Psychological Science

http://pss.sagepub.com/

When Is *Four* Far More Than *Three*? : Children's Generalization of Newly Acquired Number Words

Yi Ting Huang, Elizabeth Spelke and Jesse Snedeker

Psychological Science 2010 21: 600 originally published online 9 March 2010
DOI: 10.1177/0956797610363552

Research Article

Psychological Science
21(4) 600–606

DOI: 10.1177/0956797610363552
http://pss.sagepub.com

When Is *Four* Far More Than *Three*? Children's Generalization of Newly Acquired Number Words

Yi Ting Huang, Elizabeth Spelke, and Jesse Snedeker
Harvard University

Abstract

What is the relationship between children's first number words and number concepts? We used training tasks to explore children's interpretation of number words as they acquired the words' meanings. Children who had mastered the meanings of only the first two or three number words were systematically provided with varied input on the next word-to-quantity mapping, and their extension of the newly trained word was assessed across a variety of test items. Children who had already mastered number words to *three* generalized training on *four* to new objects and nouns, such that their representation of the newly learned number was approximate. In contrast, children who had mastered only *one* and *two* learned to apply *three* reliably within a single count-noun context (e.g., *three dogs*), but did not generalize training to new objects labeled with different nouns (e.g., *three cows*). Both findings suggest that children fail to map newly learned words in their counting routine to the fully abstract concepts of natural numbers.

Keywords

counting, word learning, numerical concepts, cognitive development

Received 5/29/09; Revision accepted 8/26/09

The relationship between number words and concepts has drawn substantial attention from cognitive and developmental psychologists (Carey, 2009; Dehaene, 1997; Gelman & Gallistel, 1978, 2004; Mix, 2002; Wynn & Bloom, 1997) in light of two curious observations. First, children learn the verbal counting list before they understand that number words refer to specific, unique, and exact cardinal values (Condry & Spelke, 2008; LeCorre & Carey, 2007). Second, children learn number-word meanings very slowly (Wynn, 1990, 1992b), in contrast with the meanings of many other kinds of words (e.g., Carey, 1978). When asked for a specific number of objects (the "give-*N*" task), most 2-year-old children produce one object when asked for *one*, but produce no consistent amount when asked for larger numbers (*one-knowers*). By 2.5 years of age, most children give two objects when asked for *two*, but grab a handful for larger quantities (*two-knowers*). Several months later, children respond appropriately to *three* (*three-knowers*), and by their fourth birthday, most children master the logic of verbal counting.

What hypotheses do children entertain in learning the meanings of number words? From the start of number-word learning, children might hypothesize that each number word maps onto an abstract numerical magnitude (Dehaene, 1997; Gelman & Gallistel, 1978; Wynn, 1998). Thus, a child who is taught the meaning of *two* in one context would apply the word to any set of two individuals regardless of their kind. This possibility gains plausibility from findings that infants show capacities to enumerate visible objects, sounds, and actions (see Feigenson, Dehaene, & Spelke, 2004) and to detect numerical correspondences between sets of objects and sounds (Izard, Sann, Spelke, & Streri, 2009; Jordan & Brannon, 2006). Alternatively, children might initially learn more narrow number-word meanings: A child who learns *two* in one context (e.g., *two dogs* applied to a set of two dogs) may extend the word only to objects (e.g., to pairs of horses, but not pairs of sounds or actions) or only to entities named by the same count noun (i.e., to other dogs). This second possibility is consistent with findings demonstrating that children often fail to apply number words broadly across distinct contexts (Mix, 1999, 2002; Mix, Huttenlocher, & Levine, 1996).

Corresponding Author:
Yi Ting Huang, Harvard University, Psychology Department, 33 Kirkland St., Cambridge, MA 02138
E-mail: ythuang@email.unc.edu

In the present research, we attempted to tease apart these two accounts of early number-word meanings by exploring children's interpretation of newly trained number words. Training paradigms have been used productively to probe children's mapping of concepts onto adjectives and spatial terms (Kiblanoff & Waxman, 2000; Shusterman & Spelke, 2003). Such paradigms may be particularly useful for studying number-word learning because they can reveal intermediary conceptual representations during the acquisition process (Griffin & Case, 1996; Siegler, 2007). In three experiments, we trained children who had mastered number words up to either *two* or *three* on the next word in their count list. We then tested what meanings the children attributed to these words by assessing their generalization of the trained word to new entities and new linguistic contexts.

