How children interface number words with perceptual magnitudes

Denitza Dramkin (ddramkin@psych.ubc.ca)
Department of Psychology, University of British Columbia
Vancouver, Canada

Darko Odic (darko.odic@psych.ubc.ca)
Department of Psychology, University of British Columbia
Vancouver, Canada

Abstract

How do children map symbolic number words to continuous and noisy perceptual magnitudes? We explore how 5- to 12-year-olds attach novel units to number, length, and area by examining whether verbal estimation performance is primarily predicted by access to number words, the precision of children's underlying perceptual systems, or a more general process in structurally aligning number words with perceptual magnitudes. We find that from age five onward, children can readily form novel mappings between number words and perceptual magnitudes, including dimensions they have no experience estimating in (e.g., length, area), and even when faced with completely novel units (e.g., mapping a collection of *three* dots to "one" unit for number). Additionally, estimation performance was poorly predicted by the noise in their underlying perceptual magnitudes and number word access. Instead, we show that individual differences in children's abilities to translate continuous perceptual signals into discrete categories underlie verbal estimation performance.

Keywords: Number words, estimation, perceptual magnitudes

Introduction

Language and perception are distinguishable in their representational formats: while words carry discrete and symbolic meanings, much of perception is richly continuous and imprecise. Despite this, we frequently – and effortlessly – access and reason across these two systems, attaching known words to our perceptual experiences. What psychological mechanisms support this interface between discrete words and analog perceptual experiences?

An important case-study in our understanding of how we connect language and perception is the link between number words (e.g., "fifteen") and our intuitive, perceptual sense of number (i.e., the Approximate Number System; ANS). While adults can perceive a collection of objects on the screen and easily attach a number word to them (e.g., responding "eight" when briefly shown some dots), the development of this interface is far from trivial. Even after young children have learned the meanings of number words, they do not immediately interface them with their ANS (Le Corre & Carey, 2007; Odic et al., 2015). Their mappings – at least initially – appear to be *associative* in nature and require a lot of practice: children must first learn to map small collections of objects with "one," then "two," then "three," etc. (Le Corre & Carey, 2007). In turn, children take a long time to develop *structural* mappings between number words and the ANS (e.g., reasoning about how to map number words to the ANS through analogy or their shared ordinal structure; Gentner, 1983; Izard & Dehaene, 2008; Sullivan & Barner, 2013, 2014). Achieving the latter is pivotal as it allows children to eventually attach number words to quantities that they have never encountered or practiced (e.g., being able to spontaneously guess that there are *three-hundred-and-forty-seven* candies in a jar). Here, we focus on understanding what factors predict performance in verbal number estimation.

Broadly speaking, children's verbal number estimation performance could be predicted by at least three (non-mutually exclusive) factors. First, since children's number estimates are at least partly the consequence of the imprecision of their ANS (Izard & Dehaene, 2008), children might become more accurate and less variable in attaching number words to their number sense primarily due to the developmental maturity of their perceptual representations. Second, children may become better number estimators as their knowledge and rapid access to a broader set of number words becomes deeper and more reliable (e.g., children who know numbers words above "ten" may show better estimation performance when estimating more than 10 dots, compared to children who have not get acquired these number words). Finally, with practice and maturity, children may become better in the actual process of mapping number words to perceptual states through structure mapping, allowing their estimation performance to improve, even if their knowledge of number words and access to precise perceptual states stays relatively constant.

The challenge in disentangling the contributions of these factors is that in typical number estimation tasks – whereby children are asked to *verbally* report "how many" items they saw – is that the precision of the ANS, *and* accessibility of number words, *and* practice with this interface all improve with age (Odic et al., 2015; Libertus et al., 2016). To tackle this, we take a different approach: comparing how well children attach number words to their perceptual representations of not only number, but also length and area, which improve along distinct developmental trajectories (Odic, 2018).

Much like the ANS, adults have access to an automatic and intuitive perceptual representations for length and area, and can map number words to these dimensions (e.g., estimating that a line is 25cm long or that a room is 520ft^2). However, our perception of length and area is significantly more precise and develops quicker than the ANS: while a typical 5-yearold can reliably discriminate a small ratio of two line lengths (e.g., 12cm vs. 11cm), they struggle to discriminate that same ratio in number perception (e.g., 12 dots vs. 11 dots) (Odic, 2018; Starr & Brannon, 2015). Therefore, if *perceptual acuity* is the primary driver of estimation performance, we should find that once children can map number words to length and area, their estimation performance in these dimensions should be more precise than for number.

