When is *four* far more than *three*? Children's generalization of newly acquired number words

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

Acknowledgments: This work benefited from conversations with members of the Laboratory for Developmental Studies and extensive, insightful comments from Susan Carey. The research was supported by NIH Grant HD23103 to ESS and NSF-BCS-0623845 to JS.

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 their 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 mastered number words to *three* generalized training on *four* to new objects and nouns, with approximate accuracy. In contrast, children who had mastered only *one* and *two* learned to apply *three* reliably within a single count noun context (*three dogs*) but did not apply it reliably to new objects labeled with different nouns (*three cows*). Both findings suggest that children fail to map newly learned words in their counting routine to fully abstract concepts of natural number.

Key words: counting, word learning, numerical concepts, cognitive development

The relationship between number words and concepts has drawn substantial attention from cognitive and developmental psychologists (Gelman & Gallistel, 1978, 2004; Wynn & Bloom, 1997; Dehaene, 1997; Mix, 2002; Carey, 2009) in light of two curious observations. First, children learn the verbal counting list before they understand that number words pick out specific, unique and exact cardinal values (Condry & Spelke, 2008; LeCorre & Carey, 2007). Second, children's learning of number-word meanings occurs slowly (Wynn, 1990, 1992a). When asked for a specific number of objects (the "Give-N" task), most 2-year-olds produce one object when asked for *one* but fail to produce a consistent amount when asked for larger numbers ("one-knowers"). By 2.5 years, most children give exactly 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 may hypothesize that each number word maps onto an abstract numerical magnitude (Gelman & Gallistel, 1978; Dehaene, 1997; 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 (for review see Feigenson, Dehaene & Spelke, 2004) and detect numerical correspondences between sets of objects and sounds (Izard, Sann, Spelke & Streri, 2009; Jordan & Brannon, 2006). Alternatively, children may initially apply number words more narrowly: a child who is taught the meaning of *two* in one context (e.g., *two dogs* applied to a set of two dogs) may extend the word only to objects (i.e., to pairs of horses but not of sounds or actions) or only to sets of entities named by the same count noun (i.e., to other dogs). This possibility is consistent with the finding that children fail explicit cross-modal numerical matching tasks until they master number word meanings (Mix, Huttenlocher & Levine, 1996).

The present research attempts 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), and they may be particularly useful for studying number word learning because they can reveal intermediary conceptual representations during the acquisition process (Seigler, 2007; Griffin & Case, 1996). In three experiments, we trained children who have mastered number words up to *two* or *three* on the next word in their count list, under systematically controlled conditions. We then asked what meanings children attributed to these words by assessing their generalization of the trained word to new entities and new language contexts.

Experiment 1

In Experiment 1, we explored the range of hypotheses that two-knowers and three-knowers assign to the meaning of the next word-to-quantity mapping (*three* or *four)*. We began by looking at a very limited form of generalization: Could children extend a word trained on one set of pictures of animals to a new set of pictures of different animals?

Methods

Participants. From a sample of 38 children, we selected the first 16 children categorized as two-knowers (2;6-3;6, M=3;2) or as three-knowers (3;2-3;9, M=3;7) on the Give-N task described below. All children were monolingual and were accompanied by a parent.

Counting: Children were given 10 objects and encouraged to count them. All children engaged in the counting routine and counted to *ten* without error.

Give-N Task: Children then 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 onto higher numbers in a pseudo-random order. When children failed to produce a quantity correctly, the experimenter asked for the preceding number before returning to the incorrect number. If children produced the correct quantity in one instance but an incorrect one in the other, they were asked for the number a third time to determine the maximum level of reliable knowledge. Children were assigned to a training condition based on their knower level.

Two-knower training: During training, two-knowers were shown cards depicting eight different kinds of animals. First, they saw two trials where a single card featuring three animals was labeled with a count phrase (*"This card has three cows!"*). Next, they were given six trials where a card with *three* was contrasted with a card depicting another quantity (*"This card has three birds! This card does not have three birds!"*). These contrasts included numbers that the children had mastered (3 vs. 1 or 2) and numbers they had yet to master (3 vs. 4, 5, or 10). Finally, children were presented with the same card pairs and asked to select the one with three items (*"Can you give me the card with three birds?"*). Errors were infrequent and were corrected.

