Children learn non-exact number word meanings first

Junyi Chu
junyichu@ucsd.edu
Department of Psychology
UC, San Diego

Katie Wagner
kgwagner@ucsd.edu
Department of Psychology
UC, San Diego

David Barner
barner@ucsd.edu
Department of Psychology
UC, San Diego

Abstract

Children acquire exact meanings for number words in distinct stages. First, they learn *one,* then *two,* and then *three* and sometimes *four.* Finally, children learn to apply the counting procedure to their entire count list. Although these stages are ubiquitous and well documented, the foundation of these meanings remains highly contested. Here we ask whether children assign preliminary meanings to number words before learning their exact meanings by examining their responses on the Give-a-Number task to numbers for which they do not yet have exact meanings. While several research groups have approached this question before, we argue that because these data do not usually conform to a normal distribution, typical methods of analysis likely underestimate their knowledge. Using non-parametric analyses, we show that children acquire non-exact meanings for small number words like *one, two, three, four* and possibly for higher numbers well before they acquire the exact meanings.

Keywords: Number word learning; Counting; Cognitive Development; Language acquisition

Introduction

How do children acquire the meanings of large exact numbers like "47"? Although humans are able to represent quantity in absence of language, attested systems of number representation are either limited to representing small arrays of objects precisely, or to representing larger quantities approximately (Feigenson, Dehaene, & Spelke, 2004). No attested non-verbal system is able to represent large numbers precisely (Laurence & Margolis, 2005). Here, we investigated how children acquire labels for exact numbers by exploring the very first meanings they exhibit. To do this, we analyzed previously collected data in a simple but novel fashion and found that, before children learn exact meanings for words like *one, two, three, four,* and perhaps larger numbers, they first learn non-exact meanings.

In the US, children learning English generally begin number word learning around age 2, by learning to blindly recite a small subset of number words in sequence (*one, two, three,* etc.), much like they learn the alphabet, though in tandem with procedures that involve pointing at objects in sequence (Gelman & Gallistel, 1986; Fuson, 1988; Wynn, 1990, 1992). Not long after, they learn an exact meaning for the word *one*, and at this time can give one object when asked for *one*, but cannot reliably give larger numbers when asked. These children are often referred to as "1-knowers". After a long delay – around 6 to 9 months later – children learn an exact meaning for *two,* and are called "2-knowers". Next, they learn *three* ("3-knowers"), and sometimes *four* ("4-knowers"), in sequence. After this, they appear to discover something more general about numbers: That the counting procedure which they began with can be used to label and generate larger sets. Having noticed this, children become able to give any set within their count list and map number words to large precise cardinalities (Wynn, 1990, 1992; see also Schaeffer, Eggleston, & Scott, 1974). These children are often referred to as Cardinal Principle knowers or "CP-knowers" (for discussion see Davidson, Eng, & Barner, 2012; Lipton & Spelke, 2005; Wagner, Kimura, Cheung, & Barner, 2015).

Several recent studies have explored the role of different perceptual systems in this learning trajectory. According to one class of accounts (e.g. Carey, 2009; Klahr & Wallace, 1976; Le Corre & Carey, 2007), the fact that children are limited to learning labels up to *three* or *four* in absence of counting suggests that these meanings are learned from a capacity limited system of object representations. Consistent with this, studies using several distinct experimental paradigms have found that preverbal infants can keep track of a maximum of 3 objects in parallel, a capacity sometimes dubbed "parallel individuation" (Feigenson & Carey, 2003, 2005). Other researchers, however, have argued that number words might be constructed instead from representations in the approximate number system, or ANS (e.g. Cantlon, 2012; Spelke & Tsivkin, 2001; Wagner & Johnson, 2011). The ANS allows humans to represent sets approximately and compare them according to their ratio, such that 20 vs. 40 and 200 vs. 400 are equally discriminable, whereas 200 vs. 220 is not. These representations appear to be available from birth (e.g. Brannon, 2002; Xu & Spelke, 2000), and become sensitive to increasingly finer numerical differences well into elementary school (Halberda & Feigenson, 2008).