If knowledge and *access to number words* is the primary predictor of performance in estimation tasks, we should expect that – given an identical set of target number words – estimation performance should be largely identical across dimensions. In other words, we should expect that the very moment children have acquired number words and understand their cardinality, they should be equally good at estimating across number, length, and area, especially if the range of presented values is held constant across the three dimensions.

Finally, if estimation performance is primarily predicted by the ability to align, and maintain alignment between, perceptual states and number words (Sullivan & Barner, 2013; Yeo & Price, 2020), we should find that estimation performance should strongly correlate across number, length, and area. That is, we should find that how well a child is able to structure map (i.e., to maintain an interface between perception and number words) in one perceptual magnitude correlates with how well they are able to do this for other perceptual magnitudes.

Here, we report data from 5- to 12-year-olds, who each completed a set of perceptual discrimination tasks (measuring their ANS, length, and area perceptual acuity, and confirming that ANS precision is much worse), followed by a set of estimation tasks. We show that from age five onward, children can instantly generate an interface between number words and perceptual magnitudes, even in dimensions that are not well practiced (e.g., length, area; Experiments 1 and 2) and for novel units (i.e., mapping *three* dots to "one" number; Experiment 2), and that neither perceptual acuity nor access to number words are sufficient for explaining differences in estimation performance. Instead, we find correlations in children's estimation ability across the three perceptual dimensions, broadly supporting the idea that verbal estimation performance is primarily driven by children's ability to structure map number words to perceptual states.

Experiment 1

Methods & Procedure

Participants Ninety children (47 males) between the ages of 5- to 12-years-old (M = 8;9 [years; months], range = 5;1 to 12;11) completed two tasks: a Discrimination Task (to test the precision of participants' perceptual representations) and an Estimation Task (to explore the quality of the interface between these perceptual representations and number words). We chose this age range as some work suggests that from age five onward, children are able to map their intuitive representations (e.g., ANS) to language (e.g., number words) (Le Corre & Carey, 2007; Odic et al., 2015). Moreover, we include a wide age range for the primary purposes of exploring whether the overall patterns in discrimination and estimation performance (e.g., whether number is better/worse than length and/or area across tasks) differ across development, though our focus is not on charting the developmental time course for improvement.

To create more balanced categories in age (as is necessary in order for the assumptions of the ANOVAs to be met), we binned participants into 4 primary age categories: "5-year-olds" (range = 5;0 – 6;11, M = 6;0, n = 20), "7-year-olds" (7;0 – 8;11, M = 7;11, n = 25), "9-year-olds" (9;0 – 10;11, M = 9;11, n = 31), and "11-year-olds" (11;0 – 12;11, M = 11;11, n = 14). These group sample sizes are consistent with typical estimation tasks (Libertus et al., 2016; Sullivan & Barner, 2014; Izard & Dehaene, 2008). An additional 14 children were tested, but failed to complete both tasks in full and were excluded. All participants were individually tested in a quiet room on a 13" MacBook Air running custom-made Psychtoolbox-3 scripts.

Discrimination Task Participants were first presented with a 2AFC discrimination task across the three dimensions in which they had to judge which set of items was more numerous (i.e., "which side has more dots?", "which line is longer?", "which blob is bigger?") modelled after Odic (2018). Participants were asked to indicate their answers verbally or by pointing. To minimize the influence of motor control on performance, children's responses were recorded on the computer by the experimenter who pressed buttons corresponding to the side of the screen children chose.

Stimuli were shown on the screen for 500 milliseconds. Previous research has demonstrated that this is sufficiently long enough for children to view the display, but also sufficiently short to prevent counting (Odic, 2018). Following 6 practice trials (2 consecutive trials per dimension), participants completed 192 trials (64 per dimension) in an intermixed order across dimensions, to eliminate potential order effects. To control difficulty, each trial varied in one of four ratios, which were identical across dimensions: 2.0 (e.g., 20 vs. 10 dots), 1.5, 1.20, and 1.07 (e.g., 16 vs. 15 dots). Consistent with previous work (Libertus et al., 2016; Wang et al., 2016; Odic, 2018), computer-generated auditory feedback was given based on the participants' performance on the task (e.g., "Good job!" for correct answers, "Oh, that's not right!" for incorrect answers). This was done to help maintain attention (Wang et al., 2016), and, particularly for the number dimension, to discourage children from using non-numeric cues (e.g., area) to make judgments (Dramkin et al., 2020; DeWind & Brannon, 2012). For each dimension, the dependent variable was accuracy (i.e., the portion of trials in which participants correctly identified the side with the greater quantity).