Experiment 1

In Experiment 1, we trained two-knowers and three-knowers on the next word-to-quantity mapping (*three* or *four)* using pictures of animals as training stimuli. Then we tested for a very limited form of generalization: Could children extend the trained word to pictures of different kinds of animals?

Method

Participants. All the children were monolingual English speakers and were accompanied by a parent. From a sample of 38 children, 16 children were categorized as two-knowers on the give-*N* task (mean age = 3 years 2 months, range = 2 years 6 months–3 years 6 months; 8 boys and 8 girls), and 16 were categorized as three-knowers (mean age = 3 years 7 months, range = 3 years 2 months–3 years 9 months; 8 boys and 8 girls). The remaining children were categorized as one-knowers or four-knowers and were not tested further.

Counting. The children were given 10 objects and encouraged to count them. All the children produced the count list to *ten* without error.

Give-N task. The children then completed the give-*N* task in order to categorize whether they were two-knowers or three-knowers. The children were shown small plastic fish and were asked to put different quantities from one to six into a basket ("Can you make ___ fish jump into the pond?"). The experimenter began by asking for "one fish" and continued on to higher numbers in a pseudorandom order. When a child failed to produce a quantity correctly, the experimenter asked for the number directly below it before returning to the failed number. If the children produced the correct quantity for the failed number on the second attempt, they were asked for the number a third time to determine their maximum level of reliable knowledge (knower level). Each child was assigned to a card-pair training condition based on his or her knower level.

Two-knower training and testing. During training, two-knowers were shown cards depicting eight different kinds of animals. First, they saw two trials in which a single card featuring three animals was labeled with a count phrase (e.g., "This card has three cows!"). Next, they were shown six trials in which a card with *three* was contrasted with a card depicting another quantity (e.g., "This card has three birds!" and "This card does not have three birds!"). These contrasts included numbers that the children had mastered (1 or 2) and numbers that they had yet to master (4, 5, 6, or 10). The arrays of animals varied in their spatial arrangement (rows vs. triangles) but contained objects of constant size and shape. Thus, continuous variables such as summed area and summed contour length were correlated with number. In the final phase of training, children were given the same card pairs again and were asked to select from each pair the card with three items (e.g., "Can you give me the card with three birds?"). Errors were infrequent and were corrected.

During the test phase, two-knowers were shown 10 new card pairs, each featuring new kinds of animals, and were asked to select the member of each pair that matched the number word indicated by the experimenter. On two noncritical trials (*known-known trials*), the card pairs contrasted two known quantities (1 vs. 2: e.g., "Can you give me the card with two horses?"). On four additional noncritical trials (*trained-known trials*), a card with three animals was paired with a known quantity (3 vs. 1 or 2). On the four critical trials (*trained-unknown trials*), a card with three animals was paired with a card showing a quantity children did not have a word for (3 vs. 4, 5, 6, or 10, for a total of four contrast ratios). Regardless of which number *three* was paired with, children were asked for *three* (e.g., "Can you give me the card with three pigs?"). To discourage responding on the basis of nonnumerical information, we varied both the arrangements and the sizes of the items in the paired test cards such that the two quantities in a pair were matched for total continuous extent (e.g., three large chickens vs. five smaller chickens).

Three-knower training and testing. The procedure and controls for three-knowers were similar to those for two-knowers, except that children were trained and tested on *four*. The known numbers used in training and testing were 1 to 3, and the unknown numbers used in training and testing were 5, 6, 10, and 16.

Results and discussion

Preliminary analyses revealed no effects of gender. All analyses were therefore collapsed over this variable.