Currently, the relative roles of these two candidate systems in number word learning is uncertain. First, it is unclear in principle how either system *could* supply the meanings of number words. As argued by Laurence and Margolis (2005), parallel individuation is a system for representing the properties and locations of individual objects, and not the properties of sets – like cardinality. Further, parallel individuation, though at best useful for learning small numbers, could not define meanings for numbers beyond three or four. Meanwhile, the ANS, though not limited to representing small numbers, is non-exact, and cannot be used to reliably distinguish large quantities like 817 and 820. Perhaps more important, neither system supplies the type of logical representations that characterize counting and the foundations of arithmetic, such as exact equality and the successor function (i.e., that the successor of any natural number n is equal to $n+1$).

In recent years, many researchers have examined children's number word knowledge to determine if children have preliminary meanings prior to exact ones, and if so, whether these meanings show signatures of parallel individuation and/or the ANS. If preliminary meanings are rooted in parallel individuation, such meanings would be limited to *four* since higher numbers are beyond the capacity limit of this system. In contrast, if children have preliminary meanings rooted in the ANS, these meanings could extend beyond *four* and should show a distinct characteristic of the ANS. Specifically, the standard deviation of children's errors should increase as the target numerosities increase (e.g., Le Corre & Carey, 2007; Wagner & Johnson, 2011).

Although several research groups have pursued these questions, differences in procedures and analysis have led to oftentimes conflicting conclusions. Early studies examined whether children who can count have mappings between the ANS and large number words (up to 100) using dot-array estimation tasks. These studies found that many CP-knowers (Le Corre & Carey, 2007) and less-skilled counters (Lipton & Spelke, 2005) lack such mappings.

More recent studies have probed the role of the ANS by asking whether, before learning to count, children already show evidence of mappings between number words and approximate magnitudes for at least a limited set of numbers beyond their knower level. For instance, Sarnecka and Lee (2009) argued that if children map number words to the ANS, then errors on the Give-a-Number (Give-N) should be normally distributed around the requested numerosities, with higher means for larger requests. Contrary to this prediction, they found that the means of incorrect responses beyond knower level were not higher for larger requests. However, as we show in the present study, excluding correct responses and analyzing only errors introduced systematic bias that underestimates knowledge. Assuming that children's responses are random, we expect the mean of children's responses (both correct and incorrect) to be the same across all requested numbers, as assumed by Sarnecka and Lee in their analysis. Critically, however, removing correct responses alters this expectation, such that a negative slope is expected. This is because the size of responses removed from analyses necessarily increases with the size of the requested number (e.g., correct responses to requests for *ten* are larger than for *one*). This makes it challenging to draw conclusions from incorrect responses only.

A more common approach is to examine the means of all responses beyond children's knower levels, both correct and incorrect (e.g., Barner & Bachrach, 2010; Gunderson, Spaepen, & Levine, 2015; Wagner & Johnson, 2011). Using data from Give-N and a set labeling task called "What's on this Card?", studies by Barner and Bachrach (2010) and Gunderson et al. (2015) found that children have non-exact meanings for the number just beyond their knower level (KL+1). However, responses for larger numbers (KL+2 and above) produced a flat slope, suggesting undifferentiated meanings (see Gunderson et al., 2015, for evidence that Wagner & Johnson's data support the same conclusion).

Whereas some studies (Barner & Bachrach, 2010; Le Corre & Carey, 2007) find that different tasks generate identical patterns of data, others find differences across tasks. For example, Gunderson et al. only find evidence of non-exact meanings using the What's on this Card task, which requires labeling a presented set. Shaeffer, Eggleston, and Scott (1974) find the opposite, with better performance generating sets than labeling them, a result replicated by Odic, Le Corre, and Halberda (2015). However, in the cases of Shaeffer et al., Odic et al., and Wagner and Johnson, comparisons across studies are nearly impossible, since no common measure was used across studies (e.g., Wynn's Give-N criteria for knower level assignments). In the one case where knower levels were assessed (Odic et al.), this was done with a counting What's on this Card task, which might have underestimated or overestimated children's knower levels relative to the standard Give-N task.

