This is the accepted version of the article, but might differ from the final archived version, which can be found at: https://doi.org/10.1016/j.cognition.2022.105149

Children's Understanding of Most is Dependent on Context

Michelle A. Hurst* & Susan C. Levine

University of Chicago

*Corresponding Author:

Michelle A. Hurst
hurstm@uchicago.edu
hurst.mich@gmail.com
5848 S University Ave
Chicago IL 60637

Abstract

Children struggle with the quantifier "*most*". Often, this difficulty is attributed to an inability to interpret *most* proportionally, with children instead relying on absolute quantity comparisons. However, recent research in proportional reasoning more generally has provided new insight into children's apparent difficulties, revealing that their overreliance on absolute amount is unique to contexts in which the absolute amount can be counted and interferes with proportional information. Across two experiments, we test whether 4- to 6-year-old children's interpretation of *most* is similarly dependent on the discreteness of the stimuli when comparing two different quantities (e.g., who ate most of their chocolate?) and when verifying whether a single amount can be described with the term *most* (e.g., is most of the butterfly colored in?). We find that children's interpretation of *most* does depend on the stimulus format. When choosing between absolutely more vs. proportionally more as depicting *most*, children showed stronger absolute-based errors with discrete stimuli than continuous stimuli, and by 6-years-old were able to reason proportionally with continuous stimuli, despite still demonstrating strong absolute interference with discrete stimuli. In contrast, children's yes/no judgements of single amounts, where conflicting absolute information is not a factor, showed a weaker understanding of *most* for continuous stimuli than for discrete stimuli. Together, these results suggest that children's difficulty with *most* is more nuanced than previously understood: it depends on the format and availability of proportional vs. absolute amounts and develops substantially from 4- to 6-years-old.

Keywords: *proportion; most; quantifiers; numerical interference*

Children's Understanding of Most is Dependent on Context

1. Introduction

Children's understanding of quantifiers – terms that denote amounts of quantities, such as *most, some,* and *all* – has been of deep interest to linguists and numerical cognition researchers for decades. This work has led to substantial insight into children's language learning, such as their use of scalar implicature and pragmatics (e.g., Barner et al., 2011; Horowitz et al., 2018; Noveck, 2001), as well as children's developing understanding of numbers and quantities (e.g., Barner, Chow, et al., 2009; Hurewitz et al., 2006). The quantifier *most* appears to be particularly challenging, both for children and for researchers. One reason for this difficulty is that *most* cannot be expressed using first-order logical predicates and instead requires comparing relative magnitudes (Barwise & Cooper, 1981). Although researchers have recognized this relational structure of *most,* their focus has been on comparisons of numerical quantities, whereas the question of how children's understanding of *most* is related to their conceptualization of proportional information has been overlooked. Importantly, a recent surge of research in young children's proportional reasoning has provided insight into the features that can facilitate or hinder proportional reasoning (Begolli et al., 2020; Davis, 2003; Jeong et al., 2007; Möhring et al., 2016). For example, one specific feature known to impact children's proportional reasoning is whether the proportional information is based on discrete numerical quantities (e.g., sets of dots) or on continuous area-based quantities. Children are less likely to make proportion-based judgements with discrete proportions compared to continuous proportions across a range of contexts, including when the proportions represent juice mixtures (Boyer et al., 2008), probabilistic game spinners (Hurst & Cordes, 2018; Jeong et al., 2007), and sharing scenarios (Hurst et al., 2020). In the current study, we test whether children's interpretation of the quantifier *most* also depends on this perceptual feature (i.e., differs for discrete versus continuous quantities), thereby addressing both methodological and theoretical questions about the development of children's understanding of quantifiers in language and children's understanding of proportional reasoning – two topics that are central to human cognition.

1.1 The Quantifier *Most*

Research investigating the development of children's understanding of *most* has provided conflicting evidence, with varying accounts of the age at which children understand its meaning. Some work has shown success at understanding *most* as early as age 3 (Halberda et al., 2008). For example, when given a set of objects with the majority blue or yellow, Halberda and colleagues found that 3-year-old children were readily able to answer the question "Are most of the objects blue or yellow?" In contrast, other work suggests that children do not reach an adult-like understanding of *most* until after 6-years-old. When asked to provide *most* objects in a set, 2- to 5-year-old children were unable to do so (Barner, Chow, et al., 2009; Barner, Libenson, et al., 2009). Furthermore, when verifying sets and indicating whether or not they can be described with *most* many 6- to 8-year-old children (about 30%) accepted values below half (e.g., 1/6) as being *most* (Papafragou & Schwarz, 2006).

One critical difference between these studies with conflicting findings is whether children were able to draw upon their understanding of *more* to make judgements about *most*. In the paradigm used by Halberda and colleagues (2008), children could respond based on which set was more (blue or yellow objects), whereas children tested by Barner and colleagues (Barner, Chow, et al., 2009; Barner, Libenson, et al., 2009) and Papafragou and Schwarz (2006) were not able to use this strategy to help them. In a recent study, Sullivan and colleagues (Sullivan et al., 2018) directly tested this hypothesis by comparing 4- to 6-year-olds' ability to select the appropriate set for *most* when it was directly pitted against an absolute *more*-based interpretation. For example, children were asked who popped most of their balloons, a character who popped 3 out of 5 or a character who popped 4 out of 9. Sullivan and colleagues (2018) found that 4- to 6-year-old children tended to select the character who popped more balloons (popped 4 balloons, rather than 3), even though that character did not pop most of their balloons (4/9 is fewer than half, whereas 3/5 is more than half). Together, this work suggests that children do not have a "more than half" relational understanding of the term *most* until beyond 6-years-old, and instead that young children may interpret *most* to mean *more* in a comparison context (as described above).

Notably, this discussion raises two distinct, but related, aspects of children's understanding of the quantifier *most*: (a) whether they can take a proportional "more than half" interpretation and (b) whether they distinguish *most* from *more.* Here, we focus on the first aspect by drawing upon recent findings from the proportional reasoning literature more broadly. Specifically, the interference between absolute amount comparisons and proportional comparisons is a well-established phenomenon in the proportional reasoning literature, even in the absence of specific quantifier vocabulary. However, we return to a discussion of the second aspect (i.e., whether children distinguish between *most* and *more*) in the General Discussion (Section 4.1.2), including how our findings speak to this issue.

1.2 Numerical Interference in Proportional Reasoning

Research findings suggest that infants and young children have sophisticated proportional reasoning abilities. For example, 6-month-olds can be habituated to specific proportions (McCrink & Wynn, 2007) and can make inferences about the probable outcome of sampling based on the proportional distribution of a population (e.g., demonstrating surprise when a low-probability outcome is randomly sampled from a bin; Denison et al., 2013). Despite this early-developing sensitivity to proportion, however, 6-year-old children show systematic errors in their proportional reasoning. Specifically, when visual proportions are presented as relations of continuous amounts, such that the proportional information is based on area and not number, 6-year-olds are able to select the larger of two proportions (Hurst & Cordes, 2018; Jeong et al., 2007) and match equivalent proportions (Boyer et al., 2008). However, when visual proportions are presented as relations of discrete sets, such that proportional information is conveyed with countable numbers of objects or parts, 6-year-olds make systematic errors, such as erroneously deciding that 4/9 is more than 3/5 or matching 4/9 with 4/5 rather than with 8/18 (Boyer et al., 2008; Hurst & Cordes, 2018; Jeong et al., 2007). These errors are often attributed to heightened attention to absolute numerical information, potentially through counting strategies, which interfere with children's proportional reasoning (Boyer et al., 2008). Furthermore, this numerical interference in the context of discrete quantities hinders proportional reasoning in a range of contexts, for example comparing the probability a game spinner will landing on a given outcome (Hurst & Cordes, 2018; Jeong et al., 2007), matching equivalent tasting mixtures of juice and water (Boyer et al., 2008), and making judgements of the niceness of characters based on the proportional versus absolute amount they shared (Hurst et al., 2020). Thus, the existing literature on children's proportional reasoning is consistent with the error pattern found by Sullivan and colleagues (2018) in children's understanding of *most* because the task they employed used proportions of discrete countable sets, just the context that impedes children's reasoning about proportions, providing a more general explanation of their difficulty that is not specific to a misunderstanding of *most*.

1.3 The Current Study

In the current study, we ask how children's understanding of *most* is impacted by the perceptual saliency or availability of numerical information. Based on the findings that have emerged from the proportional reasoning literature, we hypothesized that children's difficulty with a proportional interpretation of *most* in discrete contexts is due to the discreteness of the stimuli, which leads to numerical interference, rather than a difficulty understanding *most* in general. Thus, in Experiment 1 we used the same paradigm as Sullivan et al. (2018) to contrast a proportional vs. absolute comparison interpretation of *most* but did so by comparing performance in a discrete context that emphasized numerical information (an almost exact replication of Sullivan et al., 2018) and a continuous context that removed countable numerical information. In Experiment 2, we replicate this same phenomenon with new stimuli and further test whether this difference between discrete and continuous stimuli arises exclusively in comparison contexts, in which there is the opportunity for interference from an absolute amount comparison, or whether it arises even in the absence of this conflict, and thus reflects a more general difference in the processing of discrete vs. continuous information. Thus, the goal of Experiments 1 and 2 is to investigate whether children can demonstrate a proportional interpretation of the quantifier *most.* If our hypothesis is correct, and children do have a proportional understanding of most but fail to demonstrate this understanding when numerical information interferes, then in the comparison tasks in Experiments 1 and 2 children should show a greater proportion-based interpretation of *most* with continuous stimuli (i.e., when numerical interference is not available) than with discrete stimuli (i.e., when an absolute numerical response is available). In contrast, in the verification task in Experiment 2, when the absolute number response is not in competition with the proportional response, children should be able to reason about *most* as "more than half" in the context of discrete as well as continuous quantities. Alternatively, if our hypothesis is not correct and children in the age range tested do not yet have a proportional interpretation of *most*, then even though they may succeed on the verification task in Experiment 2 (as children can succeed using non-proportional strategies), they should fail on the comparison task in Experiments 1 and 2 when comparing both continuous and discrete stimuli.