Two-knowers. Two-knowers performed well on known-known trials, accuracy = 89%, $t(15) = 7.01$, $p < .001$, $d = 3.62$, and trained-known trials, accuracy = 91%, $t(15) = 8.06$, $p < .001$, $d = 4.16$. However, they performed at chance levels on the trained-unknown trials, accuracy = 47%, $t(15) = 0.44$,

$p > .60$, $d = 0.23$. A one-way analysis of variance (ANOVA) revealed a significant difference across the trial types, $F(2, 30) = 17.02$, $p < .001$, $\eta^2 = .53$. Performance did not differ between the known-known and trained-known trials ($p > .80$) and was significantly better on these trials than on the trained-unknown trials (both *p*s < .01). Within the trained-unknown trials, there were no reliable differences in performance across the different comparison quantities ($p > .30$).

The failure of two-knowers to generalize numbers to different kinds of animals on the critical trials contrasts with their consistent selection of the correct cards during the final phase of training. It is unlikely that this discrepancy reflects a memory failure, because the transfer test immediately followed training. Instead, it suggests that two-knowers employed one of two strategies. First, they may have mapped *three* onto the exact features of the corresponding training card, without extracting a more general relation between *three* and a numerical value. Second, children's generalizations may have been restricted to particular count nouns (or object classes) that were evaluated as a single unit during the training phase (e.g., "three dogs" or "three fish"). Experiment 3 explored these possibilities.

Three-knowers. Three-knowers performed at ceiling on known-known trials, accuracy = 100%, Wilcoxon signed-rank statistic = 136, $z = 3.49$, $p < .001$, and well above chance on trained-known trials, accuracy = 84%, $t(15) = 7.86$, $p < .001$, $d = 4.06$, and trained-unknown trials, accuracy = 71%, $t(15) = 3.77$, $p < .01$, $d = 1.95$. Nevertheless, there were reliable differences across the three trial types, $F(2, 30) = 12.25$, $p < .001$, $\eta^2 = .45$. Performance was better on known-known trials than on either type of trial with the trained number (*p*s < .001), and did not differ between the latter two trial types ($p > .10$).

The performance of three-knowers on the critical, trained-unknown trials was influenced by the numerical magnitude of the contrasting array, $F(3, 45) = 4.03$, $p < .05$, $\eta^2 = .21$ (see Fig. 1). The children reliably selected the correct card when 4 was paired with 10 or 16 ($p < .01$) but not when it was paired with 5 or 6 ($p > .30$). The children's failure to select the card with 4 objects over the card with 6 objects is striking, because infants can discriminate between arrays of 4 and 6 objects on the basis of number (Xu & Arriaga, 2007). Our finding suggests that children mapped the newly trained word *four* onto a highly imprecise representation of number.

Thus, from this brief training procedure, three-knowers were able to extract an interpretation of *four* that generalized from trained sets of animals to new sets of animals (and to new count nouns). Experiment 2 investigated whether three-knowers will generalize *four* more broadly from pictured animals to solid artifacts.

Experiment 2

In Experiment 2, three-knowers were familiarized with the meaning of *four* using the same card-pair training procedure used in Experiment 1, but they were tested with sets of concrete, household objects. The test objects therefore differed from the training stimuli in both their spatial and tactile properties and their ontological status (animals vs. artifacts). If children's initial interpretation of *four* is sufficiently abstract, they should generalize the concept across these features, such

