

Research Reports

A New Method for Calculating Individual Subitizing Ranges

Tali Leibovich-Raveh^{*ab}, Daniel Jacob Lewis^a, Saja Al-Rubaiey Kadhim^a, Daniel Ansari^a

[a] Department of Psychology & Brain and Mind Institute, The University of Western Ontario, London, Canada. [b] Department of Mathematics Education, The University of Haifa, Haifa, Israel.

Abstract

A large body of research has shown that human adults are fast and accurate at enumerating arrays of ~1-4 items. This phenomenon has been called subitizing. Above this range, enumeration is slower and less accurate. The subitizing range has been related to individual differences in variables such as mathematical abilities, working memory, etc. The two most common methods for calculating subitizing range today – bilinear fit and sigmoid fit – have their strengths and weaknesses. By combining these two methods, we overcome their biggest limitations and come up with a novel way for calculating Individual Subitizing Range (ISR). This paper introduces this new method as well as empirical studies designed to test the new method. We replicated classic effects from the literature and obtain a high correlation with the sigmoid fit method. This paper includes a Matlab code for easy calculation of ISR as well as a ready-to-use experimental file for testing ISR. We hope that these tools would be of use to researchers studying individual differences in the subitizing range.

Keywords: subitizing, individual difference, numerical cognition, counting, enumeration, bilinear fit, sigmoid fit

Journal of Numerical Cognition, 2018, Vol. 4(2), 429–447, doi:10.5964/jnc.v4i2.74

Received: 2016-08-16. Accepted: 2017-10-24. Published (VoR): 2018-09-07.

*Corresponding author at: The University of Haifa, Education building, Room 641 Haifa, Israel. E-mail: talil@edu.haifa.ac.il



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When asked to enumerate a small number of items (usually up to 4), responses are very fast and accurate. When one plots response times (RT) as a function of the number of the to-be-enumerated items, the increase in RTs as a function of number is found to be very small for up to about 4 items, creating a shallow slope, or even a slope of zero (Simons & Langheinrich, 1982). This phenomenon was named "subitizing" (Kaufman, Lord, Reese, & Volkman, 1949). For numbers greater than ~4, however, the slope becomes steeper. The point in which the slope changes, called the discontinuity point, is thought to mark the end of the subitizing range (Trick & Pylyshyn, 1994). For example, if the slope becomes steeper at the Number 4, the highest number of items that a participant is able to 'subitize', or to enumerate in a fast and accurate manner is 4. Put differently, their individual subitizing range (ISR) can be said to be 4.

While most studies have found that, on average, the subitizing range of adults is between 3-4, there does exist evidence for individual differences in the ISR. More specifically a variety of factors has been found to affect individuals' subitizing range. For example, children with Williams syndrome, a genetic developmental disorder that is associated with severe impairments in visuospatial cognition, were found to have a lower subitizing range than age-matched typically developed children (O'Hearn, Hoffman, & Landau, 2011). Moreover, Green and Bavelier (2003) demonstrated that individuals' subitizing range increases after intense playing of computer

games that requires spatial attention. In addition, Ashkenazi et al (Ashkenazi, Mark-Zigdon, & Henik, 2013) reported that children with Developmental Dyscalculia (DD) – a learning disability specific to math, had a smaller subitizing range than typically developing children.

While there are many studies that attempt to measure differences in subitizing range at the group level, it is difficult to accurately measure the subitizing range of an individual in a way that generalizes well across participants. Evaluating the subitizing range of an individual can be important to study individual differences. For example, comparing a participant's ISR to the group's average, or measuring ISR of an individual in different time points. Moreover, having a measure of individual subitizing range will allow us to evaluate the distribution of subitizing ranges in a typically developed population, and compare this distribution to special groups, such as individuals with dyscalculia. Several methods have been previously employed; we will discuss the two most common methods: bilinear fit (also known as piecewise regression, hockey stick and broken-stick fit) and sigmoid fit.

Bilinear Fit

When plotting RT as a function of the to-be-estimated number, the resulting data can be generally described by two lines with different slopes: a slope of zero – or close to zero – in the subitizing range, and a non-zero positive slope above the subitizing range. These lines are then fit to the data and the intersection point of these two lines is taken to be the upper limit of the subitizing range (Green & Bavelier, 2003).