Estimation Task Following the discrimination task, participants were presented with different stimuli and were asked to *verbally* assign a number word to an amount shown (i.e., estimate in terms of number, length, or area "how many" items they saw). In typical number-line estimation tasks, participants indicate where a value falls with respect to specific endpoints/estimation boundaries on a number line (c.f., Siegler & Booth, 2004). This requires a three-way interface between number words, the ANS, and spatial representations, and are known to be susceptible to a host of response biases (Slusser & Barth, 2017). In contrast, our *verbal estimation* method provides no information about the range or distribution of possible responses and relies on the pure interface between number words and perception, with no intermediate mapping to space (Stevens, 1946).

To control for prior experience participants may have had with particular units (e.g., cm, in^2), we created novel units and demonstrated how to use them across 6 training trials (2 consecutive trials per dimension) – for number, a single dot was called one "toma"; for length, a single line 44px long was called one "blicket"; for area, a $111px^2$ blob was called one "modi" (**Figure 1**). Children were then trained on estimating using these novel units for target numbers 2, 3, and/or 4, though critically the training trials did not use quantities that were shown during the test trials. All three dimensions also utilized the same range of target values, thereby equating any differences in children's accessibility for number words.

Subsequently, participants completed 96 estimation trials (32 test trials per dimension) in an intermixed order. They were shown the novel "standard" unit at the bottom of the screen, and a target (e.g., a more numerous set of dots, longer line, or bigger blob) would appear on the screen for 500 ms (too quick to count) to be kept consistent with the discrimination task. Each participant was then asked to verbally estimate "how many *tomas*/*blickets*/*modies*" they saw, with the use of the quantifier (i.e., how "many") and plural syntax (e.g., mod*ies*) guiding them to providing us with a number word. Estimates were recorded by the researcher.

The target values for each test trial were 5 (e.g., a blob $555px^2$ big), 8 (e.g., a line 352px long), 13 (e.g., 13 dots), or 21. Participants were not given feedback during the test trials as previous work has shown that this can readily influence or calibrate the entire range of participants' responses (Izard & Dehaene, 2008; Sullivan & Barner, 2014). If, however, children provided nonsensical estimates, such as a combination of multiple numbers (e.g., "eleventy-four") or values that exceeded what was reasonably possible (e.g., "one billion"), or answered "one" or "one big one," participants were prompted to provide "their next best guess".

To capture differences in *estimation accuracy* we calculated the absolute average error rate (AER) (i.e., the absolute average percent difference between the true value shown and the estimate provided by children) (Crollen et al., 2011; Odic et al., 2015). Normally, negative error rates indicate under-estimation, while positive error rates indicate over-estimation. Here, we were concerned with the overall degree of under- or over-estimation, but not the direction. Hence, scores closer to 0 are indicative of a more accurate interface between number words and perceptual magnitudes.

To capture *estimation variability*, we calculated each child's average coefficient of variation (CV) (i.e., the standard deviation for the responses divided by the mean response) (Cordes et al., 2001; Odic et al., 2015). Values closer to 0 indicate less variable (i.e., "better") estimates. Note that estimates below "two" and those that were above or below 3 SD from each participant's average guess were excluded from analyses (2.8% of data). Also, due to the non-parametric nature of these variables, we report Spearman's *rho* for AERs and CVs.

Results

Discrimination Task A 4 (Age Group: 5, 7, 9, 11) x 3 (Dimension: Number, Length, Area) x 4 (Ratio: 1.07, 1.2, 1.5, 2.0) Greenhouse-Geisser corrected Mixed Measures ANOVA over accuracy as the dependent variable (DV) showed a main effect of Dimension, $F(1.65, 141.84) = 84.12$, $p < .001$, $\eta^2_G = .16$, a main effect of Ratio, $F(2.33, 200.25) = 398.091$, $p < .001$, $\eta^2_G = .51$, and a main effect of Age, $F(3, 86) = 18.60$, $p < .001$, $\eta^2_G = .08$. We also found a Dimension x Ratio interaction, $F(4.15, 357.03) = 12.44$, $p < .001$, $\eta^2_G = .06$, but no Age x Dimension interaction, $F(4.95, 141.84) = 0.83$, $p = .528$, $\eta^2_G = .005$, no Age x Ratio interaction, $F(6.99, 200.25) = 1.26$, $p = .273$, $\eta^2_G = .01$, and no Dimension x Ratio x Age interaction, $F(12.45, 357.03) = 0.87$, $p = .581$, $\eta^2_G = .01$.