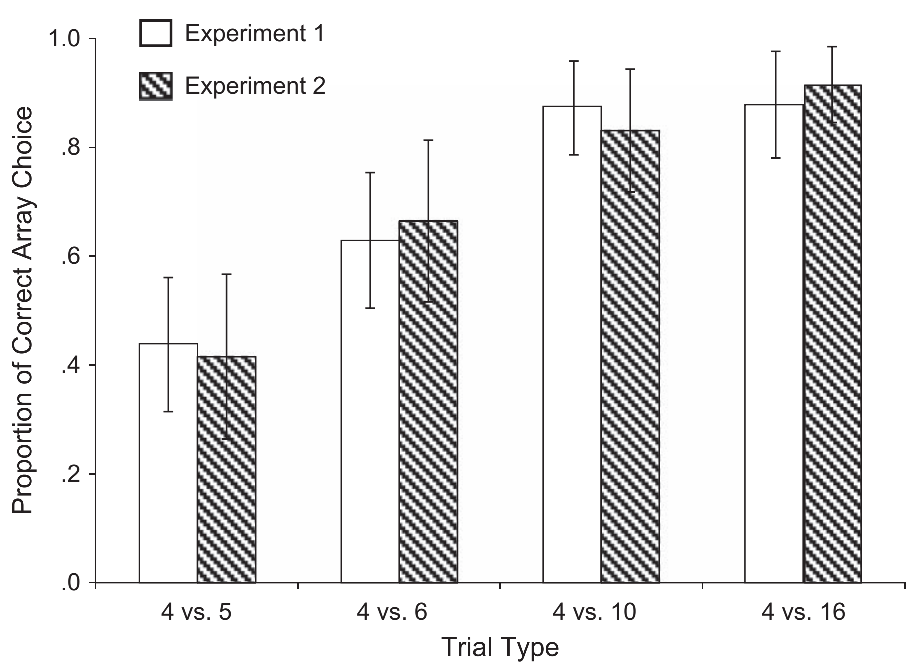

Fig. 1. Mean array-choice accuracy of three-knowers in Experiments 1 and 2, as a function of the ratio of the correct choice (the array showing 4 animals) to the incorrect choice (an array showing 5, 6, 10, or 16 animals). Error bars represent standard errors of the mean.

that performance in Experiment 2 would be similar to performance in Experiment 1. If children's notion of the next word-to-quantity mapping is restricted to a more narrow conceptual domain (in this case, animals) or to more superficial properties of the pictures we used as exemplars, then generalization would be less robust in Experiment 2 than in Experiment 1.

Method

From a sample of 29 children, we selected the first 12 who counted to *ten* and were categorized as three-knowers on the give-*N* task (mean age = 3 years 6 months, range = 3 years 1 months–3 years 10 months; 7 boys and 5 girls).

The training method was identical to that used for three-knowers in Experiment 1. The testing procedure was also the same as in Experiment 1, except that the children were tested with seven sets of objects (e.g., coins, pencils) pasted onto cardboard panels and only trained-known and trained-unknown trials were included. On the critical, trained-unknown trials, the paired sets of objects were approximately matched in surface area (e.g., 4 large vs. 10 small LEGOs).

Results and discussion

After training, children successfully selected four objects on trained-known trials, accuracy = 86%, $t(11) = 7.39$, $p < .001$, $d = 4.46$, and on trained-unknown trials, accuracy = 70%, $t(11) = 2.69$, $p < .05$, $d = 1.62$. Children performed as well in Experiment 2 as they did in Experiment 1: An ANOVA with experiment and trial type as factors revealed better performance when *four* was paired with a known than with an unknown number, $F(1, 26) = 6.47$, $p < .05$, $\eta^2 = .20$, but identified no main effect of experiment or interaction ($ps > .60$).[1] Thus, three-knowers acquired a word-to-quantity mapping for *four* that generalized from pictures to concrete objects.

As in Experiment 1, the performance of three-knowers on the trained-unknown trials was influenced by the numerical magnitude of the contrasting array, $F(3, 33) = 2.95$, $p < .05$, $\eta^2 = .26$ (see Fig. 1). Children reliably selected the correct array when 4 was paired with 10 or 16 ($ps < .05$) but not when it was paired with 5 or 6 ($ps > .20$). Experiment 2 therefore replicated the finding that children generalize the newly trained word *four* to nearby but discriminably different numerosities.

In summary, three-knowers successfully generalize *four* not only to novel pictures of animals, but also to three-dimensional artifacts. By the time children become three-knowers, therefore, their initial interpretation of a new number word generalizes to a fairly broad range of items. Children's greater success on test pairs with larger numerical differences suggests, nevertheless, that these initial number-word meanings are imprecise.