A further barrier to understanding these past studies is that, while most analyze all data, including correct responses, they focus on children's mean response as the dependent variable. This is problematic because although means provide an excellent characterization of central tendency for normally distributed data, children's responses on the Give-N task violate normality assumptions in profound ways, as shown in Figure 1, which depicts the responses of 2-knowers to requests for *four* and *five.*

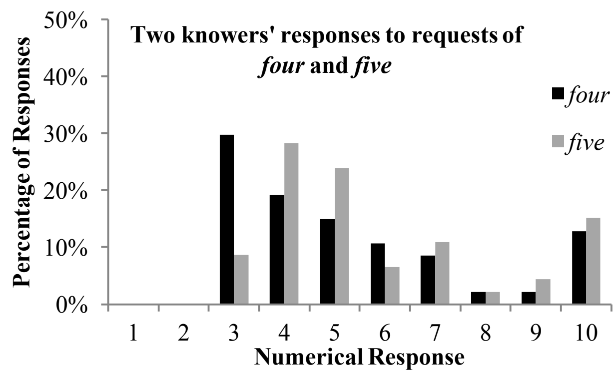

Figure 1: 2-knowers' responses for requests for *four* and *five.*

Give-N data predictably deviate from the normal distribution for two reasons. First, most obviously, normality is violated by the fact that children frequently give all available items (e.g., 10 out of 10) when they are uncertain of how to respond (see Figure 1). Second, children's responses are both upper- and lower-bounded. They are upper-bounded by the number of objects the experimenter chooses to present to the child: If the child is offered 10 objects, then only 10 can be given, even if the child might normally wish to give 12 or 15. And they are lower bounded by the child's knower level: If a child is a 2-knower then responses to *three* necessarily have a lower bound of three items. This is because the criteria for 2-knower classification require that, 2/3 of the time, the child gives two objects when asked for *two,* and further, provides two objects only when asked for *two* but not when asked for other numbers. Thus, the lower bound is given by how we choose to classify children's knowledge of numbers. Though some studies did not attempt a knower level assignment (e.g., Schaeffer et al., 1974; Wagner & Johnson, 2011) children's responses are still expected to be lower bounded because there is ample evidence that children do acquire exact meanings for number words in stages (e.g., Wynn, 1990; 1992), thereby constraining responses to requests for higher numbers.

These considerations lead us to conclude that all past studies may underestimate children's non-exact knowledge of numbers beyond their knower level, and that differences in data analysis may explain conflicting results and asymmetries between different tasks (which vary according to whether they are subject to problems of upper and lower bounds). To explore this conjecture, we analyzed data from a non-titrated variation of the Give-N task and replicate the analyses of both Sarnecka and Lee (2009) and Gunderson et al. (2015). In this variation, each child received three trials for each number tested in a pseudo-randomized order, regardless of their performance. This is important because it allowed us to examine children's responses for inquiries well beyond their knower level. We compared these analyses to non-parametric analyses, and show that children acquire non-exact meanings for small number words like *one, two, three, four* and possibly higher numbers well before they acquire their exact meanings.

Methods

Participants

Our data set combined data from a study of English monolinguals (Almoammer et al., 2013) and a study of bilinguals (Wagner et al, 2015). Only subset knowers were included from each study. Monolinguals were 44 children aged 2;0 to 4;3 years (M = 3;1). There were 29 bilinguals including 14 who spoke both English and French (M=3;4, range = 2;4 – 4;0) and 15 who spoke both English and Spanish (M=3;8, range = 2;10 – 4;11)[1]. Only English dominant bilinguals (i.e. those with higher knower-level when tested in English) were included, to ensure that stronger number knowledge in a second knowledge did not affect performance in English. Analyses below confirm no difference between monolingual and bilingual children.

