Noun condition, both 1- and 2-knowers performed significantly greater than chance, indicating that the training had an effect (Wilcoxon signed-ranks test, both p's < 0.01). In the Adjective condition, 1-knowers' performance was not significantly greater than chance, $p = 0.17$, but 2-knowers' performance was greater than chance, $p < 0.01$.

Taken together, our findings on the Learning trials show that, as predicted by the Meaning Abstraction Hypothesis, children in the Proper Noun condition were more likely to learn the identity of "Mr. 3" than children in the Adjective condition were to learn which giraffe was the one "with three spots." This was true despite the fact that the perceptual demands on the task were equal in both conditions. Moreover, only Proper Noun training allowed 1-knowers to identify "three," while their behavior was consistent with random guessing in the Adjective condition.

Transfer

Next, in order to assess whether children's knowledge of "three" extended beyond the specific exemplar they were trained on, we examined performance on the Transfer trials. As shown in Figure 2B, we found that, in the Proper Noun condition, 1-knowers chose "Mr. 3's grandma" on 27% (s.e.m. = 4%) of trials, while 2-knowers chose the target grandmother on 42% (s.e.m. = 5%) of trials. In the Adjective condition, 1-knowers chose "the grandma with three spots"

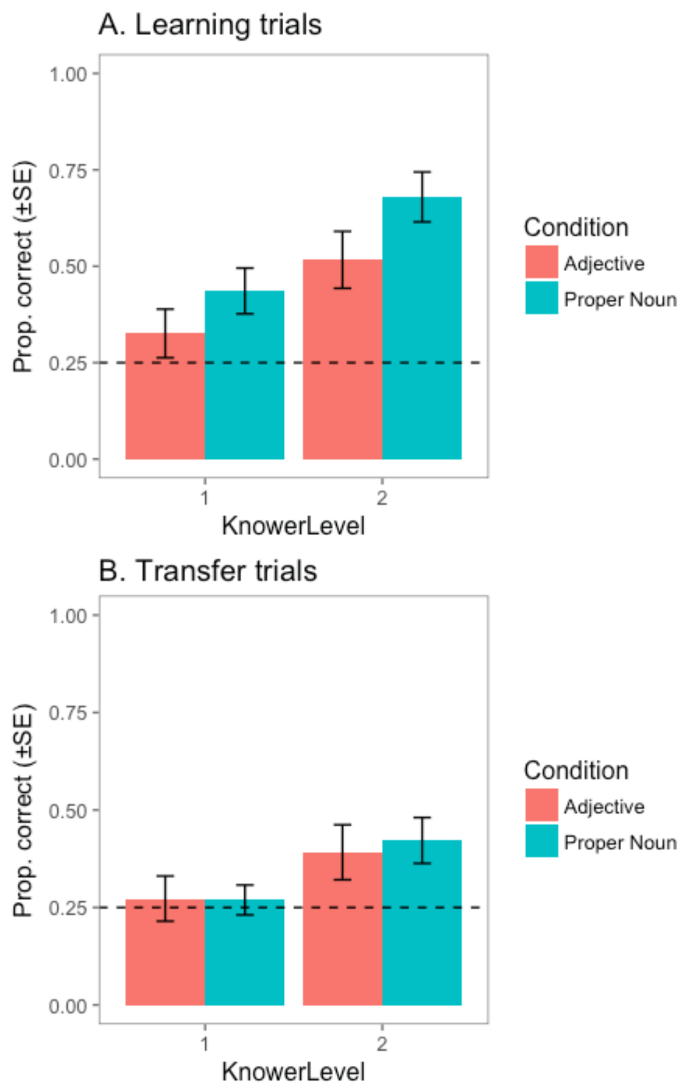

Figure 2: A. Proportion of trials on which children correctly identified either "Mr. 3" in the Proper Noun condition or "the giraffe with three spots" in the Adjective condition after training. B. Proportion of trials in which children transferred their knowledge to correctly select either "Mr. 3's grandma" in the Proper Noun condition or "the grandma with three spots" in the Adjective condition.

on 27% of trials (s.e.m.= 6%), while 2-knowers chose her on 39% (s.e.m. = 7%) of trials. A mixed-effects logistic regression with the same effects structure as that used on the Learning trials indicated that the effect of Knower Level significantly improved the fit of the model, $X^2(1) = 6.0$, $p = 0.01$. However, the effect of Condition (Proper Noun vs. Adjective) did not, $X^2(1) = 0.1$, $p = 0.7$, and there was no interaction, $X^2(1) = 0.04$, $p = 0.8$. Furthermore, we found that, in both conditions, 1-knowers' performance was no greater than chance (Wilcoxon signed ranks tests, both p's > 0.2), while 2-knowers' performance was significantly better than chance, both p's < 0.01. In other words, only 2-knowers showed evidence of transfer, and children were equally (un)likely to transfer the meaning of 'three' to other exemplars whether they were trained on the term as an adjective or a proper noun.