Post-hoc Tukey-corrected contrasts showed children performed the worst on Number (M = 77.95, SD = 9.12) compared to Length (M = 89.24, SD = 6.09), $t(172) = -10.60$, $p < .001$, and Area (M = 87.81, SD = 6.90), $t(172) = -11.95$, $p < .001$, but Area and Length discrimination accuracy did not significantly differ from each other, $t(172) = -1.95$, $p = .250$. When treated continuously, we find strong correlations between age and accuracy in Number, Pearson's $r = .42$, $p < .001$, Length, Pearson's $r = 0.43$, $p < .001$, and Area, Pearson's $r = 0.46$, $p < .001$.

These results replicate previous work (e.g., Odic, 2018) showing that perceptual magnitudes obey Weber's law (i.e., show ratio effects), improve with age, and that number perception is significantly worse, especially at harder ratios, than length and area perception.

Estimation Task A 4 (Age Group: 5, 7, 9, 11) x 3 (Dimension: Number, Length, Area) Greenhouse-Geisser corrected Mixed Measures ANOVA with absolute average error rates (AER) as the DV revealed a main effect of Dimension, $F(2.00, 171.65) = 215.07$, $p < .001$, $\eta^2_G = .06$, no main effect of Age, $F(3, 86) = 1.37$, $p = .256$, $\eta^2_G = .02$, but a Dimension x Age interaction, $F(5.99, 171.65) = 3.03$, $p = .008$, $\eta^2_G = .05$. Tukey-corrected post-hoc tests showed that at the group level, participants had the lowest AER (i.e., best estimation accuracy) on the Number trials (M = 0.09, SD = 0.07) compared to Length (M = 0.36, SD = 0.16), $t(172) = -15.62$, $p < .001$, or Area (M = 0.46, SD = 0.16), $t(172) = -19.63$, $p < .001$ (**Figure 1**). In turn, Area AER was worse than Length AER, $t(172) = 4.00$, $p = .0003$, though this effect was

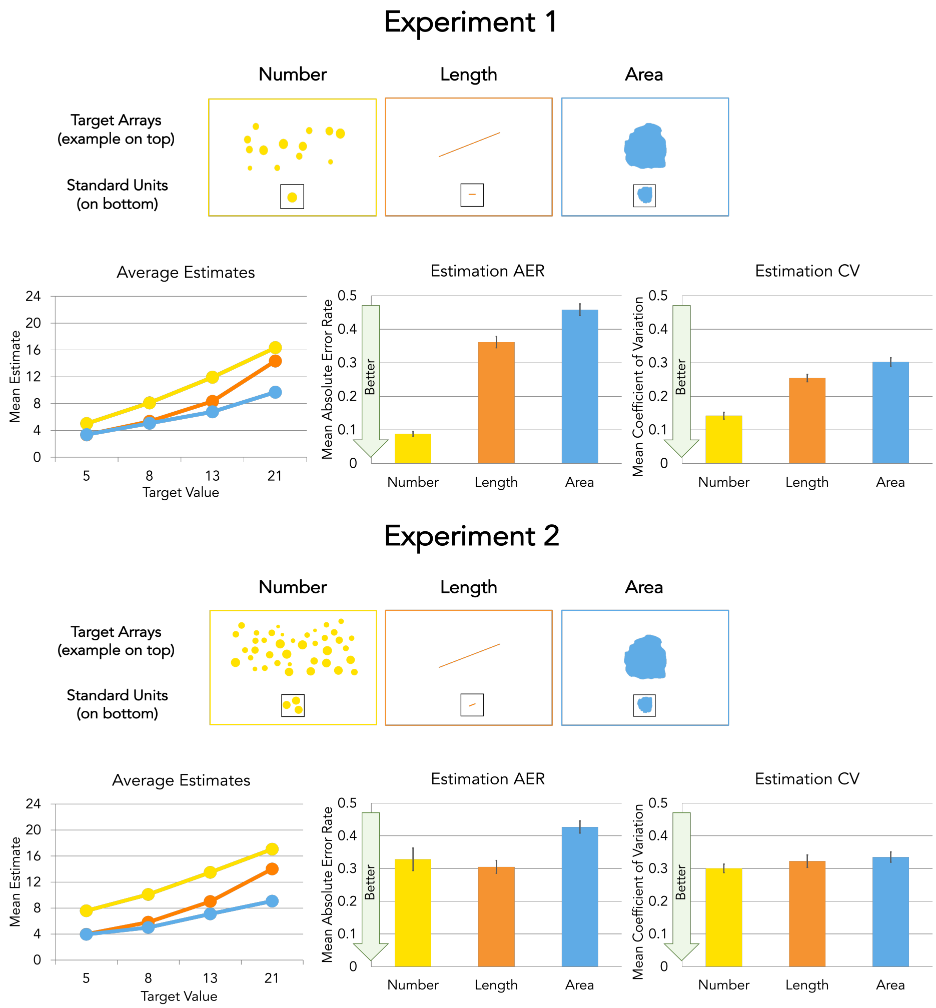

Figure 1. Estimation stimuli and results. Standard error bars for mean AERs and CVs are shown.