Procedures

Give-N Task. This task was adapted from Wynn (1990) and was used to assess children's knowledge of number word meanings in each language the child spoke. The experimenter began by presenting the child with a plate and ten similar objects. For each trial, the experimenter asked the child to place a quantity on the plate, avoiding singular and plural marking by asking, "Can you put *n* on the plate? Put *n* on the plate and tell me when you're all done." Once the child responded, the experimenter asked, "Is that *n*? Can you count and make sure?" and encouraged the child to count. If the child recognized an error, the experimenter allowed the child to change his/her response.

Children completed up to twenty-one quasi-randomized trials, consisting of three trials for each of the seven numbers tested (i.e., 1, 2, 3, 4, 5, 8, and 10). Children were defined as an *n*-knower (e.g., 3-knower) if they correctly provided *n* (e.g., 3 fish) on at least two out of the three trials that *n* was requested and, of those times that the child provided *n*, two-thirds of the times the child did so it was in response to a request for *n*. If *n* was five or higher, the child was classified as a CP-knower.

Results

Knower Level Classification

In our dataset, there were 28 non-knowers, 15 1-knowers, 16 2-knowers, 11 3-knowers, and 3 4-knowers. Because the number of 4-knowers was small (n = 3) and because prior evidence suggests CP-knowers are likely to be misclassified as 4-knowers (Wagner et. al, 2015), we took a conservative approach and excluded 4-knowers.

Further division of non-knowers. While it is likely that a large subset of non-knowers know nothing about the meanings of number words, it is also possible that some have begun developing meanings for small numbers like *one* and *two.* To assess this, we divided non-knowers into two groups. The first group (N=16) was comprised of non-knowers who gave the same number of objects in response to every request, while children in the second group (N=12) provided at least two unique responses. We excluded the first group from our analyses; for the remainder of the paper, the term *non-knower* refers only to this second group.

Replication of Sarnecka & Lee (2009)

Replicating Sarnecka and Lee's (2009) results, we found that the means of children's incorrect responses bore no positive relationship to the target value (see Table 1 and Figure 2a). The mean incorrect response was significantly different from the correct response for all but one target number, *five,* as the means for all requested numbers were close to 5. This was true for all knower levels.

Replication of Gunderson et al. (2015)

Rather than rely on incorrect responses only, Gunderson et al. (2015) analyzed the means of both correct and incorrect responses, and asked whether the slope of the means for larger numbers was positive (see Table 2). We applied this method to our data set (see Figure 2b). By comparing Figure 2a and Figure 2b, we see that a positive slope emerges once correct responses are included. In Gunderson et al. (2015)'s first experiment, they found no statistically significant positive slope from 5 to 9 amongst subset knowers but they did find a statistically significant positive slope from KL+1 to 9, though this effect was driven primarily by 3-knowers. Like Gunderson et al. we found no positive slope for subset knowers from 5 to 10 (0.11, p = 0.12, d = 0.25). However,

[1] Preliminary analyses comparing bilingual and monolingual data revealed no visual differences in the distribution of responses. For brevity, we combined both samples in the current report.

the slope from KL+1 to 10 reached statistical significance (0.13, p = 0.0031, d = 0.48). We also included an additional analysis of slope, looking at KL+2 to 10 and found that this slope also reached statistical significance for subset knowers (0.11, p = 0.037, d = 0.33). Non-knowers did not show a positive slope for KL+1 to 10, KL+2 to 10, or 5 to 10.

Table 1: Means of Incorrect Responses

N	# Trials	# Incorrect responses	Mean Incorrect	Comparison to Target		
				df	*t*	*p*
1	166	17	6.44	13	6.3	<0.01
2	167	34	6.73	18	6.5	<0.01
3	164	86	5.10	39	4.3	<0.01
4	167	129	5.20	49	3.5	0.01
5	165	139	5.50	52	0.49	0.63
8	164	153	4.93	53	8.7	<0.01
10	165	126	4.34	46	17	<0.01

Analyses of Median Responses

Having shown that we replicate previous findings in the literature using previously deployed analyses, we next sought to ask whether these studies underestimated knowledge. Already we have shown that removing correct responses from an analysis of mean responses systematically underestimates knowledge of numbers from *one* to *three.* Here we asked whether a focus on means also results in underestimation of knowledge.