Give-N post-test

Although the prior analyses indicate that many children who are not yet 3-knowers can nonetheless be trained to identify "three," this leaves open the question of whether such training changes a child's knower level, i.e., teaching the child to give exactly 3 objects when asked for 3, and not to give 3 when asked for other numbers. Following the training and test phases of the study, 118 children completed the Give-N post-test. Table 1 shows the relationship between pre- and post-training performance on Give-N. Strikingly, irrespective of their performance on the Learning and Transfer trials, only 5% of children (n = 6) were classified as 3-knowers or above after completing this study. The majority of children, 69%, were classified as having the same knower-level at both time points. Of those whose knower-level changed (n = 36), 39% improved, while 61% declined.

Table 1: Knower-levels before and after training on "three"

Condition	KL	N	Post-test knower level					
			0	1	2	3	4	CP
Proper Noun	1	34	5	24	5	0	0	0
	2	30	1	7	20	0	2	0
Adjective	1	30	6	21	3	0	0	0
	2	23	2	1	16	2	1	1

KL = knower-level prior to training.

Errors

We also performed an exploratory analysis of the types of errors children made. While the Perceptual Bottleneck Hypothesis predicts that children would be most likely to err by choosing the 4-spotted giraffe, which was most perceptually similar to the target, the Meaning Abstraction hypothesis makes no specific predictions about children's errors. Inconsistent with the Perceptual Bottleneck Hypothesis, 1-knowers chose the 4-spotted giraffe on 32% of error trials, and 2-knowers chose the 4-spotted giraffe on 34% of error trials, neither of which were significantly different from chance (33%; both $p > 0.6$).

Discussion

Why it is that a child who already has exact meanings for *one* and *two*, and can recite the count list to 10 or above, nonetheless require several more months to acquire an exact meaning for *three?* Here, we investigated this question by exploring two candidate hypotheses. According to the Perceptual Bottleneck Hypothesis, the bottleneck on early number word learning may be due to perceptual factors: either that children have a harder time visually discriminating larger arrays (e.g., 3) from their neighbors in the count-list relative to smaller arrays (e.g., 1 or 2), or that, even if discrimination is not an issue, children simply don't visually attend to the property of number when viewing a new scene. According to the Meaning Abstraction Hypothesis, children's protracted learning of new number words may instead stem from difficulties identifying the relevant dimension of the stimulus being labelled for them. For instance, a child might hear a phrase like "three spots" and think that "three" refers to the size, shape, or color of each spot, rather than the quantity of the set.

In support of the Meaning Abstraction hypothesis, we found that a linguistic intervention in which the number word "three" was presented in the context a proper noun, *Mr. 3*, helped enable both 1- and 2-knowers identify its referent, in this case a 3-spotted giraffe. Critically, unlike the adjective "three," the proper noun "Mr. 3" refers to an individual, rather than to the property of a set. In principle, children did not need to know that Mr. 3's name referenced the quantity of three dots on his tummy in order to learn his name. Our finding that removing the need to identify quantity as the relevant dimension of meaning reduced the difficulty of the task for children suggests that they may have difficulty inferring the correct referents of number words as they are typically used, as adjectives.

Because the giraffes used in our studies differed only in the quantity of spots on their tummies, in order to learn the identity of either "Mr. 3" (in the Proper Noun condition) or "the giraffe with three spots" (in the Adjective condition), the child needed to attend to the spots and perceptually discriminate three dots from, e.g., four dots. If the limiting factor on number word learning were perceptual, we would expect equally poor performance in both conditions[4]. Instead, we found that children in the Proper Noun condition were more likely to identify the correct giraffe those in the Adjective condition.