much stronger for 9- and 11-year-olds compared to 5- and 7-year-olds, whose Length AER values, $t(172) = -0.24$, $p = .969$, and Area AER values, $t(172) = 0.60$, $p = .819$, were identical. When treated continuously, we also find correlations between age and Length, Spearman's *rho* = -.29, $p = .005$, marginal correlations with Number, Spearman's *rho* = -.20, $p = .063$, but no correlations between age and Area, Spearman's *rho* = .03, $p = .787$. When controlling for age, we find that estimation accuracy (AER) is significantly correlated between Length and Area, Partial Spearman's *rho* = .42, $p < .001$, but not Number and Length, Partial Spearman's *rho* = 0.03, $p = .750$, nor Number and Area, Partial Spearman's *rho* = .08, $p = .445$.

Next, we conducted a 4 (Age Group: 5, 7, 9, 11) x 3 (Dimension: Number, Length, Area) Greenhouse-Geisser corrected Mixed Measures ANOVA with coefficient of variation (CV) as the DV. This revealed a main effect of Dimension, $F(1.86, 159.97) = 92.10$, $p < .001$, $\eta^2_G = .29$, a main effect of Age, $F(3, 86) = 10.29$, $p < .001$, $\eta^2_G = .18$, and no significant Dimension x Age interaction, $F(5.58, 159.97) = 1.26$, $p = .283$, $\eta^2_G = .02$. As shown in **Figure 1**, the patterns of estimation variability are similar to those observed for estimation accuracy (AER).

Post-hoc Tukey-corrected contrasts revealed the lowest variability (i.e., best performance) was in Number ($M = 0.14$, $SD = 0.10$), which was significantly better (i.e., lower) than Area ($M = 0.30$, $SD = 0.12$), $t(172) = -13.37$, $p < .001$, and Length ($M = 0.25$, $SD = 0.10$), $t(172) = -8.73$, $p < .001$. In turn, Area variability was much greater (i.e., worse) than Length, $t(172) = 4.64$, $p < .001$. We also find strong correlations with age (when treated continuously) across dimensions: Number, Spearman's *rho* = -.56, $p < .001$, Length, Spearman's *rho* = -.28, $p < .007$, and Area, Spearman's *rho* = -.35, $p < .001$. However, when controlling for age-related improvements, we find that estimation variability (CV) is significantly correlated between Number and Length, Partial Spearman's *rho* = .31, $p = .002$, Number and Area, Partial Spearman's *rho* = .50, $p < .001$, and Length and Area, Partial Spearman's *rho* = .33, $p < .001$.

Hence, for both estimation AERs and CVs, number was far better compared to that of length and area, and critically, estimation performance in one dimension, predicted estimation performance in another, especially in the case in of variability.

Discussion

The results of Experiment 1 demonstrate several important findings. First, although worse than number estimation, children as young as five were able to successfully map completely novel units to their perception of length and area. Since these novel units could not have had a previously established associative mapping, this implies that children who have formed an interface between number words and the ANS can readily form novel structural mappings with other perceptual dimensions as well, contrary to some theories that argue that associative mappings are pre-requisites for structure mappings to form (Yeo & Price, 2020). Moreover, how consistent children were in their structure mappings in any dimension predicted their ability to maintain those mappings in other dimensions, as demonstrated by the correlations in estimation variability across number, length, and area.

At the same time, children were far more accurate and less variable in number compared to length and area. Given that number perception is reliably *worse* than length or area perception, and that the range of target number words was identical across the tasks, this suggests that children are specifically better at interfacing number words with the ANS compared to other dimensions, and that estimation performance is not primarily predicted by the number word access *nor* by the acuity of the underlying perceptual systems (since, if it was, we would have expected area and length estimation to be more precise than number estimation).

What might drive this advantage for number? One possibility is that children have well-practiced associative mappings between their ANS representation and number words. After all, children first learn to attach number words to the ANS. Therefore, if we forced children to use entirely novel units for number, we should expect that this advantage in number estimation would disappear.