Together, these experiments allow us to address significant methodological and theoretical questions at the core of human cognition. If children's understanding of most is dependent on the perceptual characteristics of the stimuli, as we hypothesize, then theoretical accounts of children's quantifier knowledge based exclusively on discrete and countable stimuli may be missing important nuances in children's thinking and underestimate their proportional understanding of the quantifier *most*. By including both continuous and discrete stimuli and comparison and verification contexts (i.e., with and without the opportunity for numerical interference), we can build a more complete picture of children's understanding of *most*.

Moreover, given the ubiquity of proportional information in a range of cognitive domains, such as probabilistic reasoning, social cognition (e.g., sharing), and intuitions about physics (e.g., is the book likely to fall if most of it is on the table?), investigating the context-dependence of proportional reasoning can have far-reaching implications for understanding its development and how children draw upon their underlying conceptualization of proportion to guide their behavior. This is especially critical for developing theories of cognitive development that take context-dependence into account.

2. Experiment 1

2.1 Method

2.1.1 Participants

As pre-registered, our final sample consisted of 120 4- through 6-year-old children, M_{age} = 65 months, randomly assigned to one of the two block orders. Sixty children completed the continuous block of trials first: 20 4-year-olds, M_{age} = 53 months, Range: 48 to 60 months, 15 girls and 5 boys; 20 5-year-olds, M_{age} = 67 months, Range: 62 to 71 months, 11 girls and 9 boys; and 20 6-year-olds, M_{age} = 76 months, Range: 72 to 83 months, 9 girls and 11 boys; and 60 children completed the discrete block of trials first: 20 4-year-olds, M_{age} = 53 months, Range: 48 to 59 months, 9 girls and 11 boys; 20 5-year-olds, M_{age} = 66 months, Range: 61 to 70 months, 9 girls and 11 boys; and 20 6-year-olds, M_{age} = 78 months, Range: 73 to 83 months, 11 girls and 9 boys. Two additional children participated in at least some of the task but were excluded because of experimenter error (n = 1) or parental interference (n = 1).

Children were recruited from the greater Chicago, IL area, through local schools (37% of sample), a science museum (14% of sample), and our lab database (27% of sample), and tested in-person in their school, at the museum, or in our campus lab, respectively. In addition, some children, who were recruited through our lab database, Children Helping Science (childrenhelpingscience.com), and social media, were tested online via video-chat (23% of sample). Children were compensated with a small sticker or prize (except for when tested over video-chat), parents of children tested in our lab or online were compensated with $10, and schools were compensated with $50 for classroom supplies. All procedures were approved by the University of Chicago Institutional Review Board and parents provided informed consent.

Approximately 70% of the sample completed at least some of an additional demographic survey (demographic information was not collected from children at the museum and was optional for all other families). Based on this subsample, 6% reported being Asian, 20% Black or African American, 57% White, 9% more than one race (a combination of Black, African

American, American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, White, and Asian), and 7% reported Other. In addition, about 10% of children were identified as Hispanic or Latino/a/x. Most of the sample that reported demographics came from high-income homes, with 66% earning more than $100,000 per year, 20% reporting $50 - $99K per year, and 14% reporting less than $50K per year. Lastly, 84% of the parents who completed the demographic form (85% of whom were mothers) had at least a college degree.

2.1.2 Procedure and Stimuli

All children completed a Continuous block of trials and a Discrete block of trials and were randomly assigned, within their age group, to receive one of these blocks before the other. The stimuli differed across the two blocks, but the procedure was identical. Most children (77%) were tested in person and stimuli were presented on paper in a binder. The remaining 23% were tested online, due to COVID-19 halting all in person data collection, and the stimuli were presented via Microsoft PowerPoint[1]. Throughout, we note where modifications to the procedure and stimuli were needed to facilitate online data collection.

On each trial, children were told about two people who did something to some of their stuff and were asked who did that thing to *most* of their stuff, in the form of "who [verb] most of their [noun]s?", for example: "who broke most of their drums?". The discrete block was modeled after Sullivan and colleagues (2018) and introduced discrete stimuli using number words. For example, the experimenter would say "John played nine drums. Amanda played three drums. John broke four of them. Amanda broke two of them. Who broke most of their drums?"

[1] Children tested online scored higher than children tested in person. However, the online sample also included a higher proportion of 6-year-olds (25% of the in-person sample vs. 63% of the online sample), making it likely that this increase in performance is due to differences in age rather than differences in the testing medium. Importantly, however, data collection type did not interact with Block Type (the primary comparison of interest) and analyses with the in-person sample alone shows the same pattern as the combined sample.

Individual items (in this example, drums) were used to represent the objects, with the referenced subset (in this example, the broken drums) represented with X's over the objects (see Figure 1, left). The continuous block had the same structure as the discrete block but used the word "some" instead of number words and used stimuli consisting of continuous quantities. For example, the experimenter would say "John made some lemonade. Amanda made some lemonade. John sold some of it. Amanda sold some of it. Who sold most of their lemonade?" A single rectangle was used as the visual depiction of the continuous object or item (in this example, lemonade), with the corresponding referenced amount (in this case, the amount sold) shown with an X (see Figure 1, right). For in-person administration, the experimenter pointed to each quadrant while saying the phrase for that quadrant and children's verbal or pointing responses were accepted (e.g., saying "Amanda" or pointing to Amanda's quadrant). Since pointing is not feasible for online data collection, the quadrants had color coded frames around them so that John's quadrants had blue frames and Amanda's quadrants had yellow frames. While saying each phrase, power point animations had the colored frame appear to highlight the corresponding quadrant and during the question, both frames appeared on the relevant "end state" quadrants (Figure 1B). Children's verbal responses in terms of either the person (e.g., "Amanda") or the color of the frame (e.g., "yellow") was accepted. If children pointed to one of the quadrants, they were asked if they were pointing to John's blue box or Amanda's yellow box.

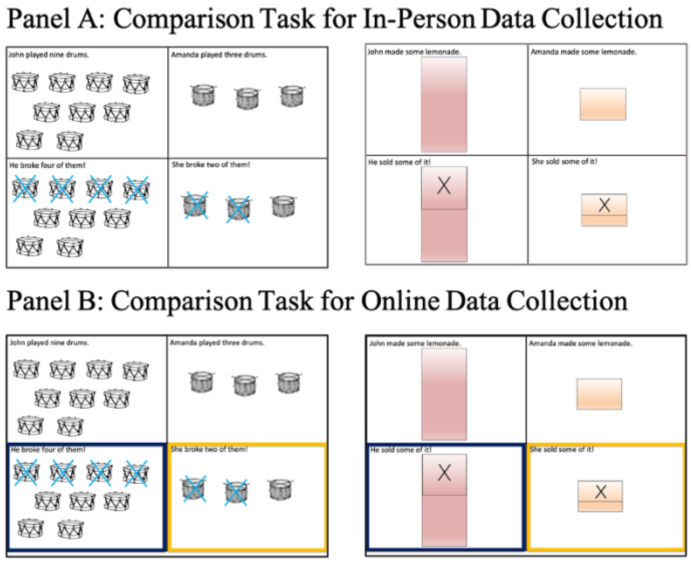

Figure 1: Example stimuli from the Discrete (left) and Continuous (right) blocks of trials, when presented in-person on paper (Panel A) and online via PowerPoint (Panel B). When presented online, the experimenter pointed to the quadrants rather than using color coded frames.

Each block had 8 trials that pit proportional responses against numerical responses. That is, one person always did proportionally the most (i.e., more than half), but numerically fewer and the other person always did numerically more, but not most of their own (i.e., less than half). The 8 discrete trials were (presented in this order and as John vs. Amanda on the left and right, respectively): 3/5 vs. 4/9 Balloons, 3/5 vs. 5/12 Paintings, 4/9 vs. 3/5 Buttons, 3/4 vs. 4/10 Drums, 4/9 vs. 2/3 Drums, 4/10 vs. 3/4 Paintings, 4/9 vs. 3/5 Buttons[2], and 5/12 vs. 3/5 Balloons. The 8 continuous trials were matched to the same proportional magnitudes as the discrete trials, but presented without specific discrete numerical information and instead just continuous area (presented in this order and as John vs. Amanda on the left and right, respectively): 0.96 / 1.6 inches vs. 1.28 / 2.88 inches "Sand" (width = 1.4 inches); 1.2 / 2 inches vs. 2 / 4.8 inches "Chocolate" (width = 1.28 inches); 2.03 / 4.57 inches vs. 1.52 / 2.54 inches "Grass" (width =

[2] This trial was accidentally repeated in the discrete block (rather than 2/3 vs. 4/9), and so differs between the discrete and continuous blocks. However, the pattern of findings is identical if this trial is removed.