During the test phase, two-knowers were shown ten new card pairs each featuring new kinds of animals and were asked to select one card using a number word. On two trials (known-known), the card pairs contrasted two known quantities (1 vs. 2: *"Can you give me the card with two horses?"*). On two further trials (trained-known), the trained number was paired with a known quantity (3 vs. 1 or 2), and on the six remaining trials (trained-unknown), it was paired with a quantity children did not have a word for (3 vs. 4, 5, 6, or 10). In both cases, children were asked for *three* (*"Can you give me the card with three pigs?"*). To discourage responding on the basis of non-numerical information, the paired test cards were matched for the total continuous extent of the objects (e.g., three large chickens vs. five smaller chickens) and the spatial arrangements of the items were varied (e.g., line of three vs. triangle).

Three-knowers: The procedure and controls were identical, except that children were trained and tested on *four*. The known numbers in training and test were 1-3 and the unknown numbers were 5, 6, 10, and 16.

Results and Discussion

Two-knowers. Two-knowers performed well on known-known trials (89%, $t(15)=7.01$, $p<.001, d=3.62$) and on trained-known trials (91%, $t(15)=8.06, p<.001, d=4.16$). In contrast, they performed at chance on the critical trials contrasting *three* to an unknown number (47%, $t(15)= 0.44, p>.60, d=.23$). A one-way ANOVA revealed a significant difference across the trial types,

$F(2,30)=17.02, p<.001, \eta^2=.53$. Performance in the known-known and trained-known conditions did not differ ($p>.80$) and was significantly better than in the trained-unknown condition (both p's<.01). In the trained-unknown condition, there were no reliable differences in performance across the different comparison quantities ($F(3,45)=1.20, p>.30, \eta^2=.07$).

The failure of two-knowers to generalize on the critical trials contrasts with their consistent selection of the correct cards during the final portion of the training phase. This discrepancy suggests that two-knowers employed one of two strategies. First, they may have learned to map each expression that included *three* onto the exact features of the corresponding training card, without extracting a more general relation between *three* and a numerical value. Alternatively, children's generalizations may have been restricted to particular count nouns (or object classes) which were evaluated as a single unit during the training phase (e.g. *"three dogs"* or *"three fish"*). Experiment 3 explores these possibilities.

Three-knowers. Three-knowers performed at ceiling on known-known trials (100%, $W=136, Z=3.49, p<.001$), and well above chance on trained-known trials (84%, $t(15)=7.86$, $p<.001, d=4.06$) and trained-unknown trials (71%, $t(15)=3.77, p<.01, d=1.95$). Nevertheless, there were reliable differences between the three trial types $F(2,30)=12.25, p<.001, \eta^2=.45$. Performance for known-known trials was higher than on the two conditions with the trained number (p's<.001), which did not differ ($p>.10$).

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 white bars in Figure 1). They reliably selected 4 when it was paired with 10 or 16 ($p<.01$) but not when it was paired with 5 or 6 ($p>.30$). Children's failure to select the card with four objects over that with 5 or 6 objects is striking, because infants can discriminate between arrays of 4 and 6 objects on the basis of number (Xu & Arriaga, 2007). It suggests that children mapped the newly trained word *four* onto a highly imprecise representation of number.

INSERT FIGURE 1 ABOUT HERE

Thus, from this brief training procedure three-knowers were able to extract an interpretation of *four* which generalized to sets of animals (and to count nouns) that had not been presented in training. Experiment 2 replicates this finding and investigates whether three-knowers will generalize *four* more broadly still, from pictured animals to solid artifacts.

Experiment 2

Three-knowers were familiarized with the meaning of *four* using the same card-pair training procedure, but they were tested with sets of concrete, household objects. The test objects therefore differed from the training stimuli both in their spatial and tactile properties and in their ontological status (animals vs. artifacts). If children's initial interpretation of *four* is sufficiently abstract, it should generalize across these features and their performance should be equivalent to Experiment 1. If, however, children's notion of the next word-to-quantity mapping is restricted to a more narrow conceptual domain (i.e., animals) or to more superficial properties of the exemplars (i.e., pictures), then generalization in Experiment 2 should be less robust.