We also found large effects of knower-level: 2-knowers were more likely to learn and extend the meaning of "three" than were 1-knowers. While 1-knowers in the Adjective condition performed no better than chance overall, 2-knowers succeeded. While both groups performed better than chance in the Proper Noun condition, the 2-knowers outperformed 1-knowers. Furthermore, regardless of training condition, only 2-knowers extended their learning of "three" to another exemplar. Interestingly, prior studies suggest that while n-knowers do not, by definition, have exact meanings for n+1, they may nonetheless have inexact, lower-bounded meanings for n+1 (e.g., Barner & Bachrach, 2010). If so, having a partial meaning for "three" might help explain why 2-knowers performed better in the Adjective condition than did

[4]Note that perceptual limitations are equal in both conditions regardless of the particular perceptual strategy children deploy. However, it is possible that the Proper Noun condition prompted more children to attend to the overall triangle shape formed by the dots, as opposed to their quantity per se, and this perceptual grouping could have aided their memory.

1-knowers. It remains more mysterious, however, why 2-knowers *also* outperformed 1-knowers in the Proper Noun condition, since success in this condition should not, in principle, require knowing any number words. One possibility is that that more 2-knowers than 1-knowers picked up on the connection between Mr. 3's name and the number of spots on his tummy, and that existing partial knowledge of "three" helped guide their choices in the test trials. On the other hand, our finding that 2-knowers in the Proper Noun condition still outperformed 2-knowers in the Adjective condition may suggest that not all 2-knowers have partial meanings for three, or that an inexact meaning is not always sufficient to map "three" to a precise quantity in this task.

While children in the Proper Noun condition had an advantage over those in the Adjective condition in the Learning trials, this was not the case on the Transfer trials. The lack of a Proper Noun advantage on these trials may be related to the specificity of names. Another possibility is that children lacked the intuition that that "Mr. 3's grandma" might share some of his perceptual features (i.e., the same number of spots). If so, this genetic inference component of the task could have masked a Proper Noun advantage we might have otherwise observed in the Transfer trials.

A prior study by Huang, Spelke, and Snedeker (2010) also attempted to train subset-knowers on number words above their knower-level. In that study, n-knowers were trained on sets of $n + 1$. Consistent with the results of our Adjective condition, they found that 2-knowers could be trained to identify particular sets of three, like "three dogs." However, 2-knowers did not generalize "three" reliably to other sets, such as "three cows." Huang et al. took their findings in 2-knowers as evidence that they had failed to form a mapping between "three" and an abstract numerical quantity during training, but, instead created an association with a specific stimulus. Here, we *did* find evidence that 2-knowers in the Adjective condition (but not 1-knowers) generalized the meaning of "three," although our transfer was 'closer': from one giraffe/set of spots to another, rather than across noun categories.

In addition to the long delays between subset-knower stages, a second puzzle regarding early number-word learning is why number words are always learned in sequence. our knowledge, our Proper Noun condition represents the only successful attempt in the literature to train n-knowers on a number word higher than $n + 1$. The finding that removing the need to represent quantity as the dimension of meaning allowed 1-knowers to learn "three" also suggests that the stage-like progression through the knower-levels may reflect difficulties identifying the intended referents for number words, rather than the increased perceptual similarity between larger sets that differ by one.

References

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *PNAS, 110*(46), 18448-18453

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cogn Psychol*, *60*(1), 40-62.

Barner, D., Chow, K., & Yang, S. J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cogn Psychol, 58*(2), 195-219.

Barner, D., Libenson, A., Cheung, P., & Takasaki, M. (2009). Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *J. Exp Child Psychol*, *103*(4), 421-440.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162-173.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Dev Sci*, *6*(5), 568-584.

Huang, Y. T., Spelke, E., & Snedeker, J. (2010). When is four far more than three? Children's generalization of newly acquired number words. *Psychol Sci*, *21*(4), 600-606.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cogn Dev*, *3*(3), 299-321.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395-438.

Le Corre, M., Li, P., Huang, B. H., Jia, G., & Carey, S. (2016). Numerical morphology supports early number word learning: Evidence from a comparison of young Mandarin and English learners. *Cogn Psychol*, *88*, 162-186.

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. *Dev Sci*, *17*(4), 553-563.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662-674.

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of 'one','two', and 'three'in English, Russian, and Japanese. *Cogn Psychol*, *55*(2), 136-168.

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cogn Psychol*, *6*(3), 357-379.

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Dev*, *65*(3), 918-937.

Wagner, K., Kimura, K., Cheung, P., & Barner, D. (2015). Why is number word learning hard? Evidence from bilingual learners. *Cogn Psychol*, *83*, 1-21.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155-193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cogn Psychol*, *24*(2), 220-251.