1.41 inches); 1.2 / 1.6 inches vs. 1.6 / 4 inches “Chocolate” (width = 1.28 inches); 1.28 / 2.88 inches vs. 0.64 / 0.96 inches “Lemonade” (width = 1.39 inches); 1.15 / 2.88 inches vs. 0.86 / 1.15 inches “Lemonade” (width = 1.4 inches); 0.5 / 0.75 inches vs. 1 / 2.25 inches “Sand” width = 1.4 inches); 1.65 / 3.96 inches vs. 0.99 / 1.65 inches “Grass” (width = 1.41 inches). John was always on the left and Amanda was always on the right to make it easier for children to keep track of the stories. However, the proportional answer was on the left (i.e., John) for 3 out of 8 trials in the discrete block (because of the accidental repeat trial, see footnote 2) and on 4 out of 8 trials in the continuous block. Children were scored on the proportion of trials on which they selected the proportional response, within the discrete block and the continuous block separately.

2.1.3 Transparency, Sample Size, and Data Analysis

The study was pre-registered on the Open Science Framework (OSF; https://osf.io/s25fd). All materials, deidentified raw data (Hurst & Levine, 2022), and analysis code is also available on the OSF at https://osf.io/vhj5q/.

Our sample size was chosen a priori based on power analyses and simulations reported in Brysbaert (2019). Although our primary research question is about the within-subject comparison between Discrete and Continuous trials, prior work suggests there may be an interaction or main effect involving the order in which children completed the blocks (e.g., Boyer & Levine, 2015; Hurst & Cordes, 2018). Thus, we used a sample size to allow for 80% power to detect an interaction effect of approximately d = 0.6 and test our primary hypothesis with a between-subject comparison on the first block of trials (d ~ 0.6, power ~ 80%), if needed. If there is not a significant effect of the counterbalanced block order, this design also provides substantial power (> 90%) to detect a within-subject effect of d = 0.4.

Data was analyzed using R 4.0.2 (R Core Team, 2020) in RStudio (R Studio Team, 2016). Data organization and wrangling was done with *tidyverse* (Wickham, 2017). Statistical analyses were computed using base R and *rstatix* 0.6.0 (Kassambara, 2020) for ANOVA, *t*-tests, and correlations, as well as *cocor* 1.1.3 (Diedenhofen & Musch, 2015) for comparing correlations. Data visualizations were created using *ggplot2* 3.3.2 (Wickham, 2016).

2.2 Results

Following our pre-registered analysis plan, we used an ANOVA with Block (2: Continuous, Discrete) as a repeated measure and Order (2: Continuous First, Discrete First) as a between-subject factor (see Figure 2). The ANOVA revealed a significant main effect of Block, $F(1,118) = 33.30$, $p < .001$, $\eta^2_{\text{partial}} = 0.22$, with children making more proportion-based responses on the Continuous trials, $M = 0.54$, $SD = 0.40$, than on the Discrete trials, $M = 0.33$, $SD = 0.39$. There was not a significant effect of Order, $F(1,118) = 3.66$, $p = .058$, $\eta^2_{\text{partial}} = 0.03$, nor an Order x Block interaction, $F(1,118) = 2.77$, $p = .099$, $\eta^2_{\text{partial}} = 0.02$. As described in our pre-registration plan, we were concerned there might be order effects and planned on doing the more conservative between-subject comparison on only the first block of trials if either the main effect or interaction involving order were significant. Although not significant, the main effect of order is only just outside the threshold (0.058 > 0.05), and so to test the robustness of the effect we are reporting the more conservative between-subject test as well. This analysis was consistent with the within-subject analyses, suggesting that children again made more proportion-based responses on the first Continuous block, $M = 0.57$, than the first Discrete block, $M = 0.24$, $t(118) = 4.93$, $p < .001$, $d = 0.90$.

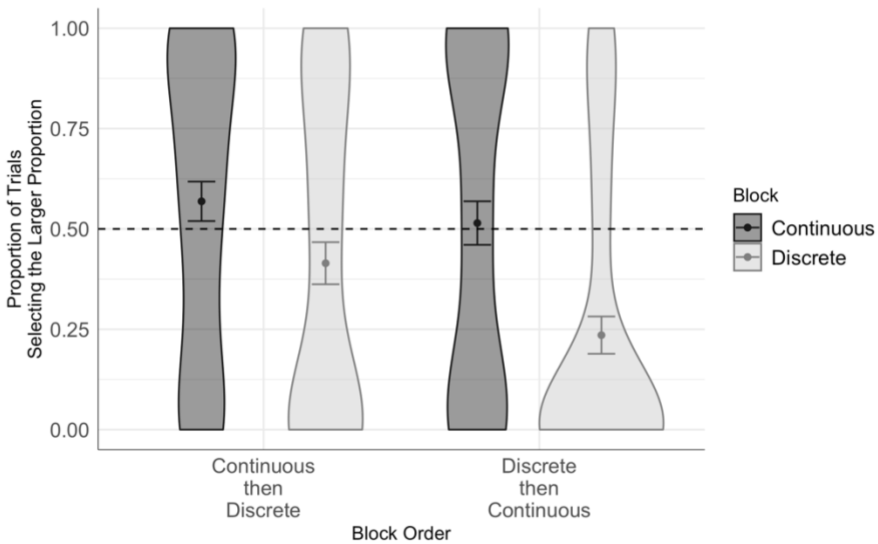

Figure 2: In Experiment 1, the proportion of trials children selected the larger proportion, for the continuous (dark grey; left) and discrete (light gray; right) trials, separated based on the order in which children received the blocks (x-axis). Points represent means, error bars are standard error of the mean, and violin plots display a smoothed kernel density plot of the underlying distribution.

2.2.1 Exploratory Age Differences

As additional exploratory analyses, we were interested in whether there were age differences in the pattern of performance (see descriptive statistics in Table 1). When age group (4-, 5-, or 6-year-olds) is included in the ANOVA described above, there is a significant Age x Block interaction (p = .042, $\eta^2_{partial}$ = 0.05). Thus, we analyzed each age group separately.

Table 1: Mean (Standard Deviation) proportion of trials selecting the larger proportion, separated by Age Group, Block Type, and Order of Blocks

	Continuous Block First		Discrete Block First	
	Continuous Trials	Discrete Trials	Continuous Trials	Discrete Trials
4-year-olds	.49 (.40)	.38 (.39)	.49 (.41)	.43 (.42)
5-year-olds	.58 (.36)	.35 (.40)	.44 (.42)	.11 (.27)
6-year-olds	.64 (39)	.51 (43)	.61 (.43)	.17 (30)

Analyses on 4-year-old children's performance did not reveal any significant effects: Block, $F(1,38) = 3.7$, $p = .064$, $\eta^2_{partial} = 0.09$, Order, $F(1,38) = 0.06$, $p = .816$, $\eta^2_{partial} = 0.001$, or Block x Order, $F(1,38) = 0.25$, $p = .623$, $\eta^2_{partial} = 0.006$. In contrast, both the 5- and 6-year-olds show a different pattern. For 5-year-olds, there was a significant main effect of Block, $F(1,38) = 15.72$, $p < .001$, $\eta^2_{partial} = 0.29$, and Order, $F(1,38) = 4.18$, $p = .048$, $\eta^2_{partial} = 0.10$, but not a significant interaction, $F(1,38) = 0.63$, $p = .43$, $\eta^2_{partial} = 0.02$. Specifically, five-year-old children made more proportion-based responses on the Continuous trials than the Discrete trials and children who completed the Continuous block first made more proportion-based responses than those who completed the Discrete block first. For 6-year-olds, there was a significant main effect of Block, $F(1,38) = 15.99$, $p < .001$, $\eta^2_{partial} = 0.30$, and a significant Block x Order interaction, $F(1,38) = 4.64$, $p = .038$, $\eta^2_{partial} = 0.11$, but not a main effect of Order, $F(1,38) = 3.53$, $p = .068$, $\eta^2_{partial} = 0.09$. Follow up analyses suggest that the 6-year-olds performed similarly on the Continuous trials, regardless of if it was their first block or their second block, $t(38) = 0.29$, $p = .774$, $d = 0.09$. However, they made significantly more proportion-based responses on the Discrete block when it followed the Continuous block than when it was the first block, $t(38) = 2.91$, $p = .006$, $d = 0.92$.

In summary, and as can be seen in Figure 3, the difference in performance between the continuous and discrete trials tends to increase across the three age groups. To provide further insight into this pattern, we also looked at children's performance relative to chance. Four-year-olds did not score significantly different from chance on either trial type, *p*s > .10. Five-year-olds scored significantly below chance on the discrete trials, $p < .001$, but not different from chance on the continuous trials, $p = .88$. Six-year-olds scored significantly below chance on the discrete trials, p = .018, and slightly, though not significantly, above chance on the continuous trials, p = .058.