Methods

Participants were 12 monolingual English-speaking children between 3;1 and 3;10 (mean 3;6) who counted at least to *ten* and who were categorized as three-knowers on the Give-N task. They were drawn from a sample of 29 children.

The training procedure and materials were identical to those used for the three-knowers in Experiment 1. Children were tested with seven sets of concrete objects pasted onto cardboard panels (coins, pencils). They received both trained-known and trained-unknown trials, following the procedure of Experiment 1 and using the same contrasting quantities (5, 6, 10, and 16). On the critical trained-unknown trials, the paired sets of objects were approximately matched in surface area (e.g., 4 large legos vs. 10 small legos).

Results and Discussion

After training, children successfully selected *four* objects on trained-known trials (86%, $t(11)=7.39$, $p<.001$, $d=4.46$) and trained-unknown trials (70%, $t(11)=2.69$, $p<.05$, $d=1.62$). Three-knowers performed no worse in Experiment 2 than in Experiment 1. An ANOVA with Experiment and Trial type as factors revealed better performance when *four* was paired with a known number than an unknown number ($F(1,26)=6.47$, $p<.05$, $\eta^2=.20$) but no main effect of Experiment or interaction across the experiments, p's>.60. Thus three-knowers acquired a word-to-quantity mapping for *four* that generalized from pictures to concrete objects.

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

In summary, three-knowers successfully generalized *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 is fairly broad. Children's greater success on test pairs with larger numerical differences suggests that they mapped the trained word onto an approximate numerical representation like that found in animals, infants and adults in diverse cultures (Feigenson et al., 2004). We consider this finding in the General Discussion.

But first, we return 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 about *three* was restricted to particular object kinds, 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 just this pattern of conservative generalization. When 3-year-olds hear a bumpy horse described with a novel adjective ("*a very blickish horse*"), for example, they successfully generalize *blickish* to other bumpy horses but not to animals from different basic-level categories such as a bumpy rhinoceros (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 either the same or a different category.

Methods

Participants were 16 monolingual English-speaking children between the ages of 2;3 and 3;5 (mean 3;1) who counted at least to *ten* and were categorized as two-knowers by the Give-N task. Children were drawn from the same sample as in Experiment 2.

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 larger sets of dogs (3 vs. 4, 5, or 10), and both were labeled with respect to the trained number ("*This card has/does not have three dogs!*"). Following this demonstration, children again were presented with the same card pairs and asked to select the one with the trained number. Errors were infrequent and were corrected.

During test, children were presented with 12 card pairs in which the target was presented with a higher, unknown number (4, 5, or 10). There were four trial types with different kinds of target card: (1) the original target (small dogs in a triangle), (2) the target set transformed in size and spatial configuration (large dogs in a row), (3) different target objects from the same basic-level kind (dogs of a different breed in a triangle), and (4) 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, 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 each block 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 Figure 2). While children performed equally on the original target, size/space variant, and within-category variant trials (all p's>.40), they performed significantly less well on the between-category variant trials (all p's<.01). Children reliably identified *three* when presented with the original target cards ($t(15)=4.88, p<.001, d=2.52$), new cards in which the targets changed in size and spatial arrangement ($t(15)=6.78, p<.001, d=3.50$), and new cards depicting objects from a different subordinate class within the same basic-level kind ($t(15)=6.00, p<.001, d=3.10$). However, children were at chance when the cards depicted animals from a different basic-level category ($t(15)=0.68, p>.50, d=.35$). These findings provide evidence that two-knowers' initial interpretation of *three* is limited with respect to the particular category and/or noun that is quantified.

INSERT FIGURE 2 ABOUT HERE

Collapsing across the four trial types, we again found no overall effect of numerical distance on correct card selection, $F(2,30)=0.39, p>.60, \eta^2=.02$ (see Figure 3). 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.

INSERT FIGURE 3 ABOUT HERE

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 (Experiment 1) and even to solid artifacts (Experiment 2). These children, however, generalized *four* in an approximate manner in both experiments, applying the word to sets of 5 or 6 objects despite contrastive training with these numbers. These findings suggest that three-knowers relate the newly-trained number word to a fairly general, but approximate, representation of numerical magnitude.