In Experiment 3, we returned to the mysterious performance of the two-knowers in Experiment 1. Did the children's success in the training phase and failure in the test phase arise because they memorized the training cards, or because their generalization of *three* was restricted to particular object categories, designated by particular noun phrases?

Experiment 3

Experiment 3 tested whether two-knowers' initial interpretation of *three* is restricted to particular lexical or conceptual contexts. Young children's interpretation of newly learned adjectives shows such a pattern of conservative generalization. When 3-year-olds hear a bumpy horse described with a novel adjective (e.g., "a very blickish horse"), they successfully generalize the adjective to other bumpy horses, but not to bumpy animals from different basic-level categories, such as rhinoceroses (Klibanoff & Waxman, 2000). Perhaps children's initial meanings for number words are similarly restricted. To explore this question, we compared two-knowers' generalization of a trained number word to novel test materials from the same category versus a different category.

Method

From the same sample as in Experiment 2, we selected the first 16 children who counted to *ten* and were categorized as two-knowers by the give-*N* task (mean age = 3 years 1 months, range = 2 years 3 months–3 years 5 months; 9 boys and 7 girls).

During training, two-knowers saw multiple target cards presenting the same picture of three small dogs arranged in a triangle. These target cards were paired with cards showing larger sets of dogs (4, 5, or 10), and all cards were labeled with respect to *three*, the trained number ("This card has [does not have] three dogs!"). Following this demonstration, children were presented with the same card pairs a second time and were asked to select from each pair the card with the trained number. Errors were infrequent and were corrected.

During testing, children were presented with 12 card pairs, each comprising a target card, which presented 3 items, and a distractor card, which presented a higher, unknown number of items (4, 5, or 10). There were four trial types, defined by the kind of target card: This card showed (a) the original target (small dogs in a triangle), (b) an array transformed in size and spatial configuration (large dogs in a row), (c) an array of different items from the same basic-level category (dogs of a different breed in a triangle), or (d) an array of animals from a different basic-level category (small sheep in a triangle). In all cases, the distractor card contained items of the same kind and size as the target card, but the number of items was larger, and they were in a different configuration. For the first three trial types, children were asked for the card with "three dogs." For the fourth, they were asked for the card with "three sheep." The four trial types were blocked, and the presentation order of the blocks was randomized between subjects.

Results and discussion

A one-way ANOVA revealed a significant difference in children's performance across the four trial types, $F(3, 45) = 7.09$, $p < .01$, $\eta^2 = .32$ (see Fig. 2). Although children performed equally well when the target card showed the original target items, their size and configuration variant, and the within-category