To assess this, our first step was to summarize data by median response (including both incorrect and correct responses) as in Figure 2c. In place of an analysis of slope, we conducted sign tests comparing children's median responses on KL+1 to 10, KL+2 to 10 and 5 to 10 (see Table 3). None of these comparisons were statistically significant for non-knowers. However, for subset knowers, sign tests comparing both KL+1 to 10 and the KL+2 to 10 reached statistical significance, both ps < 0.01. Although subset knowers also showed a positive slope from 5 to 10, these differences were not significant, p = 0.11.

A second approach to addressing this question compared children's responses for requests beyond their knower-level to chance. To do this, we created a null distribution of responses which accounted for the fact that children give certain quantities more frequently than others. For example, 2-knowers gave ten objects on 0.17 of trials for requests above two. Therefore, if 2-knowers responses are independent of the requested number, we should expect that on 17% of requests for *ten,* 2-knowers will give 10 objects. We used binomial tests to compare children's actual responses (e.g., 2-knowers provided ten objects on 0.31 of *ten* trials) to chance (e.g., 0.17) to ask whether they were more likely to provide correct responses to requests for numbers from KL+1 to 4, and separately from 5 to 10. We did not include analyses for KL+2 to 4 since such an analysis would be limited to responses of *three* and *four* by 1- knowers and responses to *four* by 2-knowers.

The responses of non-knowers did not differ from chance for requests between KL+1 and 4 or between 5 to 10. In contrast, subset knowers were more likely than chance to respond with a correct amount for both requests from KL+1 to 4 from 5 to 10 (see Table 4). To eliminate the possibility that correct responses for 5 to 10 were driven by children giving all items (which happened to be 10), we also separately analyzed requests for 5 and 8 only and found that subset knowers were still correct more frequently than chance (0.12; chance = 0.085, p = 0.04).

Discussion

Well before children acquire the exact semantics of number words, they first acquire non-exact meanings both for numbers below five, and larger numbers too. To show this, we analyzed data from the Give-N task, and conducted analyses that included both correct and incorrect responses, that focused on median responses rather than means, and that defined chance responding in a way that respected the constraints of the Give-N tasks (adjusting for children's base rates of responding). Specifically, we replicated the findings of past studies using their analyses, and then showed that these studies systematically underestimate knowledge. Consequently, we provide the first evidence that standard tasks like Give-N, which have been used for over 40 years, can detect children's non-exact meanings for number words well beyond their knower level.

Our findings are compatible with the idea that children's preliminary number meanings are rooted in the approximate number system. However, evidence from other studies suggests that even much older children – e.g., 7 year olds – lack strong associative mappings between the ANS and number words beyond 5 (Sullivan & Barner, 2014). To the extent that these older children are able to make accurate estimates, these are driven by a global structure mapping wherein the relation between number words (e.g., *ten*) and particular magnitudes (e.g., 10) are highly malleable across contexts, unlike the meanings of the positive integers, which remain fixed independent of context.

An alternative explanation consistent with both the current data and the notion that estimates above 5 are driven more by structure mappings than mappings between number words and the ANS is that children's responses are guided by principled knowledge that numbers later in the count list denote greater quantities (the "later-greater" principle). Contrary to this hypothesis, however, several previous studies have found that no knowledge of this principle in subset knowers and even in many CP-knowers (Le Corre, 2014; Condry & Spelke, 2008; Davidson et al., 2012).

These studies in older children point to the possibility that the knowledge driving children's inexact meanings (whether ANS mappings or the later-greater principle) may not extend to their entire count list. Consistent with this idea is data showing that other types of numerical knowledge, specifically the knowledge of numerical successors, is first acquired incrementally long before children abstract the