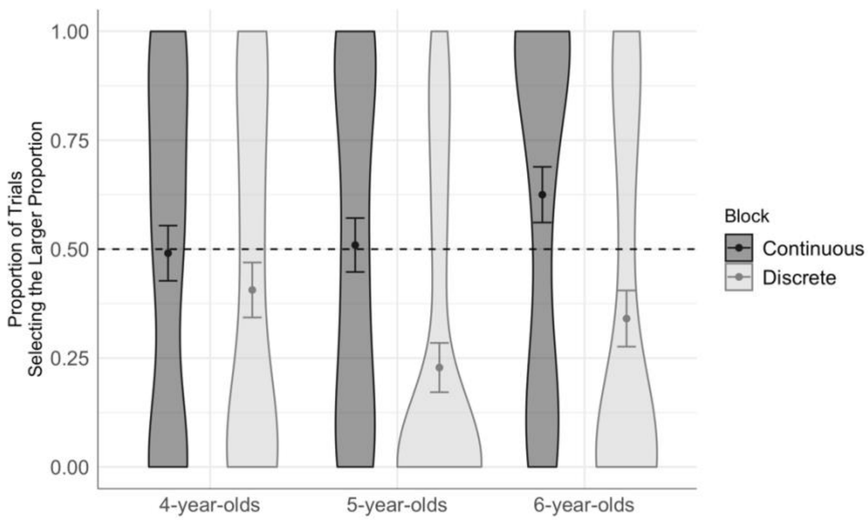

Figure 3: Proportion of trials on which children selected the larger proportion on the continuous (dark grey; left) and discrete (light gray; right) blocks, separated by age group (x-axis). Points represent means, error bars are standard error of the mean, and violin plots display a smoothed kernel density plot of the underlying distribution.

2.3 Discussion

Data from Experiment 1 replicated and extended Sullivan et al., (2018): children's absolute *more*-based interpretation of *most* is exacerbated in discrete contexts where numerical information is available, relative to continuous contexts. Furthermore, exploratory analyses across the age groups in our sample suggest that children's first interpretation of *most* as absolutely more may be exclusively applied to discrete contexts where absolutely more is based on greater number; strikingly, none of the age groups in our sample reliably used an absolute area comparison in their interpretation of *most* on the continuous trials. Moreover, a proportional interpretation of *most,* even in the continuous contexts where proportional reasoning is typically facilitated (e.g., Boyer et al., 2008; Hurst & Cordes, 2018), was not fully evident in our sample, although there was some evidence that it might be emerging around 6 years old. Together, these results indicate that there is numerical interference in the interpretation of *most* in the context of discrete quantities, but also that children have difficulty making a truly proportional interpretation of *most* even in the context of continuous quantities.

In Experiment 2, we aim to conceptually replicate Experiment 1, but we modified the study in four ways, with the goal of improving the experimental stimuli and allowing us to address additional questions. First, it's possible that the increased attention to numerical information with discrete stimuli was not caused by the visual availability of numerical information, but rather by using number words during the description of the stimuli. In other words, the verbal number words might have highlighted a comparison of the absolute cardinalities. To address this possibility, in Experiment 2 we did not use number words to introduce the stimuli and instead rely on generic quantifiers (i.e., this much and this many).

Second, we address the possibility that chance performance with continuous stimuli was due to children's difficulty understanding the continuous version of the task. The continuous stimuli were more abstract than the discrete stimuli (e.g., colored rectangles representing a glass of lemonade versus images of drums representing drums) and the visual depiction of the amount being discussed (e.g., the amount sold) was represented as a partitioned section of the rectangle with an "X" in it. Although the "X" was also used on individual objects in the discrete stimuli, the subtle partition and "X" within a single object may have been more difficult to understand than an "X" to cross out a whole object. Thus, in Experiment 2, we used stimuli that provide a more natural part-whole interpretation and can be represented using actual images of the stimuli that correspond to the sentence provided rather than abstract shapes (e.g., a butterfly that's partially colored, an image of an actual glass of water that's partially filled).

Third, we address the possibility that the difference in performance between discrete and continuous trials in Experiment 1 was due to more general differences in how continuous and discrete contexts are interpreted, rather than due to numerical interference per se. That is, given that children's performance remained around chance on the continuous trials, it may be that they were not able to interpret the word *most* in this context at all. To investigate this possibility, we measured children's interpretation of *most* using a comparisons task as well as a non-comparison verification task on which children can rely on absolute counts or amounts to determine whether a stimulus is consistent with *most.* If children's difficulty with *most* in discrete contexts is specifically due to numerical interference, then we should not see a relative benefit of continuous vs. discrete stimuli on the verification task because numerical interference is eliminated. On the other hand, if the pattern of results from Experiment 1 is attributable to children's specific difficulty with *most* in discrete contexts *and* general confusion around the term *most* in continuous contexts, then we would expect a difference between discrete and continuous stimuli in both a verification task that does not involve numerical interference (i.e., is this most?) and a comparison context that does (i.e., which is most?).

Fourth, and finally, the exploratory age analyses suggest that the older children in the sample might have had a more proportional understanding of *most,* demonstrating the largest effects and the most interesting developmental pattern. Thus, in Experiment 2 we focus on just 5- and 6-year-olds (and not 4-year-olds), allowing us to investigate the emergence of a proportional interpretation of *most* and condition differences with increased power.

3. Experiment 2

3.1 Method

3.1.1 Participants

As pre-registered, our final sample is 120 5- and 6-year-old children, M_{age} = 71 months, randomly assigned to one of the two block orders. Sixty children completed the continuous block of trials first: 30 5-year-olds, M_{age} = 65 months, Range: 60 to 71 months, 8 girls, 16 boys, 6 undisclosed gender; and 30 6-year-olds, M_{age} = 77 months, Range: 72 to 83 months, 12 girls, 13 boys, 5 undisclosed gender. Sixty children completed the discrete block of trials first: 30 5-year-olds, M_{age} = 66 months, Range: 60 to 71 months, 14 girls, 11 boys, 5 undisclosed gender; and 30 6-year-olds, M_{age} = 77 months, Range: 73 to 83 months, 11 girls, 14 boys, 5 undisclosed gender. An additional four children completed at least some of the task but are not included in the analyses because of technology issues (n = 2), they did not complete the task (n = 1), or because they were out of the target age range (n = 1).

Children were recruited from our lab database, social media, and Children Helping Science (childrenhelpingscience.com) and tested online via video-chat. Families were compensated with $5. All procedures were approved by the University of Chicago Institutional Review Board and parents provided informed consent.

Approximately 88% of the sample completed at least some of an additional demographic survey. Based on this subsample, 17% reported being Asian, 9% Black or African American, 65% White, 9% more than one race, and < 1% reported Other that was not one of the previous categories. In addition, about 17% of children were identified as Hispanic or Latino/a/x. Most of the sample that reported demographics came from high-income homes, with 54% making more than $100,000 per year, 33% reporting $50 - $99K per year, and 13% reporting less than $50K per year. Lastly, 92% of the parents who completed the demographic form (81% of whom were mothers) had at least a 4-year bachelor's degree or equivalent.

3.1.2 Procedure and Stimuli

All children completed a Verification Task and a Comparison Task (similar to Experiment 1) with both discrete and continuous stimuli. The discrete and continuous stimuli were presented in two separate blocks with the order counterbalanced across participants. Within each block, the Verification Task always preceded the Comparison Task.

Children participated online over video chat (via Zoom) and both tasks were programmed and administered in PsycoPy3 v3.2.4 (the task was also sometimes administered using PsychoPy2020 v2020.1.2) run on the experimenter's computer and shared via screen sharing. Prior to beginning the task, children were shown a screen with an image in each corner and asked to describe what they saw. This initial task was included as a warmup, given children's different levels of comfort with video chat, and to help troubleshoot any issues with screen visibility.

3.1.2.1 Verification Task.

The verification task included 24 trials in which children were shown an image that showed a set of items (Discrete Block) or a single item displaying a continuous amount (Continuous Block) and asked whether *most* of the set or amount met the condition (see Figure 4A). On the Discrete block, each image was a set of items, some of which fulfilled a property and some of which did not: full vs. empty glasses of water, purple vs. white butterflies, full vs. empty jars of sand, full vs. empty bags of cereal, and full vs. empty boxes of pizza. Children were then asked questions of the form "Are most of the [noun verb phrase]?", such as "Are most of the butterflies colored in?". On the Continuous block, each image was a matched single item that varied in the amount presented: glass of water, colored butterfly, jars of sand, bags of cereal, and pizza boxes. Children were asked questions of the form, "Is most of the [noun verb phrase]", such as "Is most of the butterfly colored in?". Children's verbal yes or no response was recorded by the experimenter using the keyboard (y or n key respectively).

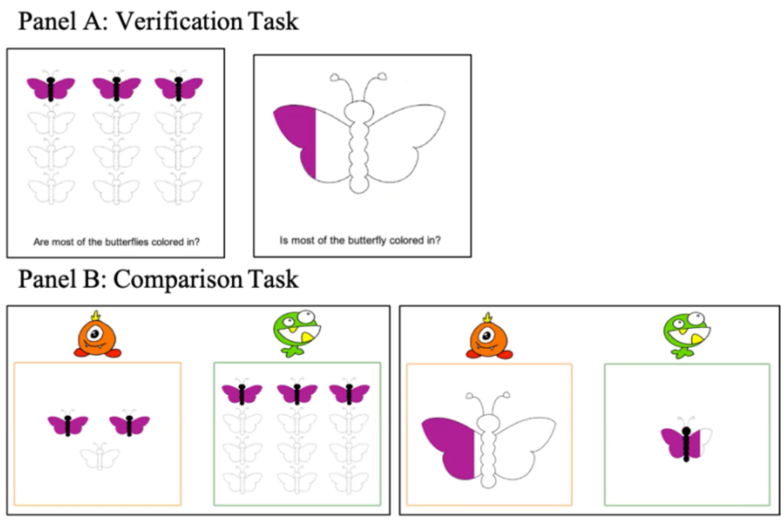

Figure 4: Example stimuli from the Verification Task (Panel A) and Comparison Task (Panel B) of Experiment 2, with discrete stimuli on the left and continuous stimuli on the right.