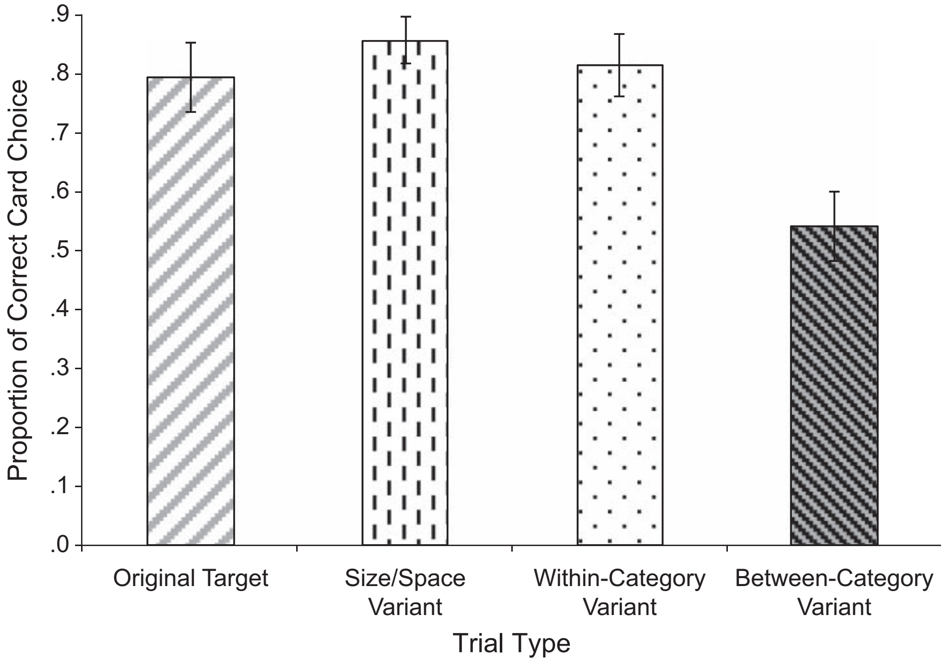

Fig. 2. Mean card-choice accuracy of two-knowers in Experiment 3, as a function of trial type. In this experiment, the target was a card used in training, a card showing the same items but in a different size and configuration, a card depicting different animals from the same basic-level category, or a card depicting animals from a different basic-level category. The distractor card on a given trial contained items of the same kind and size as the target card, but the number of items was larger, and they were in a different configuration. Error bars represent standard errors of the mean.

variant (all *p*s > .40), they performed significantly worse when the target card showed the between-category variant (all *p*s < .01). Children reliably identified *three* when presented with the original target cards, $t(15) = 4.88$, $p < .001$, $d = 2.52$; when presented with new cards in which the target items had a different size and spatial arrangement, $t(15) = 6.78$, $p < .001$, $d = 3.50$; and when presented with new cards depicting objects from a different subordinate class within the same basic-level category, $t(15) = 6.00$, $p < .001$, $d = 3.10$. However, children performed at chance level when the cards depicted animals from a different basic-level category, $t(15) = 0.68$, $p > .50$, $d = 0.35$. These findings provide evidence that two-knowers' initial interpretation of *three* is limited to the particular category or noun that is quantified.

After collapsing the data across the four trial types, we again found no effect of numerical distance on correct card selection ($p > .60$; see Fig. 3).[2] Unlike three-knowers, two-knowers did not appear to map the meaning of their trained number word to an approximate numerical magnitude. Instead, they learned to apply "three dogs" to arrays of exactly three dogs, regardless of their size, spatial configuration, or breed.

General Discussion

Three experiments explored children's hypotheses about the meanings of new number words. Our findings highlight two striking patterns. First, children who had mastered the meanings of number words up to *three* acquired a fairly broad understanding of the meaning of *four* after training under restricted conditions. When they were shown that *four* applied to pictured sets of animals, they readily generalized the word to new kinds of animals and even to solid artifacts. These children, however, generalized *four* in an approximate manner in both experiments, applying the word to sets of five or six objects despite contrastive training with these numbers. This pattern is

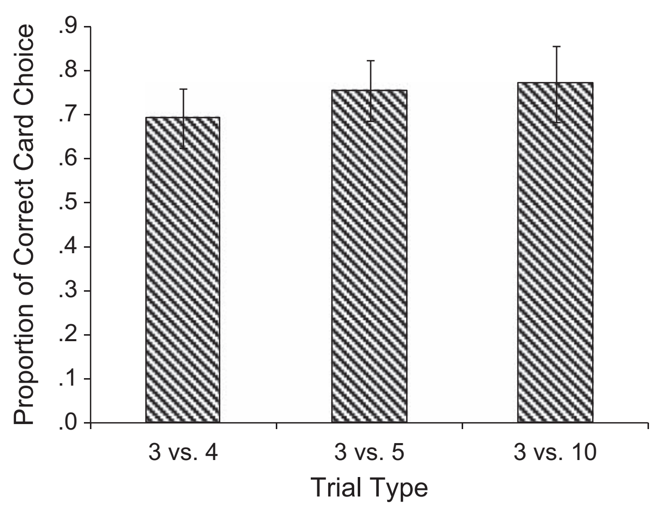

Fig. 3. Mean card-choice accuracy of two-knowers in Experiment 3, as a function of the ratio of the correct choice (the card showing 3 animals) to the incorrect choice (a card showing 4, 5, or 10 animals). Error bars represent standard errors of the mean.

unlikely to stem from counting errors: The participating children did not count out loud in either the give-*N* task or the test phase. Furthermore, counting errors would be expected to produce a generalization gradient around the correct value of four, and not categorical rejection of three with acceptance of five and six. Instead, the children's performance suggests that they mapped the trained word onto an approximate numerical representation similar to the representations found in animals, infants, and adults across diverse cultures (Feigenson et al., 2004).

