Preschoolers' Number Knowledge Relates to Spontaneous Focusing on Number for Small, but not Large, Sets

Sophie Savelkouls, Michelle A. Hurst & Sara Cordes

Boston College Department of Psychology

Author Note

Sophie Savelkouls and Sara Cordes, Department of Psychology, Boston College.

Michelle A. Hurst, Department of Psychology, University of Chicago.

Correspondence concerning this article should be addressed to Sophie Savelkouls, Department of Psychology, Boston College, 140 Commonwealth Avenue, 300 McGuinn Hall, Chestnut Hill, MA 02467. E-mail: sophie.savelkouls@gmail.com

Abstract

Much research has examined the reciprocal relations between a child’s spontaneous focus on number (SFON) in the preschool years and later mathematical achievement. However, this literature relies on several different tasks to assess SFON with distinct task demands, making it unclear to what extent these tasks measure the same underlying construct. Moreover, prior studies have investigated SFON in the context of small sets exclusively, but no work has explored whether children demonstrate SFON for large sets and how this relates to children’s math ability. In the current study, preschoolers were presented four distinct SFON tasks assessing their spontaneous attention to number for small (Experiment 1) and large (Experiment 2) sets of number. Results revealed performance across the four distinct SFON tasks was unrelated. Moreover, preschooler’s SFON for small sets (1-4 items) was significantly stronger than that for large sets (10-40 items) and analyses revealed that number knowledge was only associated with SFON for small sets, but not large. Together, findings suggest that SFON may not be a set-size independent construct, and instead may hinge upon a child’s number knowledge, at least in the preschool years. The role of number language and how it relates to children’s SFON are discussed.

Keywords: SFON; Number Knowledge; Salience; Set Size

Preschoolers' Number Knowledge Relates to Spontaneous Focusing on Number for Small, but not Large, Sets

There are documented individual differences in children's tendencies to pay attention to number in their natural environment, with some children naturally attending to number in their day-to-day lives more than others (Hannula-Sormunen & Lehtinen, 2005; Hannula et al., 2007). A child's propensity to focus on number without any prompting has been called Spontaneous Focusing on Number (SFON; Hannula & Lehtinen, 2005). In the literature, an important distinction is made between knowing specific numerical skills (e.g., how to count) and knowing that these skills are relevant to the task at hand (e.g., that counting might be a relevant strategy for a particular task). Although the majority of research in numerical cognition tends to focus on the former (what mathematical knowledge do children possess and how can we teach it to them?), recent studies investigating SFON have also begun to focus on the latter (when do children realize that number might be relevant?).

Studies have examined the reciprocal relation between SFON and math skills, finding not only that individual differences in SFON in the preschool years predict later measures of math achievement (Hannula-Sormunen, Lehtinen, & Räsänen, 2015; Hannula-Sormunen, Lepola, & Lehtinen, 2010; McMullen, Hannula-Sormunen, & Lehtinen, 2015), but also that numerical abilities (such as counting abilities) also predict SFON a few years later (Hannula-Sormunen & Lehtinen, 2005). For example, SFON around ages 3-4 predicts children's enumeration and counting skills at ages 5-6 (Hannula-Sormunen & Lehtinen, 2005; Hannula et al., 2007) and rational number knowledge at age 12 (McMullen et al., 2015). Moreover, SFON at age 6 is positively correlated with arithmetic skills but not reading skills two years later, suggesting that

SFON is domain-specific to number (Hannula-Sormunen et al., 2010) and not simply a proxy for domain-general capacities such as IQ or working memory.

Despite this research, there are still open questions regarding SFON. In the current study, we investigated two different research questions:

(1) *Are individual differences in SFON reliable across distinct SFON tasks?* Multiple tasks have been used to assess SFON, some that are verbally based and others that are behavioral (i.e., not reliant upon language), yet little is known about whether these tasks tap into the same SFON construct. There is some evidence to suggest that a child's SFON may vary substantially across distinct SFON tasks (Batchelor, Inglis, & Gilmore, 2015), however, this prior work focused on only two SFON tasks (Picture and Imitation tasks), leaving open questions about other widely used SFON task variations and how they all relate to each other. A systematic comparison of performance on different measures of SFON can therefore help clarify how context and task demands may affect a child's likelihood of spontaneously focusing on number.

(2) *How does set size affect SFON in preschoolers?* To date, investigations of SFON in preschoolers have mostly been limited to small sets (<5 items) which young children are more likely to be able to enumerate and count compared to larger sets. Yet the literature characterizes SFON as a generalized attention to any numerical information in the environment. Do children spontaneously attend to number even when they may be unable to accurately track the exact number of items present? If so, are individual differences in this large-number SFON meaningful? Or might evidence of SFON be dependent upon children's enumeration abilities? Getting an answer to these questions is important in furthering our understanding of the construct of SFON and the nature of its relation with other numerical abilities.

Distinct SFON Measures

There are three types of tasks that have traditionally been used to measure SFON: A) Imitation tasks, B) Picture tasks, and C) Choice tasks. Although all three tasks presumably assess a child’s spontaneous attention to numerical information, they involve very distinct task demands. Yet, all three have been used interchangeably across studies, making it important to investigate whether these tasks assess the same underlying SFON construct.

In Imitation tasks, children as young as 3 years are shown a series of actions by an experimenter and are asked to imitate the experimenter (Hannula-Sormunen & Lehtinen, 2005; Hannula-Sormunen et al., 2015, 2007). For example, Hannula and Lehtinen (2005) demonstrated an experimenter putting a certain number of letters in a toy mailbox and the child was asked to “do exactly like I just did” without any mention of number. The dependent variable in this task is whether or not the child imitates the number of actions the experimenter undertook, or alternatively, whether they use number words while doing their imitation. Thus, although children in this task can get credit for providing a linguistic response, providing a behavioral, nonverbal response is sufficient for getting credit for attending to number.

Picture tasks have typically been used with slightly older children (4-6-year-olds) compared to the Imitation task (Batchelor et al., 2015). In the Picture task, children are asked to describe what they see in a picture. In this case, the dependent variable is whether or not the child uses any number or quantity words in their descriptions. Importantly, both of these tasks hinge upon the ability to track exact number either verbally or behaviorally. Thus, any measure of SFON obtained from either the Picture task or the Imitation task necessarily requires an ability to encode exact number, something thought to be dependent upon number word learning.

Choice tasks have more recently been designed as another way to measure SFON (Cantlon, Safford, & Brannon, 2010; Chan & Mazzocco, 2017). Choice tasks involve an ambiguous match-to-sample task on a computer where children are asked to select the picture that “best matches” a sample picture (typically involving an array of items). On critical test trials, one of the choice pictures matches the sample picture in the number of items in the picture (e.g., two circles match two squares), while the other matches the sample on another quantitative dimension such as cumulative area (Cantlon, Safford, & Brannon, 2010), color, or shape (Chan & Mazzocco, 2017). By directly pitting number against other dimensions, these critical trials measure the child’s relative preference for number over this other dimension. Cantlon et al. (2010) found that 3- to 5-year-old children spontaneously focused on number over cumulative area at a rate higher than chance alone, a finding that has been replicated with English and Japanese populations (Cantrell, Kuwabara, & Smith, 2015). Moreover, Chan & Mazzocco (2017) found individual differences in preschoolers’ SFON in the Choice task to hold across 5 months.

Although all three tasks have been used in the literature to measure SFON, no work has explored whether these tasks tap into the same underlying construct. Notably, these tasks differ significantly in terms of their task demands, making it likely that performance on one measure will not relate fully to performance on another. In particular, the Picture task is an exclusively verbal task – requiring children to verbally describe the picture - whereas both the Imitation and Choice tasks are behavioral measures of SFON. The verbal requirements of the Picture task may prevent some children with limited communication abilities from being able to demonstrate an attention to number and furthermore it may require a level of comfort with number words that children in the preschool years simply do not have. This may explain why previous research has found that although performance on both the Picture and Imitation tasks predict arithmetic skills years later (Picture Task: Batchelor et al., 2015; Imitation Task: Hannula-Sormunen et al., 2010), performance on the two tasks do not correlate with one another (Batchelor et al., 2015; Rathé, Torbeyns, Hannula-Sormunen, & Verschaffel, 2016). This suggests that these tasks may tap into distinct aspects of SFON (verbal vs. behavioral) or alternatively, the verbal demands of the picture task may mask individual differences in underlying attention to number that children are unable to fully express verbally.

Furthermore, all three tasks vary in terms of the quantity and type of other features (besides number) available to the child. For example, in the Imitation task, children could attend to the color or orientation of the cards placed in the mailbox, approximate estimates of magnitude like “some letters” or “many letters”, the facial expressions or specific motor mannerisms of the experimenter, or any other of many possible features. Similarly, in the Picture task, anything in the picture is fair game. Stimuli in the Choice task, on the other hand, carefully control for other confounding features, such that the task carefully measures the relative salience of number when pitted against only one or two other dimensions (e.g., number pitted against cumulative area or ratio), and as such it allows for a more systematic (though somewhat limited) interpretation of SFON. We took advantage of this design feature in our study by administering two choice tasks, one in which number was pitted against cumulative area and another in which number was pitted against proportion or ratio. We included this Proportion Choice Task because although children can attend to proportional information as early as infancy (e.g., McCrink & Wynn, 2006), evidence suggests that older children (6-year-olds) attend more to number than proportion, leading to robust errors in proportional reasoning tasks (Boyer, Levine, & Huttenlocher, 2008; Hurst & Cordes, 2018). By including both a Number vs Area and a Number vs Proportion Choice task, we investigated how SFON may depend on the particular quantity-based alternatives that are available.

SFON as a Function of Set Size

One striking similarity between all SFON tasks, however, is that they have almost exclusively been used to test SFON with small sets (<5 items). This is an important limitation in the literature because if SFON, as it has been described, is a general attention to number, SFON should pertain to attention to all numerical information (regardless of set size), and should not depend upon an ability to enumerate the number of items, but instead reflect the recognition that number is a relevant dimension to attend. Since tests of SFON in preschoolers have only involved sets that they are also able to enumerate (i.e., count and identify the cardinality), it is unclear whether the ability to enumerate is a component of, or a necessary precursor to, demonstrating SFON for this age group. If SFON is truly a generalized attention to number, then it should not depend on the size of the sets involved nor on the child's ability to enumerate the sets. That is, children should demonstrate similar levels of SFON for large sets as they do for small sets.

The distinction between testing children on small and large sets is particularly important given what we know about children's developing number knowledge. Even before children master the count procedure – that is, before they acquire a full understanding of cardinality and the meaning of *all* number words – they can have an understanding of the meaning of *some* of the number words (Le Corre & Carey, 2007). The process of learning to count is a lengthy process that progresses through a series of stages over a period of 1-2 years. Research suggests that between the ages of 2.5 – 4, children start off as "subset knowers," meaning that they have an understanding of the meaning of a subset of numbers but have yet to grasp the cardinal principle more generally (i.e., that the last number word used in a count list is the cardinality of the set). Subset knowers go through a stepwise process where they learn the meaning of the number word "one", then "two" and so on. Not until after children learn the meaning of "four" or "five" do children acquire the cardinal principle – that the last number word used in a count refers to the cardinality of the set – and begin to understand the purpose of counting (thus becoming "cardinal-principle knowers" or CP-knowers; Wynn, 1992). At this point, it is expected that they now understand the meaning of all number words within their count list. This distinction between subset knowers and cardinal principle knowers is an important one because it represents qualitative differences in children's behavior. For example, cardinal principle knowers more consistently recognize that counting is an effective strategy and often use this as a strategy, whereas subset-knowers are less likely to spontaneously count in the face of a numerical task (Gordon, Chernyak, & Cordes, 2019; Le Corre & Carey, 2007; Posid & Cordes, 2018; Wynn, 1992).