On the other hand, perhaps children's advantage in number estimation stems from a more natural link between number words and the ANS. For instance, children may more intuitively draw connections between the representational content of the ANS and number words, as opposed to non-numeric dimensions (e.g., length, area). In this case, we should expect that no matter what units children are asked to use in the number estimation task, they would continue to show a robust advantage in number estimation compared to length or area.

In Experiment 2, we test these possibilities by providing children with units for number estimation that are not well-practiced or associative (i.e., labelling a set of *three* dots as a single unit, "*one* toma"), allowing us to examine whether their better interface with number abilities is a byproduct of experience children have with mapping number to single object units.

Experiment 2

Methods & Procedure

Participants Eighty-four children (44 males) between the ages of 5- to 12-years-old (M = 8;1 [years; months], range = 5;0 to 12;8) completed a number/length/area Discrimination Task and an Estimation Task. An additional 7 children were tested, but failed to complete both tasks in full. None of the participants in Experiment 2 had previously taken part in Experiment 1. As with Experiment 1, we also binned participants into 4 primary age categories for the purposes of analysis: "5-year-olds" (range = 5;0 – 6;11, M = 5;11, n = 30), "7-year-olds" (7;0 – 8;11, M = 7;11, n = 26), "9-year-olds" (9;0 – 10;11, M = 10;11, n = 19), and "11-year-olds" (11;0 – 12;11, M = 11;11, n = 9).

Discrimination Task The procedure and stimuli for the discrimination task was the same as Experiment 1. For each dimension, the dependent variable was accuracy (percentage correct).

Estimation Task The estimation task and stimuli were the same as in Experiment 1, with two exceptions. For number, participants were introduced to a three-dot "toma" to represent a single unit for number (**Figure 1**). The target values for each test trial were 5, 8, 13, or 21, which were actually represented by 15, 24, 36, and 63 dots, respectively. Thus, while the range of target number words was still identical across the three dimensions and Experiment 1, children's associative mappings for number estimation were not consistent with their prior experiences of labelling these sets. Additionally, the standard unit for length was made to rotate across trials so that the orientation matched to that of the target. This was in contrast to Experiment 1 where the standard unit stayed at the same orientation throughout all trials, which may have resulted in the targets being harder to estimate (Wenderoth & White, 1979). 2.1% of trials were excluded as outliers.

Results

Discrimination Task A 4 (Age Group: 5, 7, 9, 11) x 3 (Dimension: Number, Length, Area) x 4 (Ratio: 1.07, 1.2, 1.5, 2.0) Greenhouse-Geisser corrected Mixed Measures ANOVA over accuracy as the dependent variable (DV) revealed a main effect of Dimension, $F(1.82, 145.71) = 49.56$, $p < .001$, $\eta^2_G = .09$, a main effect of Ratio, $F(2.26, 180.53) = 315.42$, $p < .001$, $\eta^2_G = .47$, and a main effect of Age, $F(3, 80) = 31.74$, $p < .001$, $\eta^2_G = .15$. We also found a Dimension x Ratio interaction, $F(4.41, 352.74) = 4.31$, $p = .001$, $\eta^2_G = .03$, a Dimension x Age interaction, $F(5.46, 145.71) = 3.35$, $p = .005$, $\eta^2_G = .02$, and a Dimension x Ratio x Age interaction, $F(13.23, 352.74) = 1.99$, $p = .020$, $\eta^2_G = .03$, but no Ratio x Age interaction, $F(6.77, 180.53) = 1.65$, $p = .128$, $\eta^2_G = .01$.

As with Experiment 1, Tukey-corrected post-hoc contrasts showed that accuracy on Number (M = 79.58, SD = 10.30) was worse compared to Length (M = 87.69, SD = 6.22), $t(160) = -8.78$, $p < .001$, and Area (M = 87.39, SD = 7.69), $t(160) = -8.46$, $p < .001$. However, performance on Area and Length trials were not significantly different from each other, $t(160) = -0.32$, $p = .945$. This same pattern of results was observed when comparing accuracy across dimensions at the age group levels, with the only exception being the 11-year-olds, who showed no significant difference in their accuracy for Number and Length, $t(160) = -1.85$, $p = .157$. Nevertheless, we replicate that number discrimination is worse than that of either length or area. Finally, when treated continuously, discrimination accuracy correlated with age: Number, Pearson's $r = .63$, $p < .001$, Length, Pearson's $r = .49$, $p < .001$, and Area, Pearson's $r = .52$, $p < .001$.