Trials included depictions of values from 0% to 100%, presented in a random order: 0% (2 trials), 25% (5 trials), 40% (5 trials), 60% (5 trials), 75% (5 trials), and 100% (2 trials). On discrete trials, the total set size ranged from 3 to 12 and the relevant subset (e.g., full glasses of water, purple butterflies) ranged from 1 to 6. These ratios were chosen to allow us to compare children's judgements for values that are less than half and values more than half. Empty (0%) and full (100%) were included to provide some sense of children's interpretations at the extremes, but given that these trials were of less interest to our current research question, we included fewer of these trials.

3.1.2.2 Comparison Task

The comparison task included 14 trials in which children were introduced to an orange monster and a green monster, each of whom had a set or amount that depicted a given value (like those used in the verification task) and asked which option satisfied the requirement involving the quantifier *most* (see Figure 4B). On the Discrete Block, the sets were of the same structure used in the verification task and shown one-at-a-time for the orange and green monster, while the experimenter said: “The [orange/green] monster [verb] this many [noun phrase]” (e.g., “the orange monster filled this many glasses with water”). After introducing both monsters and both sets, children were then asked “who [verb] most of their [noun phrase]” (e.g., “who filled most of their glasses with water?”). The same basic structure was used on the Continuous Block using the same amounts as the verification task, but the amounts were introduced using the phrase “the [orange/green] monster [verb] [noun phrase involving “this much”]” (e.g., “the orange monster filled their glass with this much water”) and children were asked which is *most* using the phrase “who [verb] most of their [noun phrase]” (e.g., “who filled most of their glass with water?”). Children’s verbal response (orange or green) was recorded by the experimenter using the keyboard (left or right arrow key). If children pointed, they were asked whether they were pointing to the orange monster (always presented on the left) or the green monster (always presented on the right).

On both blocks, 10 (out of 14) trials included the same competition as Experiment 1, such that one option had more absolute amount than the other option (in terms of number on the discrete block and area on the continuous block) but was a value less than half, while the other option always had a lower absolute amount (number or area) but was a value more than half. The remaining four trials compared two options that had the same total set size/amount (e.g., 3/4 vs. 1/4), meaning that the option with absolutely more also had proportionally more. The specific proportional values used were matched across blocks and the trials were presented in a random order (within block). The specific stimuli included a similar range of values to those used in the Verification task, with the less than half proportions ranging from 25% to 44% and the more than half proportions ranging from 60% to 75%.

3.1.3 Transparency, Sample Size, and Data Analysis

The study was again pre-registered on the OSF (https://osf.io/52fjr). Sample size justification, data analysis, and open materials, data, and analysis are as described for Experiment 1.

3.2 Results

3.2.1 Verification Task

Trials depicting 0% and 100% were included to ensure children were asked about the full range of values and to provide some descriptive sense of children's performance at these boundaries. In general, regardless of the order in which children saw the blocks or whether the stimuli were discrete or continuous, children overwhelming said that 0% values were not *most* (*Ms* ≤ 5% of trials with "yes" response) and that 100% values were *most* (*M*s > 85% of trials with "yes" response), in line with prior work with similarly aged children (Papafragou & Schwarz, 2006). For our primary analyses, however, we focused on values between these extremes, excluding the endpoints (although, the pattern of findings is identical when these values are included). Children's acceptance of *most* across each of the values tested (0%, 25%, 40%, 60%, 75%, and 100%) are shown in Figure 5.

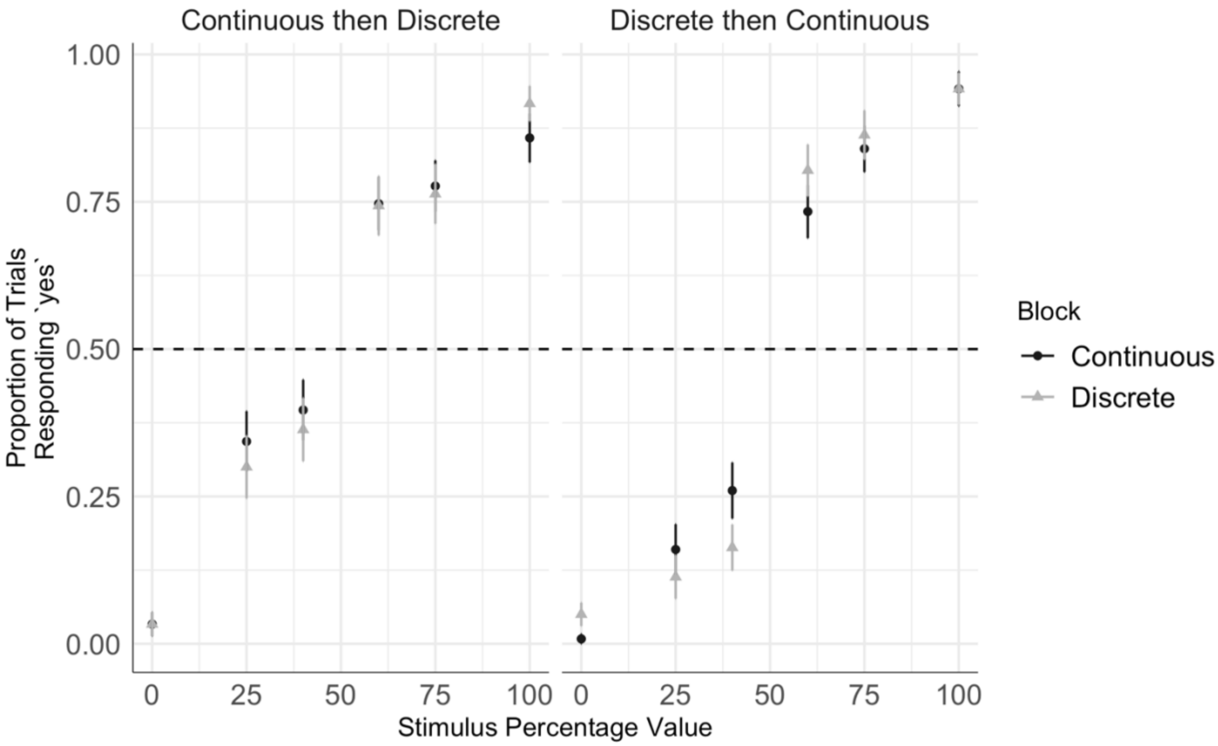

Figure 5: In the Verification task of Experiment 2, the proportion of trials on which children responded that "yes" the depicted value was *most*, for the continuous (dark grey) and discrete (light gray) trials, separated based on the percentage value depicted by the stimulus (x-axis) and the order in which the child saw the blocks (left vs. right plots). Points represent means and error bars are standard error of the mean.

As pre-registered, we analyzed the proportion of trials on which children said "yes" using a Percent Category (2: less than half, more than half) x Block Type (2: Continuous, Discrete) x Order (2: Discrete first, Continuous first) ANOVA. This analysis revealed a main effect of Percent Category, $F(1, 118) = 228.70$, $p < .001$, $\eta^2_{partial} = 0.66$, and that Percent Category interacted with both Block Type, $F(1, 118) = 6.22$, $p = .014$, $\eta^2_{partial} = 0.05$, and Order, $F(1, 118) = 11.05$, $p = .001$, $\eta^2_{partial} = 0.09$, with no other significant effects, *p*s > .100. Given the interaction between Percent Category and Order, and in line with our pre-registered analysis plan, we re-analyzed the data using the more conservative between-subject comparison of the first block of trials only to test our primary hypothesis (although the pattern of results described below is also found when analyzed within subject across the first and second block).

The Percent Category (2, within-subject: less than half, more than half) x First Block Type (2, between-subject: Continuous, Discrete) ANOVA confirmed a main effect of Percent Category, $F(1, 118) = 233.87$, $p < .001$, $\eta^2_{\text{partial}} = 0.67$, and a Percent Category x First Block Type interaction, $F(1, 118) = 18.22$, $p < .001$, $\eta^2_{\text{partial}} = 0.13$, but not a main effect of First Block Type, $F(1, 118) = 2.65$, $p = .106$, $\eta^2_{\text{partial}} = 0.02$. Children correctly responded "yes" at above chance levels when the value was above 50% on both the Discrete, $M = 0.83$, SD = 0.32, $t(59) = 8.00$, $p < .001$, and Continuous trials, $M = 0.76$, SD = 0.32, $t(59) = 6.31$, $p < .001$, which were not significantly different from each other, $t(118) = -1.22$, $p = .23$. In contrast, children responded "yes" at below chance levels (i.e., tended to correctly respond "no") when the value was below 50% on both the Discrete, $M = 0.14$, SD = 0.29, $t(59) = -9.80$, $p < .001$, and Continuous trials, $M = 0.37$, SD = 0.39, $t(59) = -2.58$, $p = .012$. In this case, there was also a significant difference between the continuous and discrete trials, with children incorrectly responding "yes" on values smaller than 50% more frequently when the values were presented continuously than when presented discretely, $t(108.22) = 3.71$, $p < .001$ (not assuming equal variances because of a significant difference in variances).