Given that subset-knowers are able to accurately identify the cardinality of some small sets, but not large ones, it is important to determine how this cardinal understanding may impact the likelihood of the child demonstrating SFON. That is, do subset-knowers demonstrate SFON at similar levels for small and large sets? Do we see differences in SFON between subset-knowers and CP-knowers? If, as has been suggested by Hannula-Sormunen and collegues (2010), SFON is a generalized attention to number (i.e., independent of set size) then we should expect 1) similar levels of SFON for small and large sets and, 2) a child's actual knowledge of number (i.e., whether they are a Subset- or CP- knower) to be similarly related to performance on SFON tasks with small and large sets. On the other hand, if number knowledge does play a role in children's level of SFON, and these two skills are not as distinct as has been suggested in the literature, then we would expect 1) children to show greater SFON for small compared to large sets and, 2) children's number knowledge to only relate to SFON in cases in which the child is able to enumerate the sets involved (i.e., for small sets)[1].

In the current study, we administered four different SFON tasks to 2.5-5-year-old children: the Picture Task (Batchelor et al., 2015), the Imitation task (Hannula-Sormunen & Lehtinen, 2005) and two Choice Tasks, one in in which number was pitted against cumulative area (hereafter referred to as Area Choice task; Cantlon et al., 2010) the other in which number was pitted against numerical proportion (hereafter referred to as the Proportion Choice Task). Following these SFON tasks, children were administered the Give-N task - a standard measure of children's number knowledge and cardinal understanding (Wynn, 1992). To explore SFON for small and large sets, we manipulated the size of the sets presented in the Area Choice task across experiments, such that children were presented with exclusively small sets (<5 items) in Experiment 1 and exclusively large sets in Experiment 2. We chose to only manipulate set size in the Area Choice task since the Imitation task would be too cumbersome if children were expected to imitate as many as 5-10 actions, the Picture task requires a specific enumeration of a large set of objects which children may not have the number words for, and proportion matched stimuli involving exclusively small sets was impossible for the Proportion Match task because no

---

[1] To our knowledge only one study has tested SFON with both small and large sets. Cantlon et al. (2010) tested children on a choice task where children could either match stimuli using number or cumulative surface area, presenting them with sets ranging from 1-12, and they found that regardless of set size, children chose the number match over the cumulative surface area match at above chance levels. Their analyses comparing SFON across set sizes did determine that children showed the highest levels of SFON when presented with comparisons involving set size one (i.e., a single item). However, because the researchers presented small and large sets intermixed throughout, it was not clear whether the degree of SFON may have differed as a function of set size and/or whether the presence of small sets may have prompted children to attend to number during large set trials. Furthermore, this study did not address whether SFON for small and large sets were equally related to the child's number knowledge (i.e., subset- versus CP-knower), providing a true test of whether SFON is a generalized numerical construct.

two equivalent proportions can be generated from exclusively numbers less than four. Instead, the Area Choice task allowed for simple stimulus creation for small and large sets and did not require the ability to track exact number, making it possible for children to match based upon an approximate estimate of the number of items within the set.

Our first aim was to examine the relation between our four SFON measures to see 1) whether they were correlated with one another and 2) which measure of SFON correlated most strongly with Number Knowledge. Our second aim was to compare preschoolers' SFON for small and large sets by examining whether there were similar levels of SFON when presented with small (Experiment 1) and large (Experiment 2) sets in the Area Choice task. Given that previous research has shown a strong relation between SFON (when tested with small sets) in the preschool years and number knowledge, we assessed whether children's SFON for small and large sets similarly relate to their knowledge of number

EXPERIMENT 1

Method

Participants

Participants were 118 preschoolers (Range 2.5-5.1 year-olds; Mean age = 3.65, SD = .65, 73 Female, 45 Male). An additional 6 participants were excluded because of experimenter error (n = 3) or for only completing a single task or less (n = 3). A power analysis for a Pearson correlation using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) using alpha of .05, a power of .08 with a medium effect size (effect size = .03) gave us a desired sample size of about 84 participants. To ensure that our final sample was roughly equally distributed across our age range (to ensure we had a substantial number of both Subset- and CP-knowers), we aimed to recruit approximately 20-30 participants for every six-month interval across our age range (e.g., 2.5-3.0 years, 3.0-3.5 years, etc.).

Of our 118 participants, 66 participants completed all five of our tasks and the remaining participants completed at least two of the tasks (See Table S1 in Online Supplemental Materials for a breakdown of participants included in each task). Given that a substantial number of participants were not able to complete all tasks due to the length of all tasks combined, many of our analyses include only a subset of participants.

Participants were recruited from the Greater Boston Area and either participated in our campus laboratory or at their preschool or after school program. Of the 56% of our sample that provided demographic information, 86% of families identified as Caucasian, 5% as Asian, and 9% as biracial. Furthermore, 98% of mothers and 83% of fathers responded as having completed a bachelor's degree or higher. This study was approved by the Institutional Review Board of Boston College (protocol number: 10.064.11, The Development of Quantity Concepts) and all participants provided informed written consent.

Design

Participants completed five different tasks in the following order: Picture Task, Area Choice Task, Imitation Task, Proportion Choice Task, and the Give-N Task. The Give-N task was presented last because it is a measure of number knowledge and we did not want to cue in the participants that we were assessing number until after all the SFON tasks were administered.

Tasks & Procedure

Picture Task. *Adapted from Batchelor, Inglis, & Gilmore (2015).* Children were presented with three cartoon pictures taken from children's books, presented on laminated cardboard (21 x 21 cm). The pictures were chosen because they were fairly simple pictures that clearly contained small sets of items in a numerical range (1-4 items) that children could attend to and likely had verbal labels for. The pictures also contained many different colors, shapes, and animal characters, providing many other features to label and talk about, aside from number.

The experimenter introduced the task by saying: “This game is all about pictures. I am going to show you a picture and I want you to tell me everything you can see in the picture. Are you ready?” The researcher then put the first picture in front of the child and asked, “What do you see in this picture?” When the child was finished talking, the experimenter prompted the child twice more saying: “Great! What else do you see?” After the child responded to the final, third prompt, the researcher moved onto the next picture. Children were given three prompts to talk for each picture. If, after any of the prompts, the child said they did not see anything else in the picture, the experimenter immediately moved onto the next picture.

Area Choice Task. *Adapted from Cantlon, Safford, & Brannon (2010).* Children were first shown a single sample stimulus in the center of the tablet screen and were told: “I want you to look at this picture very carefully and when you are done remembering the picture, I want you to touch it.” If the child seemed reluctant to touch the tablet on the first few trials, the experimenter would prompt the child again or touch the tablet for them if they indicated they were done remembering the picture. Next, participants were shown two options and were asked: “Which picture best matches the one you just saw? This picture, or this one?” The experimenter would point to or circle each picture, to make it explicit what the two options were. The task was self-paced, meaning participants could choose when to move on from the sample stimulus and take as much time as needed to make their choice. Instructions were only repeated on the first few trials or whenever participants became distracted and needed re-prompting. At no point were children given any explicit instructions on how they should match the pictures.

Importantly, there were two types of trials: Standard and Probe trials. In Standard trials, one of the two choice stimuli matched the sample stimulus on *both* dimensions (i.e., number and cumulative area – only the configuration of the items changed; the correct match), while the other choice stimulus did not match on either of the two dimensions (incorrect match). In Standard trials, children were rewarded only for choosing the ‘correct match’, in this case a positive auditory and visual stimulus played on the tablet. Choosing the ‘incorrect match’ resulted in a red ‘x’ appearing on the screen with no auditory stimulus. In Probe trials, one of the choice stimuli matched the sample in terms of number (referred to as “number choice”) but not on cumulative area, while the other stimulus matched the sample in terms of the cumulative area (referred to as “area choice”), but not on number. In Probe trials, participants were rewarded regardless of their choice. See Figure 1 for an example of Standard and Probe trials.

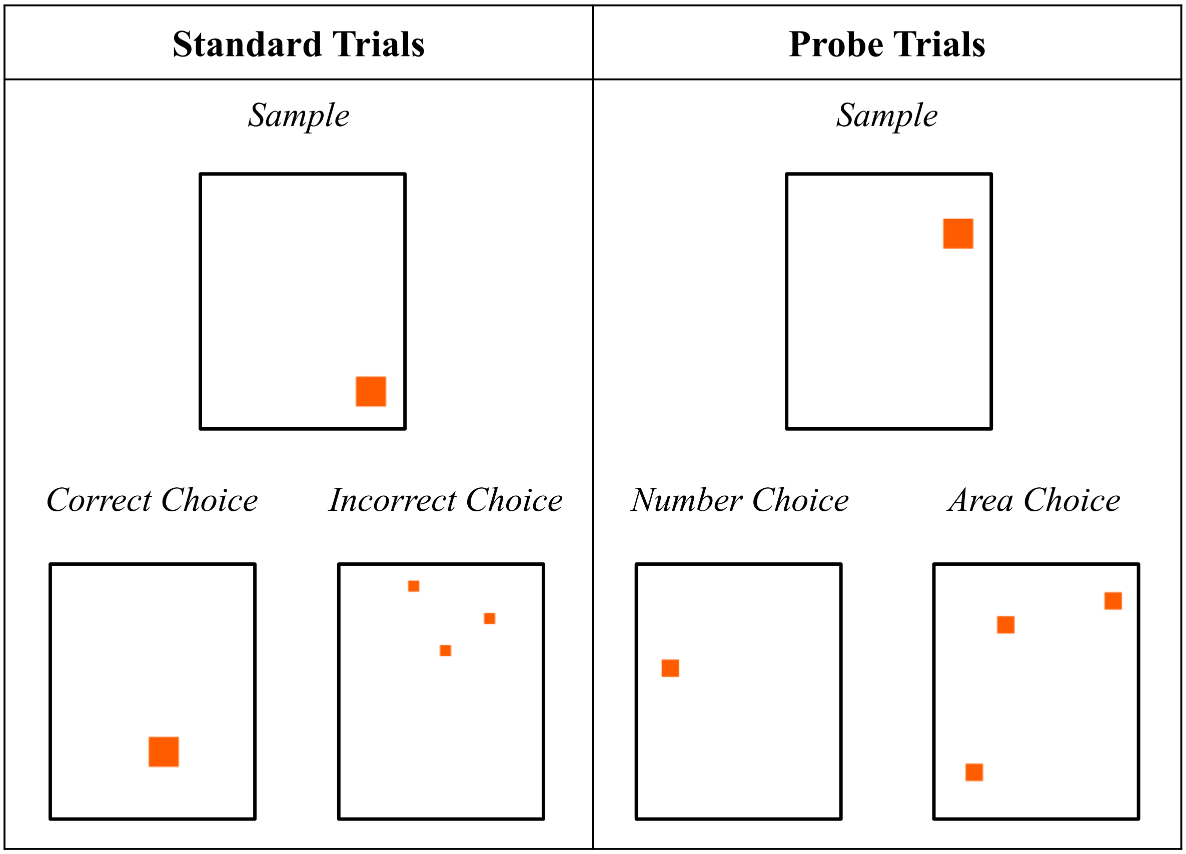

Figure 1. Stimuli from the Area Choice task from Experiment 1. For all trials, participants first saw a sample picture, followed by two choice pictures. For Standard Trials (left pane), the ‘correct choice’ matched the sample on both number *and* area and an ‘incorrect choice’ matched the sample on *neither* number *or* area. For Probe Trials (right pane), the ‘number choice’ matched the sample on number *not* area, and the ‘area choice’ matched the sample on area *not* number.