Second, children who had mastered the meanings of only *one* and *two* were extremely limited in their generalization of *three*. When trained on multiple kinds of animals, they showed no generalization to new kinds of animals. Moreover, when trained on a single kind of animal presented in a single configuration (three dogs in a small triangle), they learned to apply *three* to arrays of dogs of novel sizes, spatial arrangements, and breeds, but not to arrays of sheep. The limited performance of two-knowers is surprising: These children counted reliably to *ten*, thus producing the word *three* in the same context in which they produced *one* and *two*. Two-knowers also apply *one* and *two* to diverse entities, including both objects and actions (Wynn, 1990). Finally, across all knower levels, children's ability to produce a particular number of sounds is strongly predicted by their performance in the give-*N* task with objects (Huang, Snedeker, & Spelke, 2005). Nevertheless, two-knowers' narrow generalization of *three* posttraining suggests that their understanding differs qualitatively from that of three-knowers or adults.

This pattern of limited generalization lends itself to two distinct explanations. First, two-knowers may initially map the entire quantified phrase (e.g., *three dogs*) to a holistic representation of its meaning. In linguistic theories, numbers do not have referents, but rather are functions that take nouns to yield quantified phrases (which may have referents). To extract the number's meaning from this semantic structure, the child might have to learn several such phrases to isolate the common element. Too much input variability could prevent the initial holistic mappings, and too little input variability could hinder subsequent reanalysis.[3] This hypothesis is consistent with research highlighting the importance of linguistic context, and nouns in particular, in the acquisition of adjectives and verbs (Gillette, Gleitman, Gleitman, & Lederer, 1999; Waxman & Booth, 2001).

Second, two-knowers may extract the number word from the phrase but initially map it to a representation that includes information about basic-level object kinds. A numerically relevant representation with precisely this property has been proposed to account for infants' ability to track objects over movement and occlusion (see Xu & Carey, 1996). In very young children, this "object file" system has a capacity limit of three items (Feigenson, Carey, & Hauser, 2002; Wood & Spelke, 2005; Wynn, 1992a) and consequently could provide possible meanings for *three* but not for *four* during number-word acquisition. Moreover, it expresses quantities only implicitly in terms of individuals and their properties: An array of two dogs is expressed as [DOG, DOG] by this system. Thus, two-knowers who mapped *three* to the representation [DOG, DOG, DOG] could have inferred that the term applies only to cases involving these individuals. This hypothesis is consistent with the centrality of basic-level concepts in young children's cognition (Rosch & Mervis, 1975).

In contrast, the three-knowers' approximate generalization of *four* suggests the use of a second conceptual system that represents larger, approximate magnitudes. This system supports infant computations of large quantities across sensory domains (Brannon, 2002; Lipton & Spelke, 2003; Wood & Spelke, 2005; Xu & Spelke, 2000) and guides children's understanding of number words before they learn verbal counting (Shusterman, Carey, & Spelke, 2009; Wagner & Johnson, 2009). The set-size limit on object-file representations might lead children to shift from one representational system to the other between *three* and *four*, allowing three-knowers to entertain a broader hypothesis about the scope of the next word-to-quantity mapping. However, the second conceptual system does not provide exact representations of numerosity (Dahaene, 1997), so children would generalize the newly learned word *four* to nearby magnitudes.

The present findings may provide insight into the slow pace of children's number-word acquisition. In the absence of a single system of exact numerical representation, children cannot simply map number words onto existing concepts. Instead, they have to create conceptual representations that go beyond either of the two supporting systems (Carey, 2009). Further studies using the present training methods may help to specify how children construct such new conceptual representations.