3.2.1.2 Exploratory Age Effects

When age group is included in the ANOVA on the first block, there were not significant main or interaction effects involving age (*p*s > .05). Moreover, 6-year-olds correctly responded "yes" to values above 50% and correctly responded "no" to values below 50% at above chance levels on both continuous and discrete trials: proportion of trials saying "yes" for trials above 50%: $M_{\text{cont}} = .75$, $M_{\text{disc}} = .90$, and below 50%: $M_{\text{cont}} = .32$, $M_{\text{disc}} = .11$, all comparisons to chance *p*s < .02. Five-year-olds were similarly successful on most trial types, proportion of trials saying "yes" for trials above 50%: M_{cont} = .77, M_{disc} = .76, and below 50%: M_{disc} = .16, comparison to chance *ps* < .001, except for values below 50% presented with continuous stimuli, where 5-year-olds were not significantly different from chance, *M* = .42, *p* = .304. Thus, although there are not significant age differences (potentially because of the relatively low power to detect them), 5-year-olds may less consistently reason about *most* than 6-year-olds.

3.2.2. Comparison Task

Children performed very well on the trials that had an equal total amount, which did not pit proportion and absolute amount against each other, with both Continuous, *M* = 0.88, and Discrete *M* = 0.92, stimuli, suggesting they understood the task. Our primary analyses, however, concern the trials in which proportional amount and absolute amount were in direct conflict with each other, as was the case in Experiment 1 and prior work. We used a Block Type (2: Discrete, Continuous) x Block Order (2: Discrete First, Continuous First) ANOVA on the proportion of trials children selected the option displaying more than half (i.e., the proportional interpretation of *most*). There was a significant main effect of Block Type, $F(1, 118) = 98.90$, $p < .001$, $\eta^2_{\text{partial}} = 0.24$, and a Block x Order interaction, $F(1, 118) = 6.18$, $p = .014$, $\eta^2_{\text{partial}} = 0.02$, but not a significant main effect of Order, $F(1, 118) = 0.86$, $p = .356$, $\eta^2_{\text{partial}} = 0.005$.

Again, given the significant interaction involving block order, and as specified in our pre-registered plan, we tested our primary hypothesis using a more conservative between-subject analysis. A between-subject *t*-test on the first block children completed also revealed a significant difference between Continuous, *M* = .62, SD = .25, and Discrete, *M* = .21, SD = .36, stimuli, $t(106.2) = 7.11$, $p < .001$, $d = 1.30$ (not assuming equal variances because of a significant test of differences between variances). Moreover, children scored significantly above chance on the Continuous trials, $t(59) = 3.51$, $p < .001$, and significantly below chance on the Discrete trials, $t(59) = -6.23$, $p < .001$ (note that this same pattern is found when collapsing across the first and second test block). Thus, even with a between-subject test, we see a significant difference in proportional reasoning between discrete and continuous stimuli, with children more often successfully selecting the proportion-based response on continuous trials and incorrectly selecting the absolute-based more response (which was proportionally less) on the discrete trials. Further, the interaction between Block and Order suggests that this difference, though present for children who received either order of trials, is reduced in children who completed the continuous block first, relative to children who completed the discrete block first (see Figure 6). However, neither simple effect was independently significant: children's performance did not significantly differ on the continuous block when it occurred before, $M = 0.62$, or after $M = 0.66$, the discrete block, $t(118) = -0.95$, $p = .342$, $d = -0.17$ and similarly, children's performance on the discrete block before, $M = 0.21$, or after, $M = 0.35$, the continuous block did not significantly differ, $t(118) = 1.97$, $p = .052$, $d = 0.36$ (although notably, this small effect is just beyond the significance threshold).

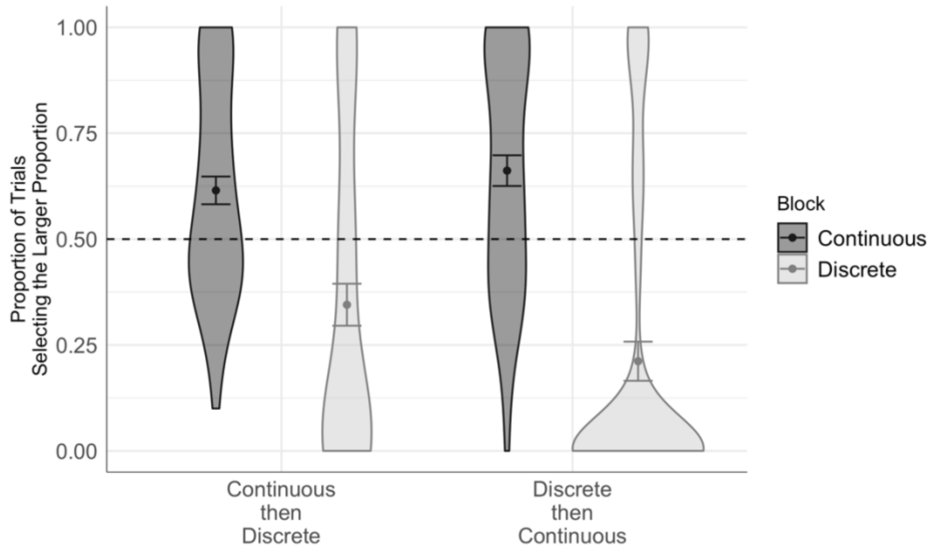

Figure 6: In the Comparison task of Experiment 2, the proportion of trials children selected the larger proportion, for the continuous (dark grey; left) and discrete (light gray; right) trials, separated based on the order in which children received the blocks (x-axis). Points represent means, error bars are standard error of the mean, and violin plots display a smoothed kernel density plot of the underlying distribution.

3.2.2.2 Exploratory Age Analyses

When age group is included in the analyses of the first block, there was not a significant main effect of age, p = .724, $\eta^2_{partial} = 0.001$, or age by block type interaction, p = .114, $\eta^2_{partial}$ = 0.021. However, although 6-year-olds performed significantly above chance on continuous trials, M = .67, and significantly below chance on discrete trials, M = .18, ps < .001, 5-year-olds' performance did not significantly differ from chance on continuous trials, M = .56, p = .211, and was significantly below chance on discrete trials, M = .25, p < .001. Thus, although there are not significant age differences (again, potentially because of the relatively low power to detect them), interpretating *most* based on proportion on continuous trials may not be as robust among 5-year-olds as among 6-year-olds.

3.2.3 Direct Comparison Across Tasks

To directly compare children's performance on the verification task to children's performance on the comparison task, we scored children on both tasks based on their interpretation of *most* as "more than half". On the verification task, children were scored correct is they said *yes* to values above 50% and *no* to values below 50%, again excluding 0% and 100% trials. On the comparison task, children were scored based on the proportion of trials selecting the proportion-based response (i.e., the larger proportion) on trials that pit proportional and absolute interpretations against each other. Given the significant interactions with order when analyzing each task individually, we used the more conservative between-subject analysis on the first block of trials only using a Task (2: Verification, Comparison) x First Block Type (2: Discrete, Continuous) ANOVA on the proportion correct, where correctness is defined as responding in terms of "more than half" (see Figure 7). This analysis revealed a significant main effect of Task, $F(1, 118) = 131.79$, $p < .001$, $\eta^2_{partial} = 0.53$, main effect of First Block Type, $F(1, 118) = 12.53$, $p < .001$, $\eta^2_{partial} = 0.10$, and an interaction of Task and First Block Type, $F(1, 118) = 79.0$, $p < .001$, $\eta^2_{partial} = 0.40$.

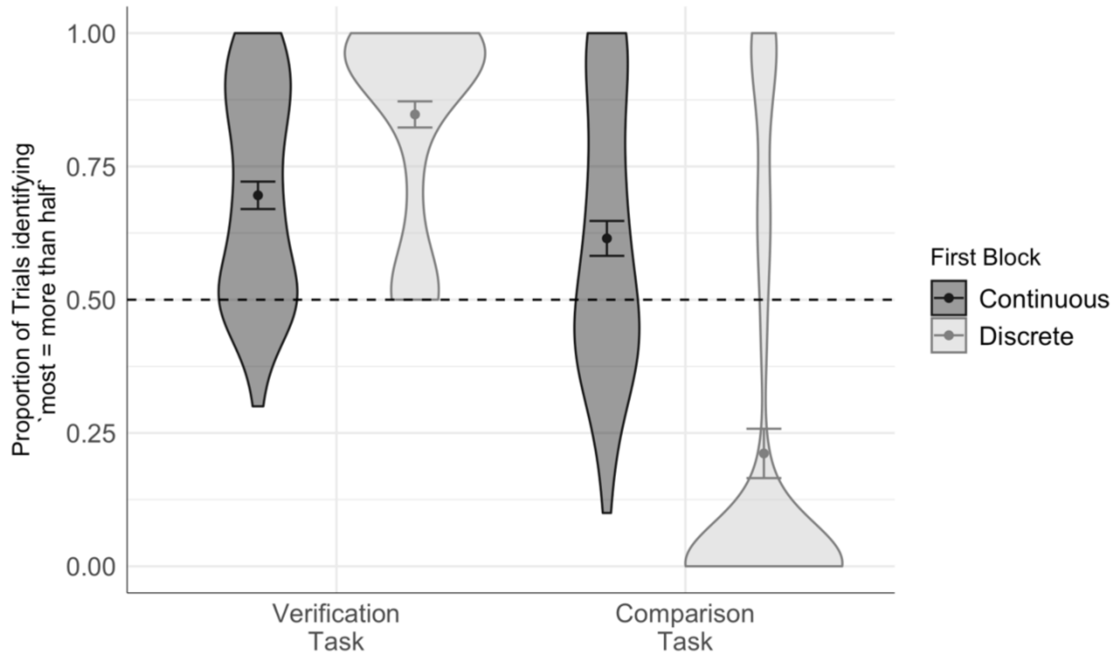

Figure 7: In Experiment 2, the proportion of trials children responded such that "most" is "more than half" for the continuous (dark grey; left) and discrete (light gray; right) trials, across the Comparison (left) and Verification (right) tasks, on the first block of trials only. Points represent means, error bars are standard error of the mean, and violin plots display a smoothed kernel density plot of the underlying distribution.