Participants first participated in 6 practice Standard trials, followed by 12 Test trials (a randomized mix of 6 new Standard trials and 6 Probe trials), for a total of 18 trials. To keep participants motivated, a short (16 sec) attractor video played halfway through each task.

Stimuli in the Area Choice task consisted of orange squares randomly placed on a white background (17 x 11 cm) with the element size of the squares homogenous within displays. All arrays contained anywhere from 1 to 4 squares, with one of three possible cumulative areas

(4800, 8800 and 12800 pixels^2). On Standard trials, the incorrect match had a different number of items (e.g., if the sample had 1 item, the incorrect match had 2, 3 or 4 items with equal frequency) and a different cumulative area, ensuring that 1) the size of the individual items were not identical to those of the sample and 2) the cumulative area was larger than the sample on exactly half of the trials. As such, the number ratio (smaller set size/larger set size) between the sample and incorrect match stimulus ranged from 0.25 to 0.75, and the cumulative area ratio ranged from 0.38 to 0.73. On Probe Trials, the number ratio between the sample and area match stimulus was the same as the cumulative area ratio between the sample and the number match stimulus (e.g., if the sample had 1 item with 4800 pixels^2, then the area match would have double the number of items (2) with the same cumulative area (4800), and the number match would have the same number of items (1) with double the cumulative area (9600 pixels^2)).

Imitation Task. *Adapted from Hannula & Lehtinen (2005).* The experimenter introduced this task by pointing to a small mailbox and 15 yellow laminated letters fanned out across the table and saying, “This is my mailbox and these are my letters. In this game, I want you to watch very carefully what I do, and then I want you to do exactly like I did.” The experimenter then proceeded by picking up one letter and putting it in the mailbox, then saying: “Now can you do exactly like I did and tell me when you are done?” If the child stopped putting letters in the mailbox, but did not tell the researcher that they were done, the experimenter would ask “Are you done doing exactly like I did?” waiting for the child’s confirmation before proceeding. For both the second and third trial, the experimenter repeated the same instructions but put 2 letters in the mailbox, one at a time, each with a separate motion. Children were allowed to put up to three letters in the mailbox but were stopped if they attempted to place more (the experimenter volunteered “Ok, now it’s my turn!”).

Proportion Choice Task. This task was identical in design to the Area Choice Task apart from the stimuli that were used. Stimuli consisted of red and blue items randomly placed on a white background (See Figure 2). On all trials, the sample stimulus was composed of a set of red and blue stars with a red star/total star ratio varying from 0.2 to 0.4 (average of 0.31). The two choice options were always composed of sets of 2-4 red and 3-16 blue dots (for a total ranging from 5 to 20 dots) on a white background. Additionally, we ensured that the ratio (small value/ smaller value) of the two choice stimuli was always .5 in terms of both proportion (smaller proportion/larger proportion) and number (small number/large number). On *Standard trials,* the correct option was identical in quantity to the sample (e.g., if the sample was 3/7 (3 red, 4 blue, 7 items total), the correct option was also 3/7 (3 red, 4 blue, 7 items total)), and only differed in arrangement and type of item (dots instead of stars). On the other hand, the incorrect option had the same total number of items as the sample (i.e., denominator) but had either half as many or twice as many red items (resulting in the proportion of red in the incorrect option ranging from 0.17 to 0.86 (average = 0.56); number of red dots ranged from 2 to 8; number of blue dots ranged from 3 to 10, number of total dots ranged from 5 to 18). For example, with if the sample was 3/7 (3 red, 4 blue, 7 items total), the correct option was also 3/7 (3 red, 4 blue, 7 items total) and the incorrect option would be a 6/7 (6 red, 1 blue, 7 items total)[2].

On the *Probe* trials, the number match had the same number of red items as the sample stimulus, but the proportion match had the same proportion (though different number) of red/blue items (e.g., sample: 4/10, red-number match option: 4/5, proportion match option: 2/5).

---

[2] One of the six practice trials was presented incorrectly as: 4/20 (Sample), 4/20 (Correct), and 2/10 (Incorrect), meaning that both options matched on proportion.

Again, the two options had the same total number of items as each other (i.e., same denominator), which was either half or twice as many as the sample stimulus.

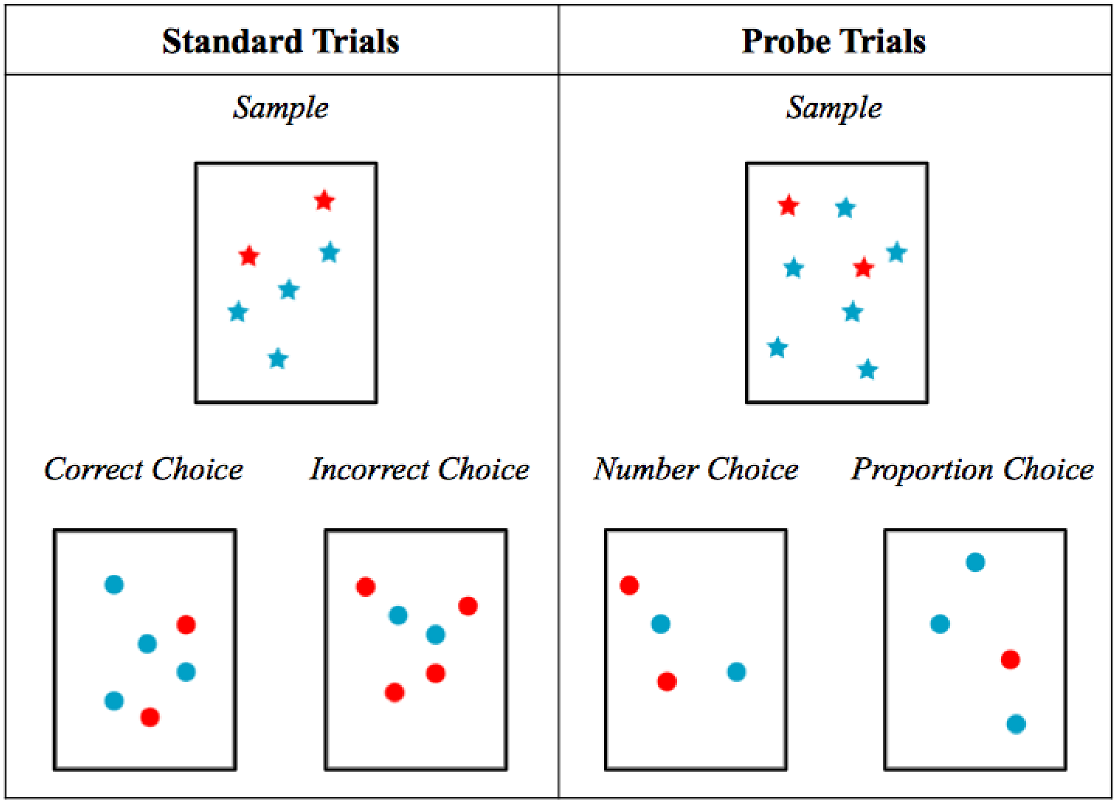

Figure 2. Stimuli from the Proportion Choice task from Experiment 1. This task was identical in procedure to the Area Choice Task. For Standard Trials (left pane), the Correct choice matched the sample on both Number *and* Proportion and an Incorrect choice matched the sample on *neither* Number *or* Proportion. For Probe Trials (right pane), the Number Choice matched the Sample on Number *not* Proportion, and the Proportion Choice matched the sample on Proportion *not* Number.

Give-N Task. *Adapted from Wynn (1992)*. Children were introduced to a pond (a small blue basket) and 20 small yellow rubber ducks and were told that the ducks like to go into the pond. The experimenter started by showing one duck jumping into the pond, and then after removing the duck, asked the child “Can you put *one* duck into the pond?” Once the child was done putting ducks into the pond, the experimenter verified “Is that *one* duck?” If the child said yes, the experimenter went onto the next trial; if the child said no, they were given an opportunity to fix what they had done until they were happy that there was *one* duck in the pond. If the child correctly put one duck into the pond, the experimenter asked for larger set sizes, with the number of ducks requested, N, increasing from 1-6. Using a titration method, each time the child successfully put N ducks into the basket, they were asked for N + 1 ducks, but if the child failed on N ducks, they were then asked for N – 1. To reduce the number of trials children had to perform, the experimenter skipped the set sizes of 2 and 5 ducks when going up the titration ladder. However, if the child failed to correctly place 3 or 6 ducks into the pond, then the experimenter asked for 2 or 5 ducks respectively. The task ended when the child: 1) succeeded in correctly placing N ducks into the pond twice and failed on N + 1 twice or 2) succeeded twice on the N = 6 trial. Children that failed to correctly put one duck into the pond twice were scored as having a knower level of 0, indicating that they were a pre-knower.

Data Processing & Coding

Picture Task. This task was transcribed using Computerized Language Analysis (CLAN), available through the Child Language Data Exchange System (CHILDES; MacWhinney, 2000). We used the CHAT (Codes for the Human Analysis of Transcripts) transcription format at the utterance level. For each of the three pictures (trials), we used CLAN software to perform a frequency count of any number or quantity-related words found in the transcripts that were said by the participants. The number and quantity words that we searched for included: the number words 1-10, many, more, less, little, lot, count, big, and small. As per Batchelor et al., (2015), on each trial children received a score of 1 if they used any number or quantity words (regardless of how many) and a score of 0 if they did not. Therefore, children could get a maximum score of 3 on this task. Twenty percent of participants were transcribed by a second coder, frequency analyses of quantity word use were done on these transcripts, and then these participants were also given a score 0-3. The level of consistency between both coders' scores were calculated using linear weighted Kappa, which resulted in a Kappa score of .81.

Choice Tasks. Performance on the Practice trials was not analyzed since these trials were designed to teach children the structure and goal of the matching game. Performance on Standard trials was used to measure participants' understanding of the task and was *not* a measure of SFON since participants could use both number and/or the other quantitative dimension (i.e., area or proportion) as a cue for matching. To measure SFON in this task we used the proportion of number matches on Probe Trials, such that higher scores reflect a greater tendency to match on number rather than on the other quantitative dimension (i.e., area or proportion).