Estimation Task. A 4 (Age Group: 5, 7, 9, 11) x 3 (Dimension: Number, Length, Area) Greenhouse-Geisser corrected Mixed Measures ANOVA with AER as the DV revealed a main effect of Dimension, $F(1.84, 146.87) = 63.32$, $p < .001$, $\eta^2_G = .19$, no main effect of Age, $F(3, 80) = 0.24$, $p = .866$, $\eta^2_G = .01$, but a Dimension x Age interaction, $F(5.51, 146.87) = 3.03$, $p = .004$, $\eta^2_G = .05$. In contrast to Experiment 1, we found that participants had the lowest AER in Length (M = 0.31, SD = 0.18), and that they were significantly more accurate in their Length estimates compared to Number (M = 0.33, SD = 0.31), $t(160) = 3.61$, $p < .001$, and Area (M = 0.43, SD = 0.17), $t(160) = -7.47$, $p = .001$, but they still performed better on Number trials compared to Area, $t(160) = -11.04$, $p < .001$ (**Figure 1**). We found that this difference in Length and Number estimation was not significant for the older children (i.e., 9-year-olds, 11-year-olds). When treated continuously, we also found that age strongly correlated with estimation accuracy in Number, Spearman's $rho = -.23$, $p = .032$, marginally with Length, Spearman's $rho = -.20$, $p = .069$, but not with Area, Spearman's $rho = .12$, $p = .263$. Finally, when controlling for age, estimation accuracy (AER) significantly correlated in Length and Area, Partial Spearman's $rho = -.47$, $p < .001$, but not Number and Length, Partial Spearman's $rho = -.17$, $p = .114$, nor Number and Area, Partial Spearman's $rho = .15$, $p = .184$.

A 4 (Age Group: 5, 7, 9, 11) x 3 (Dimension: Number, Length, Area) Greenhouse-Geisser corrected Mixed Measures ANOVA with the CVs as the DV showed a main effect of Dimension, $F(1.99, 159.46) = 3.22$, $p = .043$, $\eta^2_G = .18$, a main effect of Age, $F(3, 80) = 20.18$, $p < .001$, $\eta^2_G = .35$, and a Dimension x Age interaction, $F(5.58, 159.46) = 3.13$, $p = .006$, $\eta^2_G = .03$. Consistent with the patterns observed in Experiment 1, we also found Number (M = 0.30, SD = 0.12) to be less variable than Area (M = 0.34, SD = 0.14), $t(160) = -2.54$, $p = .033$, but not when compared to Length (M = 0.32, SD = 0.17), $t(160) = -1.35$, $p = .373$); variability for Area and Length, in turn, were not different from each other, $t(160) = 1.19$, $p = .461$ (**Figure 1**). Thus, in contrast to Experiment 1, we find that children did not robustly demonstrate a better interface with number when provided with an unpracticed unit. However, much like Experiment 1, when treated continuously, we observed improvements in estimation variability (CV) with age: Number, Spearman's $rho = -.48$, $p < .001$, Length, Spearman's $rho = -.64$, $p < .001$, and Area, Spearman's $rho = -.54$, $p < .001$. When controlling for age-related improvements in CV, Number and Length, Partial Spearman's $rho = .56$, $p < .001$, Number and Area, Partial Spearman's $rho = .50$, $p < .001$, and Length and Area, Partial Spearman's $rho = .66$, $p < .001$, were significantly correlated.

Hence, we again find that children's ability to maintain consistent mappings (CVs) are significantly correlated across dimensions, and that area estimation was still the worst; however, at best, number estimation performance is only equal to that length in Experiment 2.

To further characterize the differences in estimation performance across Experiments, we performed a 2 (Experiment: 1, 2) x 3 (Dimension: Number, Length, Area) Greenhouse-Geisser corrected ANOVA with AER as the DV, and found a main effect of Experiment, $F(1, 172) = 8.05$, $p < .001$, $\eta^2_G = .02$, a main effect of Dimension, $F(1.59, 272.95) = 73.59$, $p < .001$, $\eta^2_G = .21$, and an Experiment x Dimension interaction, $F(1.59, 272.95) = 36.165$, $p < .001$, $\eta^2_G = .111$, with Tukey-corrected post-hoc contrasts revealing a significant drop in Number AER across Experiments, $t(201) = -7.59$, $p < .001$, as well as a small improvement in Length AER, $t(201) = 2.24$, $p = .026$, which may have been due to matching the orientations between the length target and standard unit in Experiment 2, but not Experiment 1. A 2 (Experiment: 1, 2) x 3 (Dimension: Number, Length, Area) Greenhouse-Geisser corrected ANOVA with CV as the DV showed a main effect of Experiment, $F(1, 172) = 26.39$, $p < .001$, $\eta^2_G = .10$, a main effect of Dimension, $F(1.97, 339.67) = 69.91$, $p < .001$, $\eta^2_G = .09$, and an Experiment x Dimension interaction, $F(1.97, 339.67) = 26.39$, $p < .001$, $\eta^2_G = .10$, with Number CVs, $t(201) = -6.89$, $p < .001$, and Length CVs, $t(201) = -2.55$, $p = .011$, being *worse* in Experiment 2.