Follow up analyses indicate that children performed better on the verification task than the comparison task for both Continuous, $M_{\text{verification}} = 0.70$, SD = 0.20, $M_{\text{comparison}} = 0.62$, SD = 0.25, $t(59) = 2.33$, $p = .023$, $d = 0.30$, and Discrete blocks, $M_{\text{verification}} = 0.85$, SD = 0.19, $M_{\text{comparison}} = 0.21$, SD = 0.36, $t(59) = 12.24$, $p < .001$, $d = 1.58$, with this difference being notably larger on discrete trials. Moreover, on the verification task, children had higher scores with Discrete, compared to Continuous, stimuli, $t(59) = -4.18$, $p < .001$, $d = -0.54$, but on the comparison task, in contrast, children had higher scores with Continuous stimuli, compared to Discrete stimuli, $t(59) = 6.83$, $p < .001$, $d = 0.88$.

Although not pre-registered, we conducted an additional exploratory analysis to examine whether children's performance on the verification and comparison tasks was related, perhaps more so for continuous than discrete trials because of the use of different strategies across tasks for the latter. Thus, we analyzed the correlations between children's performance on the verification and comparison task for each stimulus type, again focused only on the more conservative between-subject analysis of the first block of trials. For children who received the Continuous Block of trials first, there was a significant correlation between performance on the verification and comparison tasks with continuous stimuli, $r = .32$ 95% CI [.07, .53], $p = .014$. In contrast, for children who received the Discrete Block of trials first, there was not a significant correlation between tasks with discrete stimuli, $r = .02$ 95% CI [-.23, .27], $p = .877$. However, these correlations are not significantly different from each other $z = 1.66$, $p = .096$.

3.3 Discussion

Experiment 2 replicated our findings from Experiment 1. However, with this new and more naturalistic set of stimuli and a larger sample of older children, not only were children more susceptible to an absolute *more*-based interpretation of *most* with discrete stimuli, but children also reasoned proportionally about *most* at above chance levels when amounts were presented continuously. Together with the age-related pattern in Experiment 1, this suggests that a proportional interpretation of *most* is evident at around 6-years-old when stimuli are continuous amounts.

In addition, and importantly, children's interpretation of *most* in a non-comparison context, which did not involve competition between absolute and proportional amounts, was more nuanced. Overall, in this context, children showed the expected pattern of "yes" responding for values more than 50% and "no" responding for values less than 50%. In addition, as hypothesized, children's performance did not benefit from the use of continuous stimuli. In fact, for values less than 50%, children were better able to reject the *most* phrasing for discrete stimuli than for continuous stimuli. Together, these findings suggest that the presence of numerical information impedes reasoning about *most* when a proportional interpretation is required but facilitates reasoning about *most* when an absolute interpretation is sufficient, potentially because continuous quantities require approximation whereas discrete quantities provide the opportunity to make exact judgements based on counts.

Lastly, the significant interaction between performance and block order suggests that how children draw upon their conceptualizations of proportion might be malleable: increasing the saliency of proportion and decreasing the saliency of absolute number through an initial block of continuous trials impacts children's strategy use on the subsequent discrete trials. Moreover, this same pattern has been shown in proportional reasoning tasks involving equivalent juice mixtures and probabilistic comparisons with game spinners (Boyer & Levine, 2010; Hurst & Cordes, 2018), further emphasizing that children may be drawing upon their conceptualization of proportion in similar ways across contexts, including the current context where children are asked questions about *most*.

4. General Discussion

In the current study, we report two experiments investigating children's interpretation of the quantifier *most*. Overall, we find that by age 6, children's difficulty with a proportional interpretation of *most* is likely caused by numerical interference when discrete and countable numerical information is available. In contrast, when discrete numerical information was not available, a proportional interpretation of *most* was evident by 6 years old. Furthermore, although the presence of discrete numerical information hindered a proportional interpretation of *most*, when absolute numerical information was sufficient for interpreting *most* correctly (i.e., when an absolute comparison was consistent with a proportional comparison), discrete numerical information actually facilitated children's interpretation of *most* as not applying to values less than half. In other words, it is not the case that continuous stimuli, and the absence of discrete number, always facilitates children's reasoning about *most*. When an absolute interpretation is sufficient (as in the verification task of Experiment 2) reasoning about *most* in the context of continuous areas was more error prone than in the context of discrete quantities. Moreover, in a comparison context, younger children (i.e., 5-year-olds) struggled with the interpretation of *most* in reference to continuous stimuli, performing around chance, despite the emergence of systematic (albeit, incorrect) reasoning about *most* in reference to discrete stimuli. Lastly, exploratory correlational analyses in Experiment 2 suggest that when presented with continuous stimuli, children might be drawing on the same proportional reasoning strategies regardless of whether they are required to compare across proportions or within a single proportion. In contrast, with discrete stimuli children's reasoning about *most* might be more dependent on task demands and task-specific heuristics. Together, the pattern of results across both experiments suggests that children's initial understanding of *most* is fragmented and highly context-dependent in ways that reflect their underlying knowledge of different types of quantity.

4.1 Implications

First, the current experiments highlight the importance of considering how children reason in the context of different perceptual stimuli and different types of questions to understand children's knowledge and develop theories of cognitive development. In the context of assessing children's quantifier knowledge, Sullivan and colleagues (2018) convincingly make the case for considering task demands across different paradigms (e.g., production vs. verification vs. comparison). We extend their argument to different stimuli, even within an identical paradigm – notably varying whether the stimuli consist of continuous amounts or discrete sets. Our findings, together with the growing literature on proportional reasoning, suggest that the distinction between continuous and discrete proportional amounts must be considered when studying the development of children's reasoning about quantifiers such as *most* and *more*. However, the consideration of this distinction is not unique to children's understanding of quantifiers, but instead is relevant to a range of contexts, including reasoning about mixtures (Boyer et al., 2008), probability (Hurst & Cordes, 2018), and resource distribution scenarios in social cognition (Hurst et al., 2020).

Second, these implications are not merely methodological (to the extent that methodological concerns are ever mere), but rather have theoretical implications for our understanding of children's quantifier knowledge and proportional reasoning, two important domains of cognitive development. In the case of children's understanding of the quantifier *most*, the current study provides insight into two related, but somewhat distinct, questions: whether, and when, children interpret *most* in terms of proportion and whether children distinguish between *most* and *more.*

4.1.1 Interpreting *Most* in Terms of Proportion

Overall, we find that children can apply a proportional interpretation of the quantifier *most* by around 6-years-old, but that the proportional interpretation of *most* is only evident, at least at first, with continuous amounts. One possibility for this pattern is that children are using entirely distinct interpretations of *most* across contexts, with different developmental trajectories. That is, children may first develop an absolute comparison interpretation of *most* that applies exclusively to discrete sets and is evident by around 4 years old (although other work has shown it as early as 3 years old; Halberda et al., 2008). In contrast, it is not until around the age of 6 years that children develop a proportional interpretation of *most* that applies (at least at first) exclusively to continuous area-based contexts. An alternative possibility is that children's initial conception of *most* is proportional, even in young children, but that the saliency of discrete numerical information prompts an absolute numerical comparison, interfering with children's ability to demonstrate their proportional knowledge. Investigating even younger children's interpretation of *most* with continuous stimuli, where an absolute interpretation is available but numerical interference is not (i.e., in a comparison task that pits continuous absolute area vs. proportional area), will allow us to better understand the emergence of children's proportional understanding of *most*.

In either case, however, it is clear from the current study that by 6-years-old children can interpret *most* proportionally and demonstrate this proportional interpretation in some contexts. Importantly, future work is needed to better understand the potentially distinct interpretations of *most* for different kinds of quantities, how they develop in tandem, and how children eventually come to integrate these different meanings of *most* into an adult-like interpretation that they can apply across context and stimulus type. Furthermore, although exploratory analyses comparing 5-year-olds and 6-year-olds suggest that it might be around this time that children become able to demonstrate their proportional understanding of *most* in continuous contexts, a more complete investigation of age differences with a larger sample across a wider age range would provide a better understanding of how children's knowledge of *most* develops.

4.1.2 Distinguishing *Most* from *More*

Although the current experiments did not directly address children's understanding of *more*, they contribute to a larger discussion about whether, and how, children distinguish between *most* and *more.* Notably, demonstrating a proportional interpretation of *most* in some contexts is not sufficient to conclude that children distinguish *most* and *more.* In other words, it is possible that children interpret *most* to mean *more* but interpret both *most* and *more* proportionally in some contexts. The feasibility of this hypothesis was tested in a sample of adults (see Experiment S1 in Supplemental), which found that adults did interpret both *more* and *most* proportionally with the continuous stimuli used in Experiment 2. Thus, when the quantities have a clearly defined part-whole structure, even the quantifier *more* might be interpreted proportionally. Although it is unclear whether children would show similar flexibility in their interpretation of *more*, if they do interpret *more* like adults then it may be that children in Experiments 1 and 2 did not show absolute interference in the continuous comparison task because in this context there is not competition between *more* and *most* – both quantifiers lead to the same proportional response.