Imitation Task. As per Hannula and Lehtinen (2005), children were considered to have spontaneously focused on number if participants met any of the following requirements: a) they put the same number of letters in the mailbox as the experimenter, b) their utterances included number words - regardless of whether they were the correct number words - (e.g., "I am putting in two at the same time") or quantity more generally (e.g., "How many did you put in?"), and/or c) they used gestures/fingers to denote numbers. Thus, children were given credit for displaying the correct numerical behavior (requirement a) and/or a numerical/quantitative verbal/gestural response (requirements b-c). Scores on each trial were binary such that children scored 1 if they demonstrated any or all of the above measures of SFON and 0 if they did not. Therefore, children could get a maximum score of 3 on this task (summing across the three trials). Twenty percent of participants were coded by a second coder and reliability between the two coders was calculated using linear weighted Kappa, which resulted in a Kappa score of .80.

Give-N. The child's Give-N score was the highest number, N, such that they got N correct at least twice and N+1 incorrect at least twice. If they placed 6 correctly in the basket twice, they were designated a CP knower.[3]

Results and Discussion

Individual SFON Measures

Picture Task. Overall, very few quantity words were produced in this task. The average number of trials on which children used number words was .97 out of 3 trials (See Figure 3).

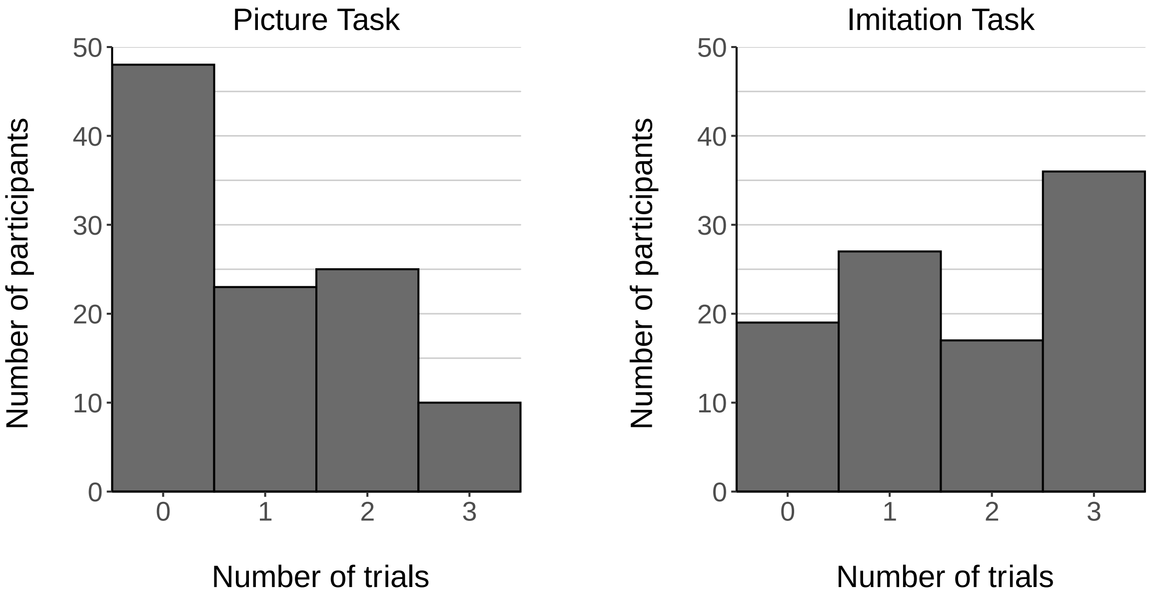

Figure 3. Histograms depicting the number of trials on which participants used quantity words (Picture Task, left) or imitated number (Imitation Task, right) in Experiment 1.

---

[3] About three quarters of our participants were tested on an extended version of Give-N in which they were asked for *N* from 1-10. For consistency, we will only be reporting participants score as 1-6. For those participants that received the extended version, we gave them a score of 6, if 1) they correctly placed 6 ducks into the pond twice or 2) if they correctly placed both 6 and 7 ducks into the pond.

Area Choice Task. Children performed significantly above chance (50%) on Standard Trials (65.5%; $t(104) = 6.05$, $p < .001$) and children chose the ‘number match’ significantly more often than chance on the Probe Trials (65.1%; $t(104) = 6.76$, $p < .001$). Thus, when number was pitted against area for small sets, children were more likely to attend to numerical information relative to area. Children’s performance on both the Standard and Probe trials correlated positively with age (Standard: $r = .41$, $p < .001$; Probe: $r = .29$, $p < .01$; See Table 1 for all correlations of Experiment 1).

Imitation Task. On average, participants scored as attending to number on 1.71 out of 3 trials with 80.8% of participants scoring as attending to number on at least one trial (See Figure 3). The majority of numerical responses involved correctly imitating the number of actions (84% of numerical responses were due to a correct imitation of the experimenter’s actions) suggesting higher evidence of SFON on the Imitation task was due to the fact that children were able to give a nonverbal behavioral response, rather than relying on their verbal skills alone .

Proportion Choice Task. Children did not perform significantly differently from chance on either the Standard (M=51.15%; $t(95) = .50$, $p = .62$) or Probe (M= 48.98% number selected; $t(95) = .43$, $p = .67$) trials. Furthermore, performance on neither trial type was significantly correlated with age (Standard: $r = -.04$, $p =.69$; Probe: $r = -.12$, $p =.25$). Because children failed to perform above chance on the Standard trials, it was unclear whether children understood the task demands. Thus, we looked at data from the subset of children who performed above chance (50%) on the Standard Trials (n = 31; M=77.2% on standard trials), as they are more likely to have understood the matching task in general. However, even these children did not select number (or proportion) at above chance levels on the Probe trials (M= 51.34% number selected; $t(30) = .32$, $p = .75$).

Relations between SFON tasks

Notably, there were no significant correlations in performance on any of our SFON tasks when controlling for age (*p*'s > .20; See Table 1)[4]. Secondary analyses explored whether the lack of correlation in performance across SFON tasks could be accounted for by the different response modes across tasks. For each participant we create a separate score for verbal and behavioral responses on the Imitation task, as this was the one task that allowed for both verbal and nonverbal behavioral responses. That is, participants received a behavioral score (out of 3) measuring correct behavioral imitation on each trial and a separate verbal score (out of 3) measuring whether they had used numerical language on each trial. Using this alternative scoring, we found that verbal responses in the Imitation Task correlated positively with (verbal) responses in the Picture task ($r = 0.22$, $p =.04$). Nonverbal behavioral responses in the Imitation task, however, did not correlate with nonverbal performance on the Probe trials of the Area Choice Task ($r = 0.11$, $p = .29$). These inconsistent findings hint at the possibility that response mode may have played some role in the level of SFON demonstrated by a child in a given task.

---

[4] For a heatmap of the simple correlations between tasks see Figure S1 in the Online Supplemental Materials.

Variable	Age	2	3	4	5	6	7
1. Picture Task	. 01 (*N*=106)	-.02 (*N*=94)	.04 (*N*=94)	-.11 (*N*=93)	.08 (*N*=87)	-.05 (*N*=87)	-.04 (*N*=87)
2. Area Choice Task– Standard	.41*** (*N*=105)		.46*** (*N*=105)	.13 (*N*=89)	.10 (*N*=93)	-.11 (*N*=93)	.25* (*N*=89)
3. Area Choice Task – Probe	.28** (*N*=105)			.04 (*N*=89)	-.04 (*N*=93)	-.20 (*N*=93)	.27* (*N*=89)
4. Imitation Task	.28*** (*N*=99)				-.10 (*N*=83)	-.01 (*N*=83)	.10 (*N*=84)
5. Proportion Choice Task - Standard	-.04 (*N*=96)					.06 (*N*=96)	.004 (*N*=83)
6. Proportion Choice Task - Probe	-.12 (*N*=96)						.003 (*N*=83)
7. Give-N	.72*** (N=97)						

Table 1. Correlation Matrix Experiment 1. The second column with the “Age” heading lists the correlations of each of our tasks with age. The rest of the table are pairwise partial correlations between our difference tasks when controlling for age.

*p< .05, **p<.01*** p<.005

The Relation Between Number Knowledge and SFON Measures

Children's performance on Give-N was quite variable ($M = 3.65$, Range: 0-6) and as expected, was highly correlated with age ($r = .72$, $p < .001$). Therefore, all analyses examining the relation between our SFON tasks and Give-N controlled for age (in months).

Surprisingly, Give-N performance was only positively correlated with performance on one SFON task - the Probe trials of the Area Choice task when controlling for age ($r = .29$, $p < .01$). None of the other SFON measures (Proportion Choice Task, Picture Task, or Imitation Task - whether using verbal, nonverbal or combined responses for the Imitation Task) correlated with Give-N performance (p's > 0.3). One-sample t-tests confirmed that both Subset- (57.5%; $t(51) = 2.46$, $p = .02$, $d = .32$) and CP-knowers (76.9%; $t(36) = 7.98$, $p < .001$, $d = 1.29$) performed above chance on the probe trials of the Area Choice Task[5].

Discussion

Findings from Experiment 1 reveal significant variability across our individual measures of SFON in preschoolers. In particular, whereas the Picture task revealed near floor performance – likely due to the verbal requirements of the task – preschoolers were significantly more likely than chance to match number on both the Imitation task and the Area Choice task. Thus, adding to other work in this domain, results of Experiment 1 reveal little to no consistency in SFON performance across distinct SFON tasks. The findings that verbal SFON on the Imitation task correlated with SFON in the verbal Picture task suggests that response mode played some role in

---

[5] As some researchers have classified children who can successfully produce a set of 5 (5-knowers) as CP-knowers, we repeated our analyses including 5-knowers as CP-knowers in the Online Supplemental Materials. Importantly, our pattern of results in this separate analysis mirrors that reported here.

determining the extent of SFON that children demonstrated across these three tasks. Otherwise, we saw no correlations between tasks that had different response modes (i.e., no correlations between the Picture and Area Task). These various SFON tasks, that have all traditionally been used to measure SFON, therefore may measure distinct aspects of a child's cognition (e.g., fluency with language), either separate from or in addition to measuring SFON.

Interestingly, despite prior reports relating SFON in these different tasks with counting and later math ability, our findings revealed that the only SFON measure that related to children's Give-N performance was the Area Choice Task. One possible reason for this is that the Area Choice task was the most straightforward of the SFON tasks and did not rely as heavily on verbal ability or other general skills required to imitate behavior than the other SFON tasks.

EXPERIMENT 2

In Experiment 2, we examined SFON for large sets of items (>4 items), specifically exploring whether the positive correlation between number knowledge and SFON is exclusive to small sets or holds across all set sizes. If SFON is a general numerical construct, then we predict (1) comparable levels of SFON for large sets as we do for small sets and (2) a relation between number knowledge and SFON for large sets – just as we find for small sets. In Experiment 2 we presented children with all the same tasks as Experiment 1, except the Area Choice task involved stimuli with large sets of items (10-40).