Table 2: Slopes of the means of all responses from KL+1 to 10 and 5 to 10

		KL+1 to 10				KL+2 to 10				5 to 10			
	N	mean	SD	p	d	mean	SD	p	d	mean	SD	p	d
Non-knowers	9	-0.006	0.39	0.84	-0.016	-0.067	0.35	0.58	-0.19	-0.120	0.66	0.60	-0.18
All Subset Knowers	**42**	**0.128**	**0.26**	**0.0031*****	**0.48**	**0.107**	**0.32**	**0.037***	**0.33**	**0.108**	**0.44**	**0.12**	**0.25**
1-knowers	15	0.030	0.21	0.59	0.14	0.030	0.26	0.66	0.11	0.070	0.46	0.57	0.15
2-knowers	16	0.192	0.21	0.0022***	0.92	0.172	0.25	0.015**	0.67	0.138	0.42	0.21	0.33
3-knowers	11	0.167	0.37	0.161	0.46	0.117	0.47	0.43	0.25	0.117	0.47	0.43	0.25

Table 3: Slope and Sign Tests of Median Responses

		KL+1 to 10		**KL+2 to 10**		**5 to 10**	
	N	Slope	Sign test p	Slope	Sign test p	Slope	Sign test p
Non-knowers	9	0.056	0.34	-0.125	0.96	0	0.89
All Subset Knowers	**42**		**<0.01**		**<0.01**		**0.11**
1-knowers	15	0.25	0.36	0.14	0.05	0.2	0.36
2-knowers	16	0.43	<0.01	0.33	<0.01	0.4	0.27
3-knowers	11	0.17	0.34	0.2	0.34	0.2	0.34

Table 4: Correct Responses for KL+1 to 10 compared to chance

	KL+1 to 4				**5, 8, 10**			
	# Trials	Proportion Correct	Chance	p	# Trials	Proportion Correct	Chance	p
Non-knowers	88	0.25	0.19	0.12	65	0.035	0.069	0.91
All Subset Knowers	**281**	**0.31****	**0.24**	**<0.01**	**402**	**0.16****	**0.11**	**<0.01**
1-knowers	130	0.31**	0.22	<0.01	129	0.10	0.084	0.28
2-knowers	118	0.26	0.22	0.23	175	0.19**	0.12	<0.01
3-knowers	33	0.45	0.36	0.16	98	0.21*	0.14	0.04

successor principle and can apply it to their entire count list (Wagner et al, 2015; Davidson et al, 2012).

Determining the source of children's inexact meanings for number words will inform what information children have at their disposal when learning the counting principles. This knowledge could be a useful stepping stone to acquiring the counting principles, but on it's own, it cannot be sufficient, given that children are forming these inexact meanings months before learning the counting principles.

Acknowledgments

Teachers, children and families at VIP Village Preschool in Imperial Beach, CA. NSF (#1420249) awarded to DB.

References

Almoammer, A., Sullivan, J., Donlan, C., Marusic, F., Zaucer, R., 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. *Cognitive Psychology, 60,* 40-62.

Brannon, EM.(2002). The development of ordinal numerical knowledge in infancy. *Cognition, 83*(3), 223-240.

Cantlon, J. F. (2012). Math, monkeys, and the developing brain. *PNAS, 109,* 10725-10732.

Carey, S. (2009). Where our number concepts come from. *Journal of Philosophy, 106* (4), 220-254.

Condry, K.F. & Spelke, E.S. (2008). The development of language and abstract concepts: The case of abstract number. *J Exp Psychol Gen, 137, 22-38*

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

Feigenson, L. & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6,* 568-584.

Feigenson, L. & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*, 295-313.

Feigenson, L., Dehaene, S. & Spelke, E. (2004) Core systems of number. *Trends Cogn Sci, 8*(7), 307-314.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Harvard University Press.

Gunderson, E., Spaepen, E. & Levine, S. (2015). Approximate number word knowledge before the cardinal principle. *J Exp Child Psychol, 130,* 35-55.

Halberda, J. & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Developmental Psychology, 44(5)*, 1457-1465.

Klahr, D. & Wallace, J. G. (1976). *Cognitive development: An information processing view.* Hillsdale, NJ: Erlbaum.

Laurence, S. & Margolis, E. (2005). Number and natural language. In P. Carruthers, S. Laurence, & S. Stich (eds), *The Innate Mind: Structure and Contents* (pp. 216-235). Oxford University Press.