There are two additional findings from the current study that are worth noting when considering whether children treat *most* and *more* as synonyms. First, although an absolute comparison was all that was necessary to succeed on the verification task in Experiment 2, the language used (e.g., "are most of the butterflies colored in?") is more ambiguous with a "*more*" interpretation than a "*most*" interpretation. In other words, although "most of the" fully implies the reference set (e.g., all the butterflies), "more of the" does not specify an explicit reference set (e.g., "are more of the butterflies colored in?" leaves open the question of "more than what?"). This contrasts with typical verification tasks that highlight both subsets and the question posed is entirely grammatical when substituting *more* for the word *most* (e.g., are most [more] of the crayons <u>blue</u> or <u>yellow</u>? Halberda et al., 2008). Although it is possible that children used a *more* interpretation and spontaneously inferred the correct reference set, requiring them to do so may have made it a more difficult task than when the reference set is specified. Future work with the goal of distinguishing *most* and *more* could investigate whether children readily infer the comparison set for *more,* or whether difficulty with this inference in our verification task provides evidence against the possibility that they are interpreting *most* as synonymous with *more*.

Second, we found that children under 6 years were not systematic in their *most* judgements of continuous amounts, neither responding primarily in terms of absolute amounts nor primarily in terms of proportion. This finding suggests that children's initial interpretation of *most* is more limited than their understanding of *more*. Prior work has found that children younger than those tested here can readily interpret *more* to compare both discrete and continuous amounts (Odic et al., 2013). Thus, if younger children were truly interpreting *most* as a synonym for *more,* then the children in our study should have made consistent judgements of *most* with both discrete and continuous stimuli, which was not the case.

Together, these patterns emphasize the need for research that separately addresses questions about children's proportional interpretation of *most* from questions about whether children treat *more* and *most* synonymously. The current study focused on the first question, concluding that children can interpret *most* proportionally by at least 6-years-old, while leaving open whether this interpretation of *most* is truly distinct from their interpretation of *more.*

4.1.3 Proportional Reasoning More Generally

Finally, in terms of proportional reasoning, numerical interference in discrete proportional reasoning has been extensively studied for both symbolic (e.g., fractions) and non-symbolic proportions (e.g., visual representations) across a range of developmental age groups (e.g., Alonso-Díaz et al., 2018; Alonso-Díaz & Penagos-Londoño, 2021; Boyer et al., 2008; Hurst & Cordes, 2018; Ni & Zhou, 2005). Our findings extend this research in two ways. First, we replicate the typical numerical interference effect in a new domain, namely interpretation of the quantifier *most*. Second, and more importantly, we show that when the possibility of numerical competition is eliminated, as in the verification task used in Experiment 2, discrete information is actually helpful, potentially because it provides a way for children to be more exact about their judgements. By understanding the context-dependence of proportional reasoning across different domains, we can better understand children's successes and failures in proportional reasoning and how they impact the many contexts of human cognition that rely on processing proportional information.

4.2 Conclusion

In summary, the current study reveals that by at least 6-years-old children can understand the semantics of a quantifier that requires proportional reasoning, but children's interpretation of *most* is nuanced and depends on the need to reason proportionally and the format of the quantities involved. When proportional reasoning is not required, as in our verification task, even young children may appear to have mature understanding of *most* and perform better when they can make comparisons based on the number of objects in two sets than when they must compare continuous amounts. In contrast, when proportional reasoning is required, children perform better when stimuli consist of continuous amounts rather than discrete quantities, because the former eliminate numerical interference, a known impediment to proportional reasoning. Thus, children may use context-specific interpretations of *most*: proportional interpretations in continuous contexts, but absolute interpretations in numerical contexts. These differing interpretations either facilitate or interfere with performance, depending on task demands. By using stimuli that include both discrete and continuous amounts, we can broaden our understanding of the development of quantifier knowledge. Furthermore, this is not unique to quantifier knowledge, but also other domains of cognitive development that rely on the ability to make inferences based on proportional information, a kind of reasoning that is ubiquitous in everyday life as well as in specialized domains within Science, Technology, Engineering, and Math (i.e., STEM).

5. Acknowledgements

This research was funded by a generous grant (#2018-0680) from the Heising-Simons Foundation to the Development and Research in Early Math Education (DREME) Network, of which S.C. Levine is a member. The funder was not involved in the conduct of the research or in the preparation of the manuscript. In addition, we would like to thank all our participants and participant partners, including the schools and organizations that participated, the Museum of Science and Industry (Chicago, IL), and ChildrenHelpingScience.com. Finally, we would like to thank two reviewers, Melissa Kibbe and Roman Feiman, for constructive feedback that improved the manuscript. We particularly thank Roman Feiman for noting that proportional responding with *most* does not imply that children differentiate *most* from *more* and for suggesting Experiment S1 (in Supplemental) to test this claim with adults. Declarations of interest: none. CRediT author statement: MAH: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization, Supervision, Project administration; SCL: Supervision, Funding acquisition, Writing – Review & Editing.

6. References

Alonso-Díaz, S., & Penagos-Londoño, G. I. (2021). The numerator bias exists in millions of real-world comparisons. *Acta Psychologica*, *213*, 103248. https://doi.org/10.1016/j.actpsy.2020.103248

Alonso-Díaz, S., Piantadosi, S. T., Hayden, B. Y., & Cantlon, J. F. (2018). Intrinsic whole number bias in humans. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(9), 1472–1481. https://doi.org/10.1037/xhp0000544

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93. https://doi.org/10.1016/j.cognition.2010.10.010

Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development☆. *Cognitive Psychology*, *58*(2), 195–219. https://doi.org/10.1016/j.cogpsych.2008.07.001

Barner, D., Libenson, A., Cheung, P., & Takasaki, M. (2009). Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *Journal of Experimental Child Psychology*, *103*(4), 421–440. https://doi.org/10.1016/j.jecp.2008.12.001

Barwise, J., & Cooper, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, *4*(2), 159–219.

Begolli, K. N., Booth, J. L., Holmes, C. A., & Newcombe, N. S. (2020). How many apples make a quarter? The challenge of discrete proportional formats. *Journal of Experimental Child Psychology*, *192*, 104774. https://doi.org/10.1016/j.jecp.2019.104774

Boyer, T. W., & Levine, S. C. (2015). Prompting children to reason proportionally: Processing discrete units as continuous amounts. *Developmental Psychology*, *51*(5), 615–620. https://doi.org/10.1037/a0039010

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*(5), 1478–1490. https://doi.org/10.1037/a0013110

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.72

Davis, G. E. (2003). From parts and wholes to proportional reasoning. *The Journal of Mathematical Behavior*, *22*(3), 213–216. https://doi.org/10.1016/S0732-3123(03)00020-8

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, *49*(2), 243–249. https://doi.org/10.1037/a0028278

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Halberda, J., Taing, L., & Lidz, J. (2008). The Development of “Most” Comprehension and Its Potential Dependence on Counting Ability in Preschoolers. *Language Learning and Development*, *4*(2), 99–121. https://doi.org/10.1080/15475440801922099

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The Trouble With Quantifiers: Exploring Children’s Deficits in Scalar Implicature. *Child Development*, *89*(6), e572–e593. https://doi.org/10.1111/cdev.13014

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the Acquisition of Numbers and Quantifiers. *Language Learning and Development*, *2*(2), 77–96. https://doi.org/10.1207/s15473341lld0202_1

Hurst, M. A., & Cordes, S. (2018). Attending to relations: Proportional reasoning in 3- to 6-year-old children. *Developmental Psychology*, *54*(3), 428–439. https://doi.org/10.1037/dev0000440

[dataset] Hurst, M.A. & Levine, S. C. (2022). Children's Understanding of "Most" for Discrete and Continuous Stimuli. *Open Science Framework.* DOI: 10.17605/OSF.IO/3N9UE

Hurst, M. A., Shaw, A., Chernyak, N., & Levine, S. C. (2020). Giving a larger amount or a larger proportion: Stimulus format impacts children's social evaluations. *Developmental Psychology*, *56*(12), 2212–2222. https://doi.org/10.1037/dev0001121

Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The Development of Proportional Reasoning: Effect of Continuous Versus Discrete Quantities. *Journal of Cognition and Development*, *8*(2), 237–256. https://doi.org/10.1080/15248370701202471

Kassambara, A. (2020). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (0.4.0) [Computer software]. https://CRAN.R-project.org/package=rstatix

McCrink, K., & Wynn, K. (2007). Ratio Abstraction by 6-Month-Old Infants. *Psychological Science*, *18*(8), 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

Möhring, W., Newcombe, N. S., Levine, S. C., & Frick, A. (2016). Spatial proportional reasoning is associated with formal knowledge about fractions. *Journal of Cognition and Development*, *17*(1), 67–84.

Ni, Y., & Zhou, Y.-D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, *40*(1), 27–52.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1

Odic, D., Pietroski, P., Hunter, T., Lidz, J., & Halberda, J. (2013). Young children's understanding of "more" and discrimination of number and surface area. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 451–461. https://doi.org/10.1037/a0028874

Papafragou, A., & Schwarz, N. (2006). Most Wanted. *Language Acquisition*, *13*(3), 207–251. https://doi.org/10.1207/s15327817la1303_3

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Studio Team. (2016). *RStudio: Integrated Development for R*. RStudio Inc. http://www.rstudio.com/

Sullivan, J., Bale, A., & Barner, D. (2018). Most Preschoolers Don't Know Most. *Language Learning and Development*, *14*(4), 320–338. https://doi.org/10.1080/15475441.2018.1489813

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wickham, H. (2017). *Tidyverse: Easily Install and Load the "Tidyverse". R package version 1.2.1*. https://CRAN.R-project.org/package=tidyverse