Method

Participants

Participants were 103 2.5 – 5.1 year-olds (Mean age = 3.70, SD = .78, 52 Female, 51 Male). An additional 5 participants were excluded for experimenter error (n=1), parental interference (n=1), or failure to complete more than one task (n = 3). Of our final sample, 74 participants completed all 5 tasks, with the rest completing a subset of the tasks (See Table S2 in Online Supplemental Materials for a breakdown of participants included in each task). Of the 45% of participants that provided us with demographic information, 74% identified as Caucasian, 8% as Asian, and 8% as biracial. Furthermore, 93% of mothers and 91% of fathers completed a bachelor’s degree or higher.

Tasks & Procedure

All tasks were identical to Experiment 1 except for stimuli in the Area Choice Task and the stimuli and structure of the Proportion Choice Task.

Area Choice Task. The procedures for this task were identical to Experiment 1, but the stimuli were purple squares involving set sizes of 10, 20, 30, or 40 items (multiples of 10 of the set sizes used in Experiment 1) with three possible cumulative areas (7200, 13200 and 19200 pixels^2; See Figure 4). All other details of the task were identical to Experiment 1. Importantly, prior research has revealed that children of this age should be able to reliably detect the numerical differences between the set sizes presented in Experiment 2 (Halberda & Feigenson, 2008). Thus, children should have been able to match sets based on number if they found it to be a relevant variable for matching.

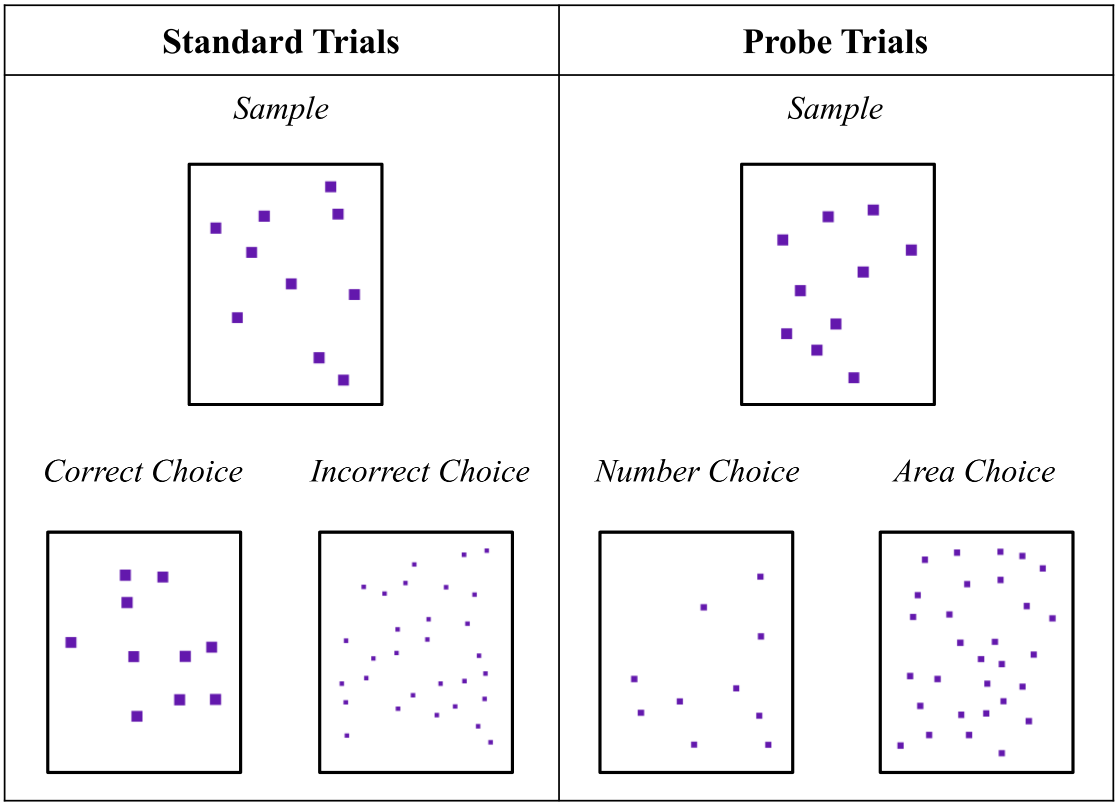

Figure 4. Stimuli from the Area Choice task from Experiment 2. The procedure was identical to the Area Choice Task from Experiment 1. The stimuli were changed such that the number of items in each display in Experiment 1 were multiplied by 10, thus creating large sets (ranging from 10-40 items).

Proportion Choice Task. Two significant changes were made to the Proportion Choice task to increase performance on this task: (1) the stimuli were simplified by presenting them as divided area models instead of discrete dots and (2) we included fewer trials of the spontaneous format used in Experiment 1[6].

---

[6] We also included a second block of trials in the Proportion Choice task to assess children's proportional reasoning in an unambiguous proportion matching task (based on Boyer et al., 2008). Because this second block of trials was included to assess children's proportional reasoning abilities, and not their SFON, description of the methods and results of this second block of trials is included in the Supplemental Materials accompanying the online article.

In the Proportion Choice task, children saw 4 Standard trials (with accurate feedback) followed by 4 Probe trials (with positive feedback regardless of selection). Each stimulus was made up of a single horizontal rectangle (width always 1.3 cm, height ranged from 2 – 8.1 cm) divided into 1.3cm x 0.7cm sized units, so that some of the units were orange and some of the units were blue. The orange units were always grouped together on the bottom of the rectangle (Figure 5). Standard and Probe trials had the same quantitative structure as in Experiment 1. On Standard trials, the sample stimulus had a range of 2, 3, or 4 orange units out of a total of 7-12 units. The proportion of orange out of the total on the sample stimulus ranged from 0.25 to 0.43 (average 0.35). On Standard trials, the correct stimulus was identical to the sample stimulus. That is, like in Experiment 1, it had the same number of orange and blue units as the sample. Notably, in Experiment 2 (unlike Experiment 1), because the blue and orange units were presented in a clear order within a rectangle, the sample stimulus and correct stimulus were visually identical (in Experiment 1 the arrangement of dots differed across stimuli). The incorrect stimulus had the same total number of units (i.e., same height rectangle), but differed in the number of orange units (number of orange: 2-8; total number of units: 7-12; proportion of orange: 0.17-0.86, average of 0.61). The two options differed in terms of the number of orange units and the proportion of orange units by an average ratio (smaller/larger) of .48 (range: .4 to .5).

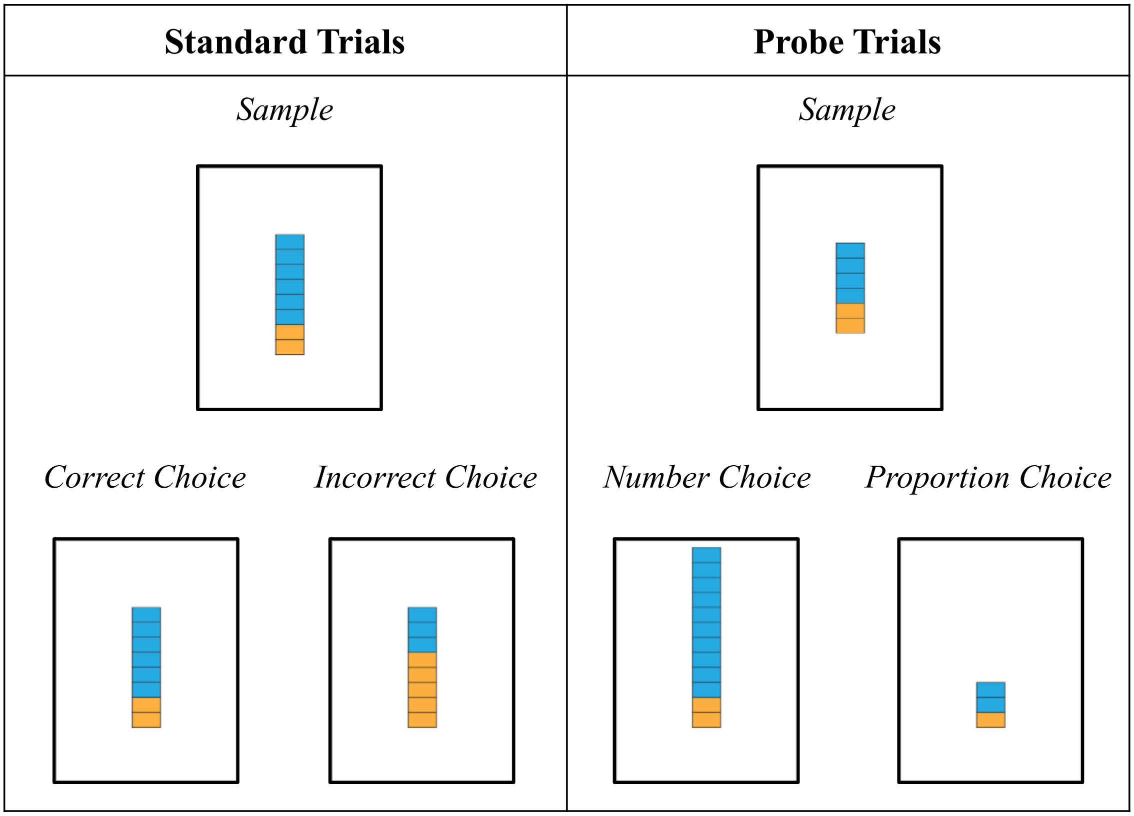

Figure 5. Stimuli from the Proportion Choice task from Experiment 2. The procedure was identical to the Proportion Choice task from Experiment 1.

Data Processing & Coding

All tasks were scored identically to Experiment 1. For the Picture and Imitation task, data from 20% of participants were transcribed and recoded by a second coder and reliability between the two coders was calculated using linear weighted Kappa which resulted in Kappa values of .93 and .87 for the Picture and Imitation tasks respectively.

Results and Discussion

Measures of SFON

Picture Task. Similar to Experiment 1, there were very few number words used during the Picture task. The average number of trials on which children used number words was .79 out of 3 trials (See Figure 6). However, unlike Experiment 1, we did find a significant correlation between number word usage on this task and age (r = .34, p < .001; See Table 2 for all correlations for Experiment 2).