General Discussion

We explored which factors predict how well children verbally estimate across number, length, and area. We report three major findings.

First, we find that from age five onward, children can readily form mappings between number words and their perceptual magnitude systems, even when given entirely new units: across all experiments and dimensions, children consistently gave higher estimates for higher values (see Figure 1). Children were even able to do this in Experiment 2, where they had to treat *three* items as a single number word (i.e., estimating that there were only "five" tomas when shown 15 dots). This is even more remarkable when considering that children had potentially competing associative mappings for number that had to be inhibited (e.g., a strong link between word "one" and *1* item), and that the youngest children in our study have not yet been formally taught how to multiply or divide. This powerfully demonstrates the importance of mechanisms akin to structure mapping that must underlie at least part of the interface between number words and perceptual magnitudes, and that robust associative links are not necessary for this interface (Yeo & Price, 2020), as children were able to rapidly and spontaneously generate estimates for novel units and uncommon target values.

Second, we find that children's estimation performance was not primarily predicted by their *perceptual acuity* nor *number word accessibility* and knowledge (which was kept identical across number, length, and area estimation and the two experiments). Even in Experiment 2, we find that number estimation abilities were still on-par – or in the case of variability, slightly better – with length and area, despite the large differences in discrimination precision between number, length, and area.

Finally, while estimation performance was predicted by children's ability to maintain structure mappings across dimensions, as demonstrated by strong correlations in estimation variability in both experiments, we found that a large portion of the advantage children showed for number estimation in Experiment 1 is likely due to well-practiced units. That is, when children were asked to estimate number using novel units in Experiment 2 (i.e., *"one* toma" referring to *three* dots), their estimation accuracy and variability became substantially worse. Therefore, while structure mappings can be rapidly deployed for novel units, we also found evidence that children benefit from well-practiced links between number words and the ANS.

Taken together, our results broadly contribute to theories of how children reason about language and perception. Beyond finding that the acquisition and the improvements in the interface are not driven by mere perception nor access to number words, we show that it is also not built in piecemeal: children can rapidly and easily create novel interfaces for dimensions and units they have never used number words for. This supports a structure-mapping account not only for at least part of interface between the ANS and number words, but also for how children interface number words with other perceptual magnitudes.

References

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*(4), 698–707.

Crollen, V., Castronovo, J., & Seron, X. (2011). Under-and over-estimation: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Experimental Psychology*, *58*(1), 39.

DeWind, N. K., & Brannon, E. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience*, *6*, 68.

Dramkin, D., Bonn, C., Baer, C., & Odic, D. (2020). *The malleable role of non-numeric features in visual number perception.* [Manuscript in preparation]. Department of Psychology, University of British Columbia.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–70.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116–20.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–47.

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

Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *Journal of Experimental Child Psychology*, *150*, 207–26.

Matthews, P. G., Lewis, M. R., & Hubbard, E. M. (2016). Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychological Science*, *27*(2), 191–202.

Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science*, *21*(2), e12533.

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, *138*, 102–21.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–44.

Slusser, E., & Barth, H. (2017). Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks. *Journal of Experimental Child Psychology*, *162*, 181–98.

Starr, A., & Brannon, E. M. (2015). Developmental Continuity in the Link Between Sensitivity to Numerosity and Physical Size. *Journal of Numerical Cognition*, *1*(1), 7–20.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-80.

Sullivan, J., & Barner, D. (2013). How are number words mapped to approximate magnitudes? *The Quarterly Journal of Experimental Psychology*, *66*(2), 389–402.

Sullivan, J., & Barner, D. (2014). Inference and association in children's early numerical estimation. *Child Development*, *85*(4), 1740–55.

Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99.

Wenderoth, P., & White, D. (1979). Angle-Matching Illusions and Perceived Line Orientation. *Perception*, *8*(5), 565–75.

Yeo, D. J., & Price, G. R. (2020). Probing the mechanisms underlying numerosity-to-numeral mappings and their relation to math competence. *Psychological Research*, 1–24.