Acknowledgments

This work benefited from conversations with members of the Laboratory for Developmental Studies at Harvard University and extensive, insightful comments from Susan Carey.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The research was supported by National Institutes of Health Grant HD23103 to E.S. and by National Science Foundation Grant BCS-0623845 to J.S.

Notes

1. The comparison between the experiments had sufficient power to detect a moderate effect of stimulus type (λ = .80 for a mean difference in accuracy of 16%, collapsed across the trial types).
2. Note that an effect of the size observed in three-knowers (mean difference in accuracy = 34%) would have been detected with virtual certainty (λ = .97).
3. Thus, the two-knowers in Experiment 1 may have failed to make the initial narrow mappings because of either input variability or an inability to extend the mappings they made to novel categories or nouns.

References

Brannon, E.M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*, 223–240.

Carey, S. (1978). The child as word learner. In M. Halle, I. Bresnan, & G.A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Condry, K.F., & Spelke, E.S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*, 22–38.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford, England: Oxford University Press.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science*, *13*, 150–156.

Feigenson, L., Dehaene, S., & Spelke, S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314.

Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gelman, R., & Gallistel, C.R. (2004). Language and the origin of numerical concepts. *Science*, *306*, 441–443.

Gillette, J., Gleitman, H., Gleitman, L.R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135–176.

Griffin, S., & Case, R. (1996). Evaluating the breadth and depth of training effects, when central conceptual structures are taught. *Monographs of the Society for Research in Child Development*, *61*(1–2, Serial No. 246), 83–102.

Huang, Y., Snedeker, J., & Spelke, E. (2005, August). *Two dogs and two barks: How abstract are children's number words?* Paper presented at the 10th meeting of the International Congress for the Study of Child Language, Berlin, Germany.

Izard, V., Sann, C., Spelke, E.S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, USA*, *106*, 10382–10385.

Jordan, K.E., & Brannon, E.M. (2006). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences, USA*, *103*, 3486–3489.

Klibanoff, R.S., & Waxman, S.R. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development*, *71*, 649–659.

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

Lipton, J.S., & Spelke, E.S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science*, *14*, 396–401.

Mix, K.S. (1999). Preschoolers' recognition of numerical equivalence: Sequential sets. *Journal of Experimental Child Psychology*, *74*, 309–332.

Mix, K.S. (2002). The construction of number concepts. *Cognitive Development*, *17*, 1345–1363.

Mix, K.S., Huttenlocher, J., & Levine, S.C. (1996). Do preschool children recognize auditory-visual numerical correspondences? *Child Development*, *67*, 1592–1608.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Shusterman, A., Carey, S., & Spelke, E. (2009, April). *Two paths in the development of counting and cardinality*. Paper presented at the biennial meeting of the Society for Research on Child Development, Denver, CO.

Shusterman, A., & Spelke, E. (2003, April). *Spatial language and spatial reorientation: A training study*. Poster presented at the biennial meeting of the Society for Research on Child Development, Tampa, FL.

Siegler, R.S. (2007). Cognitive variability. *Developmental Science*, *10*, 104–109.

Wagner, J.B., & Johnson, S.C. (2009, April). *Analog magnitude representations influence preschoolers counting development*. Poster presented at the biennial meeting of the Society for Research on Child Development, Denver, CO.

Waxman, S.R., & Booth, A.E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*, 217–242.

Wood, J., & Spelke, E. (2005). Infants' enumeration of actions: Numerical discrimination and its signature limits. *Developmental Science*, *8*, 173–181.

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

Wynn, K. (1992a). Addition and subtraction by human infants. *Nature*, *358*, 749–750.

Wynn, K. (1992b). Children's acquisition of number words and the counting system. *Cognitive Psychology*, *24*, 220–251.

Wynn, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Sciences*, *2*, 296–303.

Wynn, K., & Bloom, P. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, *24*, 511–533.

Xu, F., & Arriaga, R.I. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology*, *25*, 103–108.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*, 111–153.

Xu, F., & Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.