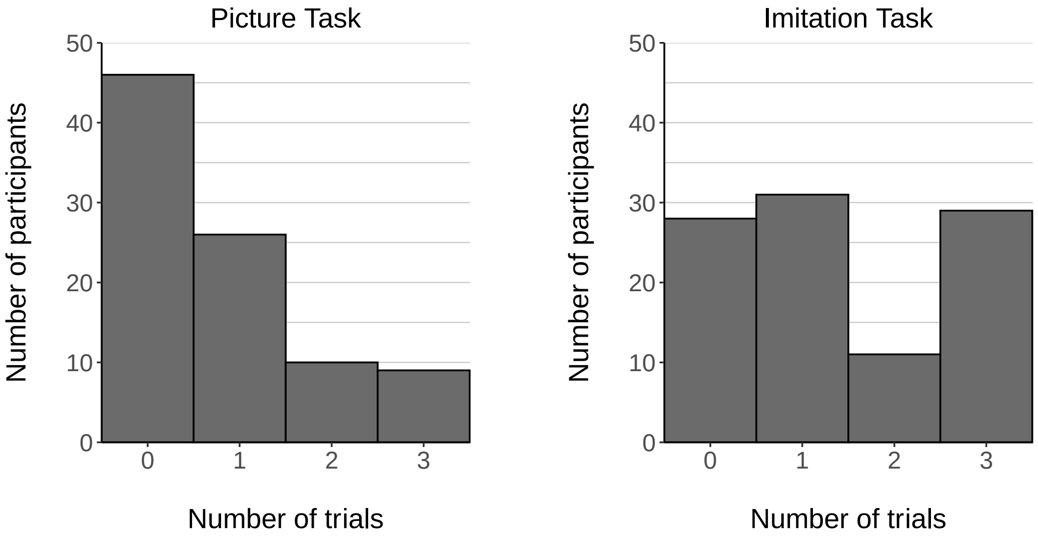

Figure 6. Histograms depicting the number of trials on which participants used quantity words (Picture Task, left) or imitated number (Imitation Task, right) in Experiment 2

Area Choice Task. In the context of large sets, participants performed significantly above chance on Standard trials of the Area Choice task (57.8%; $t(95) = 2.83$, $p < .01$, $d = .22$) and performance on the Standard trials was correlated with age ($r = .39$, $p < .001$).

However, participants were not more likely than chance to select the number match on Probe trials (50.2%; $t(95) = .09$, $p = .93$, $d = .009$), and performance was not correlated with age ($r = .08$, $p = .44$). Therefore, although children were able to make a match when both area and number were confounded in Standard trials, when number was pitted against area in the Probe trials, number did not continue to be a salient cue for matching. Thus, for large sets, participants no longer demonstrated a clear reliance on number and instead showed no preference for either number or area. Notably, a histogram of children's responses did not reveal this lack of significant preference to be the result of combining two bimodal distributions - one with children with a strong numerical preference and another with children with a strong area preference (see Figure 7). Instead, probe scores appeared to be fairly normally distributed suggesting that individual children did not reveal strong preferences for number or area.

Imitation Task. SFON on the Imitation task was again higher than that of the Picture task, the average number of trials on which children used number words was 1.41 out of 3 trials, with 72% of our participants imitating number on at least one trial (See Figure 6). Similar to Experiment 1, it appears that better performance on the Imitation task was driven by children's behavioral nonverbal responses (93.9% of SFON scores were behavioral, not verbal). Similar to the Picture task, here too performance was correlated with age ($r = .21$, $p = .03$).

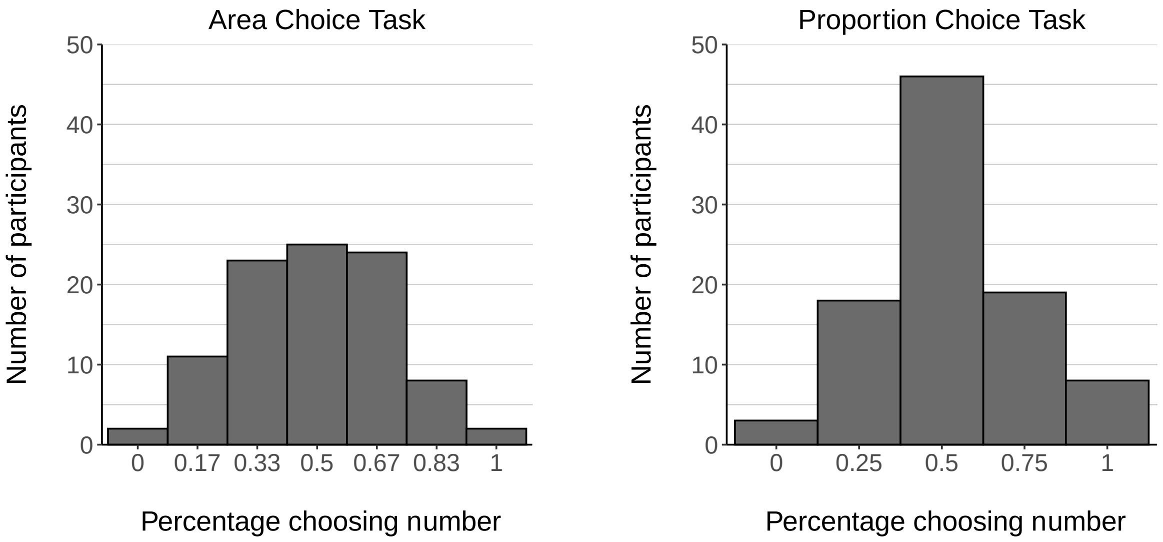

Figure 7. Histograms depicting children's preferences for number on the Area Choice Task and the Proportion Choice Task in Experiment 2.

Proportion Choice Task. In contrast to Experiment 1, children performed above chance on the Standard trials of the Proportion Choice task (61.8%; $t(94) = 4.31$, $p < 0.001$), but did not perform significantly differently from chance on the Probe trials (choosing to match on number 53.2% of the time; $t(94) = 1.33$, $p = 0.19$). Thus, children were able to match stimuli when the stimuli matched exactly (i.e., when the correct stimulus matched on proportion, amount of blue, and amount of orange), but did not reliably use a single feature when proportion and number were pit against each other. Again, a histogram of the number of children choosing number at different levels (see Figure 7) was found to be fairly normally distributed suggesting that most children randomly selected their responses on Probe Trials with no clear strategy. Lastly, performance on the standard trials was significantly correlated with age ($r = 0.57$, $p < 0.001$), but this was not the case for Probe trials ($r = 0.11$, $p = 0.3$).

Relations between SFON tasks

There were no significant correlations between any of our SFON tasks when controlling for age (*p*'s>.05; see Table 2)[7]. As in Experiment 1, for each participant we created a separate score in the Imitation task of only behavioral and verbal responses to determine the extent to which task demands may have played a role in performance on these SFON tasks. Again, verbal responses in the Imitation Task correlated positively with performance on the Picture task (*r* = 0.26, *p* =.02) however the behavioral responses did not correlate with any of the SFON tasks (*p's*>.25).

---

[7] For a heatmap of the simple correlations between tasks see Figure S2 in the Online Supplemental Materials.

Variable	Age	2	3	4	5	6	7
1. Picture Task	.34*** (*N*=92)	.22* (*N*=85)	.12 (*N*=85)	-.05 (*N*=89)	.12 (*N*=86)	.08 (*N*=86)	.16 (*N*=87)
2. Area Choice Task– Standard	.39*** (*N*=95)		.09 (*N*=95)	-.11 (*N*=92)	.09 (*N*=91)	.01 (*N*=91)	.08 (*N*=90)
3. Area Choice Task – Probe	.08 (N=95)			.08 (*N*=92)	.06 (*N*=91)	.09 (*N*=91)	.06 (*N*=91)
4. Imitation Task	.21* (*N*=100)				.01 (*N*=92)	-.05 (*N*=92)	.08 (*N*=93)
5. Proportion Choice Task - Standard	.57*** (*N*=95)					.09 (*N*=95)	.03 (*N*=91)
6. Proportion Choice Task - Probe	.11 (*N*=95)						.02 (*N*=91)
7. Give-N	.74*** (*N*=96)						

Table 2. Correlation Matrix Experiment 2. The second column with the “Age” heading lists the correlations of each of our tasks with age. The rest of the table are pairwise partial correlations between our difference tasks when controlling for age.
*p< .05, **p<.01*** p<.005

Number Knowledge and SFON

Performance on our Give-N task was again highly variable ($M = 4.13$, Range: 0-6) and strongly correlated with age ($r = .74$, $p < .001$). In contrast to Experiment 1, we found no correlation between any of our SFON tasks and number knowledge when controlling for age (p's $> .10$, r's $= .02$ to $.16$). Therefore, while number knowledge correlated with children's SFON in the Area Choice Task when sets were small (in Experiment 1), this was not the case with large sets. Moreover, subsequent analyses revealed that neither Subset Knowers (46.8%; $t(49)=1.08$, $p=.29$, $d = -.15$) or CP knowers (54.3%, $t(39)=1.23$, $p=.23$, $d = .19$) matched based on number above chance, nor did Subset and CP knowers differ significantly in their performance on Probe trials ($t(88)=1.65$, $p=.10$, $d = .33$)[8].

Discussion

Mirroring findings of Experiment 1, participants in our study did not perform consistently across our four measures of SFON. We again found that verbal responses in the Imitation Task correlated positively with performance on the Picture task, suggesting that both tasks may measure a form of verbal SFON. Otherwise none of our SFON measures were correlated with one another. Again, findings suggest that these four tasks measure distinct aspects of a child's cognition, and not a singular construct of SFON. These findings raise questions regarding the validity of these measures in assessing children's SFON since task demands and/or context differences seem to play a role in the level of SFON preschoolers demonstrate.

---

[8] Similar to Experiment 1, we conducted the same one-sample t-tests using a Subset- CP-knowers cut-off point of 5. Both Subset- and CP -knowers performed at chance on the probe trials of the Area Choice Task. See the Online Supplemental Materials for more details.

In contrast to Experiment 1, when presented with large sets in the Area Choice Task of Experiment 2, children did not focus on number over cumulative area more than chance. Given that children, of all knower levels, should have easily been able to discriminate between the large numbers we presented them (preschool-aged children have been shown to discriminate a 3:4 of change, the hardest ratio tested here; Odic, Libertus, Feigenson, & Halberda, 2013), this finding suggests that a lack of focusing on number cannot be explained by an inability to detect differences in the numerosities presented. Instead these findings lead us to conclude that when sets are large, number becomes *less* salient to children. Notably, although children did not select the numerical match at above chance levels, they also did not select the area match at above chance levels. Thus, it is not the case that children found area to be more salient in the context of large sets, but rather neither number nor area was a salient dimension in its own right.

Furthermore, we also did not see a relation between children's preference for number in the context of large sets (Area Choice task) and cardinal knowledge. Combined with findings from Experiment 1, this finding suggests the possibility that SFON may not necessarily be an independent construct of numerical attention in general, but instead may reflect an ability to attend to numbers they can quantify in the world around them. When tested only on small sets, those children with greater number knowledge focused on number. However, when tested on large sets - sets that went beyond the scope of children's number knowledge - we no longer saw this relationship between children's knowledge of number words and their SFON. A direct comparison between our findings in Experiment 1 and 2, may be able to tell us more about this different pattern of findings for Small and Large sets.

Combined Analyses Experiment 1 & 2

Preliminary analyses revealed there were no significant differences between participants in Experiments 1 and 2 in terms of children's age or Give-N performance (*p*'s >.5).

Next, we ran an ANCOVA with Age as a covariate comparing performance on the Standard trials of the Area Choice Tasks in Experiments 1 and 2 and found a significant effect of Experiment ($F(1, 197) = 5.51$, $p = .02$, $\eta = .03$), with participants performing significantly better on Standard trials when presented with small (65.5%) compared to large (57.8%) sets. A similar analysis on probe trial performance similarly revealed that participants were significantly more likely to select a number match on Probe trials when presented with small (65.1%) compared to large sets (50.2%; $F(1, 197) = 22.58$, $p < .001$, $\eta = .10$)[9].