Second, children who had mastered the meanings only of *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 for two reasons. First, two-knowers have learned to map *one* and *two* to a diverse range of entities including visible objects and audible events (Wynn, 1990). Second, these children counted reliably to *ten*, producing words like *three* in the very same contexts in which they produced *one* and *two*. Nevertheless, their narrow generalization of *three* suggests that their understanding differs qualitatively from that of three-knowers or adults.

Children's pattern of generalization for *three* lends itself to two complementary accounts. First, two-knowers may initially map the entire quantified phrase ("three dogs") to an abstract adult-like representation of its meaning (e.g., for some set x, x consists of dogs and x has cardinality 3) and are able to use this mapping to extend the phrase "three dogs" to new sets of three dogs. But breaking the quantifier loose from this semantic structure might require more nouns, more example sets of objects, or more analysis than the present experiments allowed. 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; Grimshaw, 1990; Waxman & Booth, 2001).

Second, two-knowers may parse the number word out of 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). For the purpose of quantifying sets, this "object file" system is constrained in two ways. First, it has a capacity limit of 3 objects in children (Wynn, 1992b; Feigenson, Carey, & Hauser, 2002; Wood & Spelke, 2005) and consequently could provide possible meanings for *three* but not *four* during number word acquisition. Second, it expresses quantities only implicitly in terms of the presence of individuals and their properties: an array of two dogs would be expressed as DOG and DOG by this system. Thus two-knowers who mapped *three* to the representation DOG and DOG and DOG might reasonably infer that the term only applies 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 which represents larger approximate magnitudes. This system supports infant computations of large quantities across various sensory domains (Brannon, 2002; Xu & Spelke, 2000; Lipton & Spelke, 2002; Wood & Spelke, 2005), and it guides children's understanding of number words before they learn verbal counting (Wagner & Johnson, 2009; Shusterman, Carey, & Spelke, 2009). While the object-file system has a set size limit of three, the approximate number system could provide candidate meanings for larger number words. The limits on object file representations might force children to shift from one representational system to another between *three* and *four*, potentially explaining why three-knowers entertained a broader hypothesis about the scope of the next word-to-quantity mapping. However, unlike the integers, this system does not provide exact representations of numerosity (Dahaene, 1997) and thus may encourage children to generalize the newly-learned word *four* to nearby magnitudes, even if the child can discriminate them (4 vs. 6).

This interpretation of the present findings may lend insight into the slow pace of children's number word acquisition. In the absence of a single system with the full capacity to express exact numerical representation, children cannot simply map number words onto existing concepts. Instead they have to create a conceptual representation that goes beyond either of the two supporting systems (Carey, 2009). Further studies using the present training methods may help to specify how children construct this new conceptual representation.

References

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 Science*, 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, 83-102.

Grimshaw, J. B (1990). *Argument structure*. Cambridge, MA: MIT Press.

Jordan, K. E. & Brannon, E. M. (2006). The multisensory representation of number in infancy *Proceedings of the National Academy of Sciences*, 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., Huttenlocher, J., & Levine, S. C. (1996). Do preschool children recognize auditory-visual numerical correspondences? *Child Development*, 67, 1592-1608.

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

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

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

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

Siegler, R.S. (2007). Cognitive Variability. *Developmental Science*, 10(1), pp. 104-109.

Wagner, J. B. & Johnson, S. C. (2009). 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). Children's acquisition of number words and the counting system. *Cognitive Psychology,* 24, 220–251.

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

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., & 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.

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

Figure 1: In Experiments 1 and 2, the proportion of correct card choices across four ratios by three-knowers.