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*, 395-438.

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *Brit J Dev Psychol, 32(2), 163-177.*

Lipton, J. S. & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*(5), 978-988.

Odic, D., Le Corre, M. & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition, 138,* 102-121.

Sarnecka, B. & Lee, M. (2009). Levels of number knowledge during early childhood. *J Exp Child Psychol, 103,* 325-337.

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

Spelke, E. S., & Tsivkin, S. (2001). Initial knowledge and conceptual change: Space and number. *Language Acquisition and Conceptual Development*, 70-97.

Sullivan, J. & Barner, D. (2014). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology,* 171-189.

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10-22.

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

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

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

Xu, F. & Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1). B1-B11.

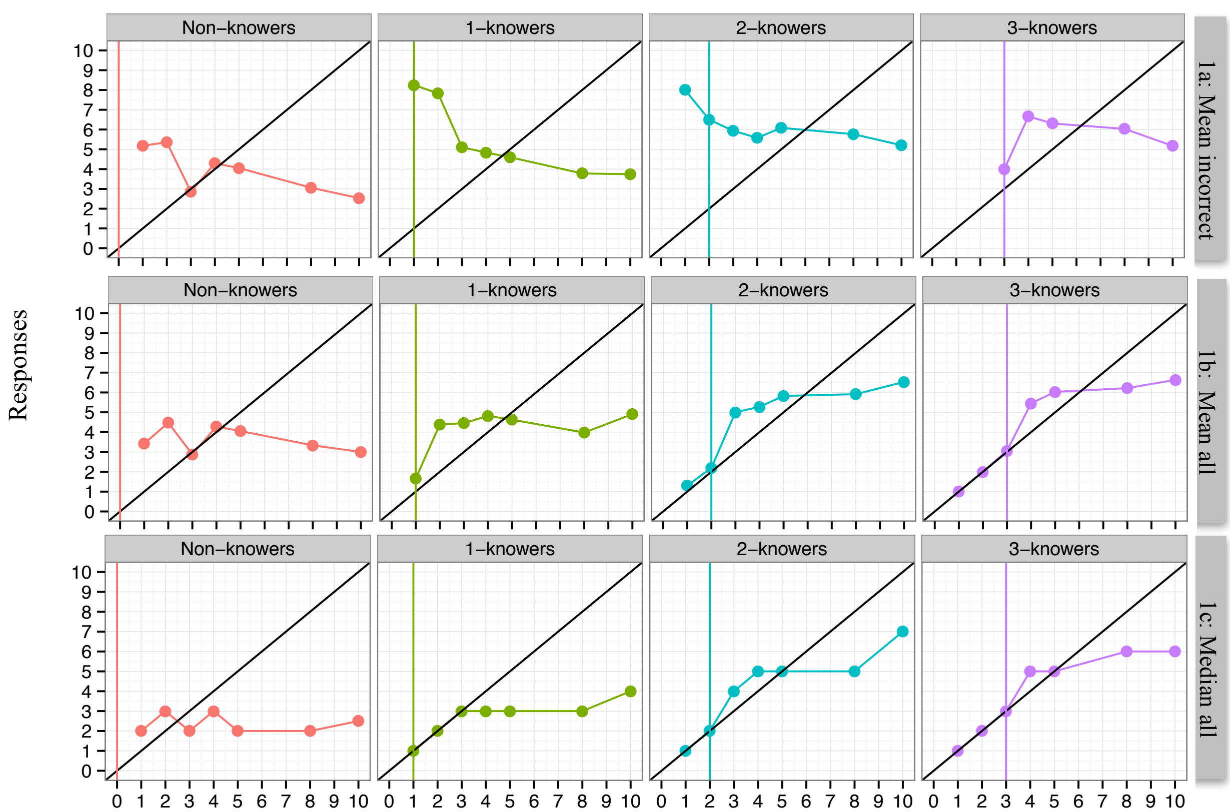

Figure 2: Responses for each knower level, using different summary statistics.
(a) Mean incorrect responses, (b) Mean responses, and (c) Median responses.