Next, we ran a regression to test to what extent performance on the Probe trials was dependent on age, set size (Experiment 1 vs Experiment 2), and number knowledge (Give-N; see Table 3 for full results). To look at the main effects of our predictors, in Model 1 we included Age, Set Size (dummy coded as 0 = Experiment 1 and 1 = Experiment 2) and Number Knowledge as predictors. Overall, the model explained significant variance in performance, $R^2=.16$, $F(3, 175) = 10.73$, $p<.001$. Furthermore, we found significant unique effects of both set

---

[9] We explored whether the better performance on small sets may have been driven by those trials that included one item in the sample (based on Cantlon et al., 2010). Thus, we reran the above analyses excluding trials that included 1 item in the sample for Experiment 1 (and the matched trials in Experiment 2 that included 10 items in the sample). An ANCOVA with Age as a covariate found no significant difference in performance on small (62.0%) and large sets (57.0%) for standard trials ($F(2, 197) = 1.80$, $p = .18$, $\eta = .01$), although participants continued to be significantly more likely to select a number match on probe trials when presented with small (66.3%) compared to large sets (52.3%; $F(2, 195) = 17.65$, $p < .001$, $\eta = .08$). Thus, performance on the Standard trials of the Area Choice Task of Experiment 1 may have been boosted by performance on trials involving a single item, but importantly, the difference in performance when presented with small and large Probe trials was not attributable to the presence of single item trials.

size and number knowledge. Specifically, better number knowledge was associated with greater matching on number on the probe trials (b = .03, SE_b =.01. ß = .25, *p*<.05) and we also found a significant main effect of set size, such that children were less likely to match on number with large sets in Experiment 2 relative to small sets in Experiment 1 (b = -.16, SE_b =.03. ß = -.33, *p*<.001).

To further understand how set size and number knowledge interacted with one another we ran a second regression model (Model 2 in Table 3) that included the Experiment x Give-N interaction variable. The inclusion of this interaction term lead to a significant R^2 change relative to Model 1 (R^2 change =.02, *F*(1, 174) = 4.43, *p*=.037), suggesting that the interaction added significant explained variance. Specifically, there was a significant interaction such that the effect of number knowledge was higher for the small sets (Experiment 1) than for large sets (Experiment 2) (b = -.04, SE_b =.02. ß = -.37, *p*=.037). In particular, when presented with small

Variable	B	Std. Error	ß	t	p	R^2	R^2 change
Model 1						.16	.14
Age	-.01	.03	-.04	.34	.74		
Experiment	-.16	.03	-.33	4.70	<.001		
Give-N	.03	.01	.25	2.42	<.05		
Model 2						.18	.02
Age	-.004	.03	-.01	.12	.90		
Experiment	-.01	.08	-.03	.18	.86		
Give-N	.05	.01	.38	3.19	<.01		
Experiment x Give-N	-.05	.02	-.37	2.11	<.05		

Table 3. Regression with Probe Trials as our dependent measure. Experiment 1 (Small sets) was coded as 0, and Experiment 2 (Large sets) was coded as 1.

*p< .05, **p<.01*** p<.005

sets (Experiment 1) each one level increase in knower level was associated with a 5% increase in performance on Probe trials (b = .05, SE_b =.01. ß = 0.38, *p*=.002). However, when presented with large sets (Experiment 2) each one level increase in knower level was not associated with a significant change in performance (based on a reversed dummy coding with Experiment 2 = 0 and Experiment 1 = 1; b = .01, SE_b =.02. ß = .09, *p*=.48). This confirms the pattern of findings in Experiments 1 and 2 separately indicating that number knowledge only had a significant effect on Probe trial performance when the Probe trials included small sets, and this was a significantly different pattern that that shown for large sets in Experiment 2 (See Figure 8).

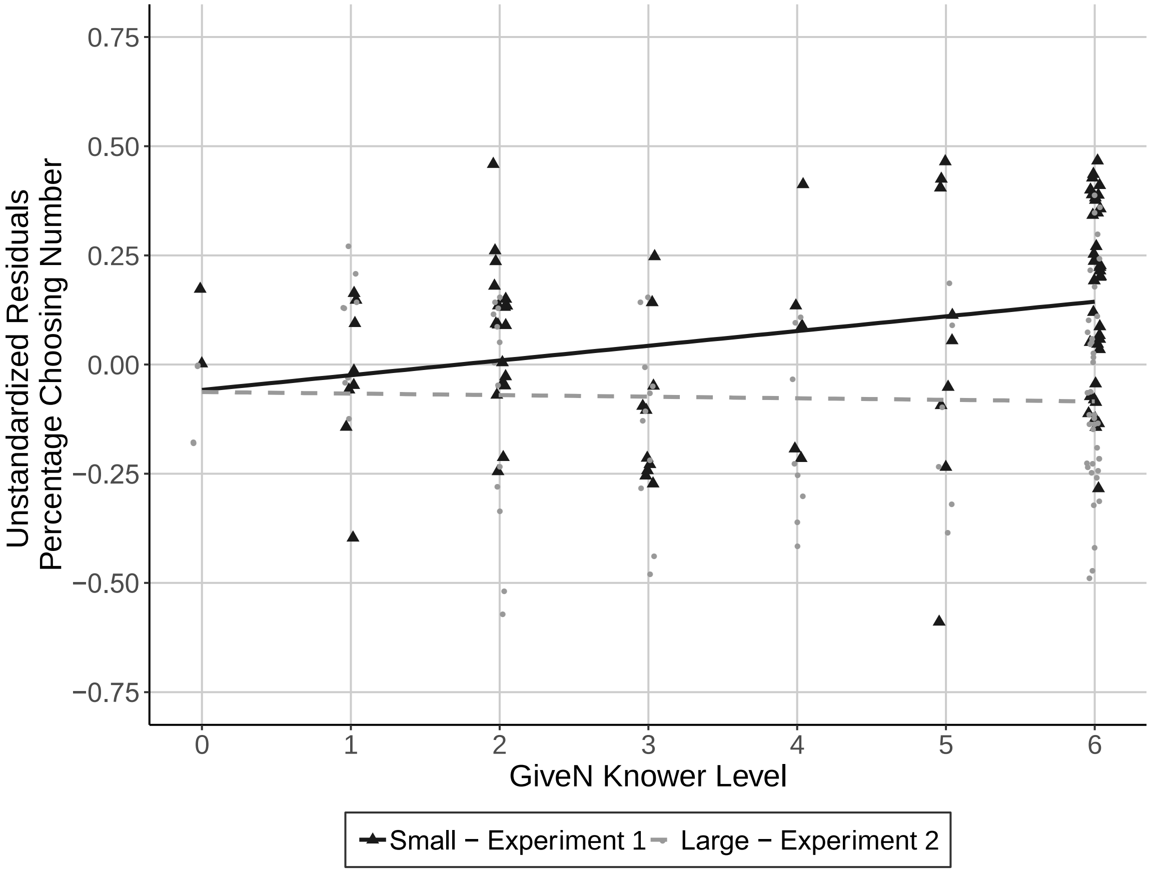

Figure 8. Scatterplot displaying the relationship between participants' Number Knowledge (measured through Give-N task) and their percentage choosing the number match on the Area Choice Task (using unstandardized residuals), when controlling for age on Experiments 2 and 2. Data points have been jittered to reduce overplotting.

General Discussion

The current study had two aims: (1) to examine the relations between four different tasks that have been previously used to measure children's SFON and their relation to a child's number knowledge; and (2) to assess whether preschoolers demonstrate similar levels of SFON in the context of small (Experiment 1) and large (Experiment 2) sets, and determine whether both are equally related to number knowledge. Given the evidence showing that SFON, tested in the preschool years, is predictive of later long-term measures of math achievement (Hannula-Sormunen et al., 2015, 2010; McMullen et al., 2015), it is important that we get a better understanding of what the limits of SFON tasks are and what construct these tasks are really measuring, to further our understanding of its relation with other numerical and mathematical abilities.

Relation Between SFON Tasks

Our first research question pertained to how performance on different SFON tasks may be related, allowing us to measure how context and task demands affect children's tendency to attend to number. We tested children on four different SFON tasks (Imitation Task, Picture Task, Area Choice Task, and Proportion Choice Task) none of the tasks correlated with one another when we controlled for age. The only exception was that when performance on the Imitation task was broken down by whether children had given a verbal or behavioral response, verbal responses alone correlated with Picture task performance in both Experiments 1 and 2. These findings therefore replicate, to some extent, previous research showing that performance on the Picture and Imitation tasks as a whole do not correlate with one another (Batchelor et al., 2015; Rathé et al., 2016), suggesting that these two tasks are tapping into unique behavioral constructs.

Instead, these findings suggest that the difference in response modes of the three tasks may play a role in the level of SFON children show. Although the Picture Task relied entirely on verbal expression of numerical information, the Imitation task allowed children to express their attention to number either verbally or through behavioral imitation. Our finding that only verbal responses, but not behavioral responses, in the Imitation task correlated with Picture task performance, and that verbal responses made up a very small percentage of numerical responses in the Imitation task (in Experiments 1 and 2 respectively, 84% and 93.9% of numerical responses were behavioral, not verbal), suggests that the verbal nature of the Picture Task may hinder children's expression of SFON. It seems likely that being able to spontaneously talk about number is a skill that develops after the ability to imitate or act using number (as in the Imitation or Choice Task). This is in line with other literature on children's use of gesture, suggesting that while children may not be able to verbally express their emerging knowledge or skills, they may be able to express it using a behavioral mode like gesture (Goldin-Meadow & Breckinridge Church, 1986). In a sample of adult participants (reported in Online Supplemental Materials only)[10] we found that adults were significantly more likely to talk about number than our child participants in the Picture Task, suggesting that the verbal limitations of the task may have played a role in the low scores of numerical focus or attention. As such, on top of measuring SFON, the different SFON tasks may also capture individual differences in other basic cognitive skills related to response mode. One way to investigate this possibility would be to

---

[10] A third experiment was conducted with adults testing them on the Pictures Task and the small and large version of the Area Choice task, along with a few measures of math abilities. This Experiment generally showed high levels of SFON on all tasks, although adults showed a significantly greater likelihood of selecting the number match for small, compared to large, sets. See Online Supplementary Materials for a full report of Experiment 3.

systematically manipulate the response mode of SFON tasks. Taking the Imitation task as example, one could ask children either to imitate (“do exactly as I did”) or describe (“tell me what I did”) the researcher’s actions. Another way to measure the role of verbal abilities on task performance would be to include a test of children’s verbal abilities in future studies using the Picture task to determine to what extent children’s verbal proficiency may impact their ability to display SFON in this task.

Another, non-mutually exclusive possibility for why participants performed so differently on these different tasks is that the different tasks, intended to be measuring SFON, may not measure the same underlying SFON construct. When we take a look at the three existing measures of SFON and what they have in common and where they differ, it becomes clear that there seem to be some discrepancies in what they are measuring. For example, it is unclear what role accuracy must play in demonstrations of SFON. In the Imitation task (apart from the rare case where children used number words during the task) the measure of SFON is also an accuracy measure. Children who may have been attending to number but fail to accurately imitate the correct exact number of actions, would not be considered to be engaging in SFON. While this is also the case in our Choice task, it was less of an issue because the particular numbers presented were expected to be within the range of values that children can track and compare using basic estimation abilities (Halberda & Feigenson, 2008). In contrast, in the Picture task, regardless of whether children’s number word use refers to the correct amount, as long as they use number words, they are given credit for engaging in SFON.