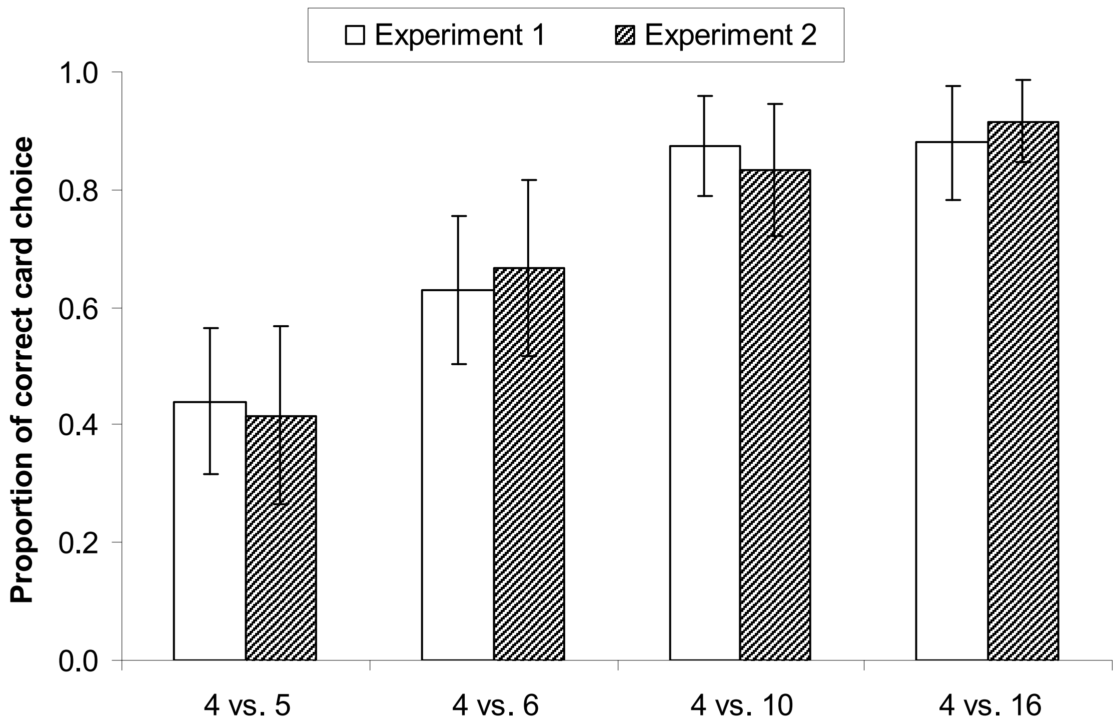

Figure 2: In Experiment 3, the proportion of correct card choices for four trial types by two-knowers.

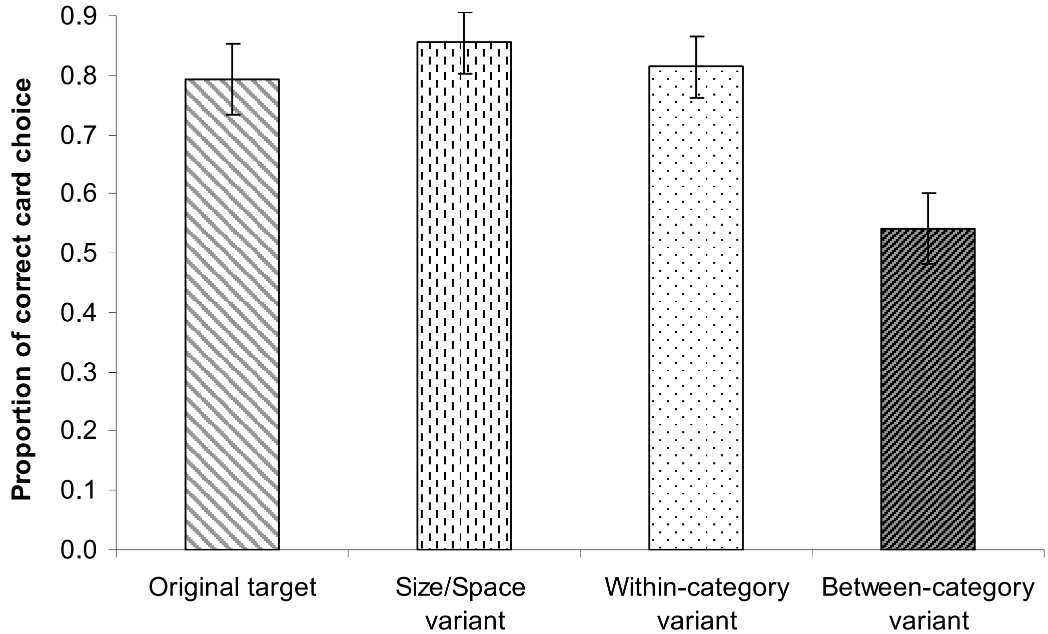

Figure 3: In Experiment 3, the proportion of correct card choices across three ratios by two-knowers.