The Impact of Set Size on SFON

Our second research question concerned how set size may impact the likelihood of a child demonstrating SFON. Prior research has only tested SFON with small sets (<4 items), sets that preschoolers are typically able to enumerate and have the number words for, making it difficult to determine whether SFON is a general numerical construct or specific to enumerable numerosities. If, as Hannula-Sormunen and collegues (2010) claim, SFON is general attention to discrete numerical information, then we should expect two things: (1) SFON should systematically differ across set sizes and (2) SFON for all set sizes, small and large alike, should relate to number knowledge. To explore these possibilities, we presented children with an Area Choice task involving small sets (Experiment 1) and large sets (Experiment 2), allowing us to compare levels of SFON across set sizes. In contrast to predictions of a single construct of SFON, preschooler's SFON (as measured by probe trial performance on the Area Choice task) for small sets (1-4 items) was significantly greater than that for large sets (10-40 items). Moreover, regression analyses revealed that number knowledge was more strongly associated with SFON for small sets, than large sets (which was, in fact, not statistically significant).

These findings suggest that children's number knowledge may play an integral role in SFON, at least in the preschool years. Support for this idea comes from research suggesting that language plays an important role in solidifying certain concepts and even remembering them across time. For example, in the domain of color, participants perform better at color discrimination tasks (Winawer et al., 2007) and have better memory for colors (Uchikawa & Shinoda, 1996) when the colors they are tested on have distinct linguistic labels (e.g., shades of green vs. blue), compared to when they are part of the same linguistic category (e.g., shades of blue). Similarly, in the domain of number, members of societies such as the Amazonian Pirahã tribe, whose language does not have words to represent exact numerical quantities, have difficulty matching the exact cardinality of large sets, suggesting that language for number may be particularly useful in terms of memory and attention for numerical information (Frank, Everett, Fedorenko, & Gibson, 2008).

Relating these findings to what we know about SFON, it is possible that in the preschool years when children are learning number words, these words allow for the encoding of numerical information in their memory, and this improved memory for, and awareness of, number may be the primary driver of individual differences in SFON at this age. Thus it is possible that at least in the preschool years, SFON in the way that it is currently assessed, may in part be a reflection of a child's ability to encode number exactly – that is, SFON may be better described as a proxy for children's enumeration abilities and not their spontaneous focusing on *any* numerical information (i.e., children's representation of large sets). Furthermore, this could mean that findings showing that SFON in the early preschool years (Hannula-Sormunen & Lehtinen, 2005) predicts later math ability, may be accounted for by the fact that number knowledge predicts later math ability (e.g., Libertus, Feigenson, & Halberda, 2011). Future studies should explore whether the relation between SFON and later math abilities holds when controlling for differences in number knowledge.

Alternatively, the finding that children's SFON for large sets was significantly smaller than that for small sets may be because small and large sets invoke distinct representational systems. Prior research suggests that children rely upon exact, set-size limited object files for tracking small sets, whereas the tracking of large sets of items is thought to be dependent upon a noisy, approximate system termed the Approximate Number System (ANS). It may be the case that children rely upon two distinct types of SFON: object-file dependent SFON for small sets and ANS dependent SFON for large sets. If so, then it may not be surprising that the SFON measured for large sets did not relate to a child's number knowledge – an ability currently thought to rely almost exclusively upon object file representations (Le Corre & Carey, 2007; Wynn, 1992; but see Wagner & Johnson, 2011). Future research should explore whether individual differences in SFON for large sets may be relevant for other numerical abilities which rely upon the ANS such as numerical discrimination or the acquisition of Arabic numerals (e.g., Halberda, Mazzocco, & Feigenson, 2008; Moyer & Landauer, 1967).

One surprising finding was that we did not replicate previous research showing a relation between children’s performance on the Imitation task and their number knowledge (Hannula-Sormunen & Lehtinen, 2005; Hannula et al., 2007). We believe this can be explained by a difference in our measures of number knowledge. Hannula-Sormunen and Lehtinen (2005) measured cardinality using the “caterpillar task” which presents children with caterpillars with a different number of legs and they are then asked to bring “just enough” socks for all the legs. This task in many ways resembles an imitation or choice task, since children are first presented with the quantity (e.g., number of legs) they need to imitate or match (e.g., number of socks). On the other hand, our measure of cardinality was the more commonly used Give-N task where children are verbally instructed to put a certain number of ducks into a pond, however this task cannot be solved through imitation or matching. It is therefore possible that the structural similarity between the caterpillar task and the SFON imitation task could explain why Hannula-Sormunen and Lehtinen (2005) found a relationship between number knowledge and the Imitation task, while we did not.

In light of our findings, how should we interpret past research showing that SFON relates to children’s later arithmetic and math achievement (Hannula-Sormunen & Lehtinen, 2005; Hannula-Sormunen et al., 2010)? Although we do not doubt that the ability to attend to number plays an important role in children’s numerical development, given that preschool-aged children only demonstrate SFON for small sets once they have reliably learned how to track number via the counting process, it is clear that acquiring a symbolic system (i.e., language) that encodes number plays a very important role in what children pay attention to. In fact, we propose that in preschool, SFON is not a truly independent construct from cardinal knowledge and enumeration and that instead it is a reflection of an ability to encode exact number (i.e., small sets). In that case, the relation between SFON and later math ability may be driven by the strong correlation between number knowledge and math ability. Furthermore, the fact that task demands played such an important role in whether or not children demonstrated SFON, further supports our hypothesis that SFON, as has been tested in the current literature, does not seem to be a distinct construct from number knowledge in the preschool years.

In breaking down how different measures of SFON relate to children's number knowledge, we have gained a better understanding of some of the informal and spontaneous practices that young children engage in with respect to number and math. Results of our study suggest that current definitions of SFON may not fully account for the task-dependent and set-size dependent nature of this construct. As such, future research in this domain should reconsider how SFON is defined and whether it assesses a general numerical attention or instead whether it is a proxy for numerical abilities.

References

Batchelor, S., Inglis, M., & Gilmore, C. (2015). Spontaneous focusing on numerosity and the arithmetic advantage. *Learning and Instruction*, *40*, 79–88. https://doi.org/10.1016/j.learninstruc.2015.09.005

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

Cantlon, J. F., Safford, K. E., & Brannon, E. M. (2010). Spontaneous analog number representations in 3-year-old children. *Developmental Science*, *13*(2), 289–297. https://doi.org/10.1111/j.1467-7687.2009.00887.x

Cantrell, L., Kuwabara, M., & Smith, L. B. (2015). Set size and culture influence children's attention to number. *Journal of Experimental Child Psychology*, *131*, 19–37. https://doi.org/10.1016/j.jecp.2014.10.010

Chan, J. Y. C., & Mazzocco, M. M. M. (2017). Competing features influence children's attention to number. *Journal of Experimental Child Psychology*, *156*, 62–81. https://doi.org/10.1016/j.jecp.2016.11.008

Chernyak, N., & Sobel, D. M. (2016). Equal But Not Always Fair: Value-laden Sharing in Preschool-aged Children. *Social Development*, *25*(2), 340–351. https://doi.org/10.1111/sode.12136

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, (41), 1149–1160.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819–824. https://doi.org/10.1016/j.cognition.2008.04.007

Goldin-Meadow, S., & Church, R. B. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*, 43–71. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.715.4570&rep=rep1&type=pdf

Gordon, R., Chernyak, N., & Cordes, S. (2019). Get to the point: Preschoolers' spontaneous gesture use during a cardinality task. *Cognitive Development*, *52*, 100818. https://doi.org/10.1016/j.cogdev.2019.100818

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, *130*, 35–55. https://doi.org/10.1016/j.jecp.2014.09.008

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465. https://doi.org/10.1037/a0012682

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(October), 665–668. https://doi.org/10.1038/nature07246

Hannula-Sormunen, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction*, *15*(3), 237–256.

https://doi.org/10.1016/j.learninstruc.2005.04.005

Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool Children's Spontaneous Focusing on Numerosity, Subitizing, and Counting Skills as Predictors of Their Mathematical Performance Seven Years Later at School. *Mathematical Thinking and Learning*, *17*(2–3), 155–177. https://doi.org/10.1080/10986065.2015.1016814

Hannula-Sormunen, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology*, *107*(4), 394–406. https://doi.org/10.1016/j.jecp.2010.06.004

Hannula, M. M., Rasanen, P., Lehtinen, E., Hannula-Sormunen, M. M., Rasanen, P., & Lehtinen, E. (2007). Development of Counting Skills: Role of Spontaneous Focusing on Numerosity and Subitizing-Based Enumeration. *Mathematical Thinking and Learning*, *9*(1), 51–57. https://doi.org/10.1207/s15327833mtl0901_4

Hurst, M. A., & Cordes, S. (2018). Children's understanding of fraction and decimal symbols and the notation-specific relation to pre-algebra ability. *Journal of Experimental Child Psychology*, *168*, 32–48. https://doi.org/10.1016/j.jecp.2017.12.003

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. https://doi.org/10.1016/j.cognition.2006.10.005

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292–1300. https://doi.org/10.1111/j.1467-7687.2011.01080.x

McMullen, J., Hannula-Sormunen, M. M., & Lehtinen, E. (2015). Preschool spontaneous focusing on numerosity predicts rational number conceptual knowledge 6 years later. *ZDM*

- *Mathematics Education*, *47*(5), 813–824. https://doi.org/10.1007/s11858-015-0669-4

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Nikoloska, A. (2009). Development of the cardinality principle in Macedonian preschool children. *Psihologija*. https://doi.org/10.2298/psi0904459n

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology*, *49*(6), 1103–1112. https://doi.org/10.1037/a0029472

Pinhas, M., Donohue, S. E., Woldorff, M. G., & Brannon, E. M. (2014). Electrophysiological Evidence for the Involvement of the Approximate Number System in Preschoolers' Processing of Spoken Number Words. *Journal of Cognitive Neuroscience*, *26*(9), 1891–1904. https://doi.org/10.1162/jocn_a_00631

Posid, T., & Cordes, S. (2018). How high can you count? Probing the limits of children's counting. *Developmental Psychology*, *54*(5), 875–889. https://doi.org/10.1037/dev0000469

Rathé, S., Torbeyns, J., Hannula-Sormunen, M. M., & Verschaffel, L. (2016). Kindergartners' Spontaneous Focusing on Numerosity in Relation to Their Number-Related Utterances During Numerical Picture Book Reading. *Mathematical Thinking and Learning*, *18*(2), 125–141. https://doi.org/10.1080/10986065.2016.1148531

Uchikawa, K., & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research and Application*, *21*(6), 430–439. https://doi.org/10.1002/(SICI)1520-6378(199612)21:6<430::AID-COL5>3.0.CO;2-X

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

https://doi.org/10.1016/j.cognition.2010.11.014

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Achievement. In *Test*.

Wynn, K. (1992). Children’s acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P