The basic bilinear fit is essentially a piecewise regression model, which joins two straight lines with an abrupt transition at the intersection point that minimizes the fitting error. More generally, the data is simply segmented into two groups and a line fit to each group of data, with the intersection of those lines being the upper limit of the subitizing range. This method requires the selection of the point at which the data is split along the x-axis into the two groups (subitizing and above) over which the two lines will be fitted (i.e., the split point). In the simplest implementation, the researchers will select the split point themselves. To do that, the researcher needs to first establish a prior assumption for which points lie within the subitizing range. This assumption is refined after fitting the two lines to the groups separated by this split point by then taking the intersection point as the upper bound of the subitizing range.

To determine the split point, various methods are used; some simply eyeball the plot to select that point (Logan & Zbrodoff, 2003), while others compared the slopes of RT over several ranges (e.g., 1-3 vs. 1-4, etc.). For example, if the slope of 1-3 is significantly smaller than the slope of 1-4, then 4 would be the split point that will define the two groups of data used to build linear equations (Allen & McGeorge, 2008). While bi-linear fit can be used to closely describe enumeration task data under some circumstances, it has some drawbacks. One of the problems with this method is that the further the 'true' subitizing upper bound for a given participant lies from the split point, the worse the fit will be, and the less accurate the calculated subitizing upper bound will be. It is also labor intensive in that it requires manual data curation for each participant/group. Alternatively, if the assumption is only made once, or calculated at the group level and then applied to all participants, it may return poor individual fits and 'hide' in-group variability.

There are other more accurate techniques for determining this splitting point for an individual, which do not face these limitations. Since the data range for subitizing calculations is small and data presentation consists of natural numbers only (e.g., 2, 3, 4, etc.), you can simply iterate over each of the possible splitting points between

natural numbers, and for each possible pair of lines, calculate the best fit and the total sum of square errors. For example, you would start at a point “between” 2 and 3, fitting a line for the RT at 1 and 2, and a line for the RT at all values at 3 or larger. You would then do the same for a point between 3 and 4, and so on until the second last point you collected (since you need at least two points to uniquely determine a line). Once this is done, you choose the split point that has produced the smallest sum of square errors for each linear fit. The intersection point for this fit is your calculated upper bound of the subitizing range (Green & Bavelier, 2003; Pagano, Lombardi, & Mazza, 2014). There are other methodological tools for attempting to ascertain the best split point, but they are not commonly used in subitizing range calculations and are thus not discussed here (Toms & Lesperance, 2003).

Bilinear fitting also presents several problems unique to ISR calculations. Due to the small number of RT data points used to fit the linear equations, individual deviations from expected response patterns can introduce large errors in the fit and skew the calculated subitizing range (see Figure 1). Because the bilinear method is very sensitive to these variations, researchers often make several simplifying assumptions to “clean” the data. Such assumptions include limiting or restricting the slope of one or both lines to a predetermined range, enforcing positive slopes, enforcing continuity, truncating the data, or restricting the range over which the break points may vary (Logan & Zbrodoff, 2003; O’Hearn et al., 2016; O’Hearn, Franconeri, Wright, Minshew, & Luna, 2013; Starkey & McCandliss, 2014). Each of these a-priori assumptions bias the resulting calculation of subitizing range.

Bilinear fit is also sensitive to experimental design. A methodology whereby stimuli presentation is time-limited may introduce a “plateau” in RT above the subitizing range when the presented numerosity become sufficiently large, effectively introducing a third linear element into the data which is not accounted for in this method (see Figure 1). This may also occur naturally, as participants may start “guessing” with larger numbers. This shifts the fit of the second line to the right, effectively reducing the slope. Depending on the methodology used to calculate the bilinear fit, this will lead to an overestimation or an underestimation of the subitizing range (Pagano et al., 2014). To account for this, some researchers will exclude these plateau points from the fitting procedure as previously described; this requires either manual curation per participant, or an a-priori range assumption which is applied to all participants regardless of whether it matches the responses of each participant (see Figure 1).

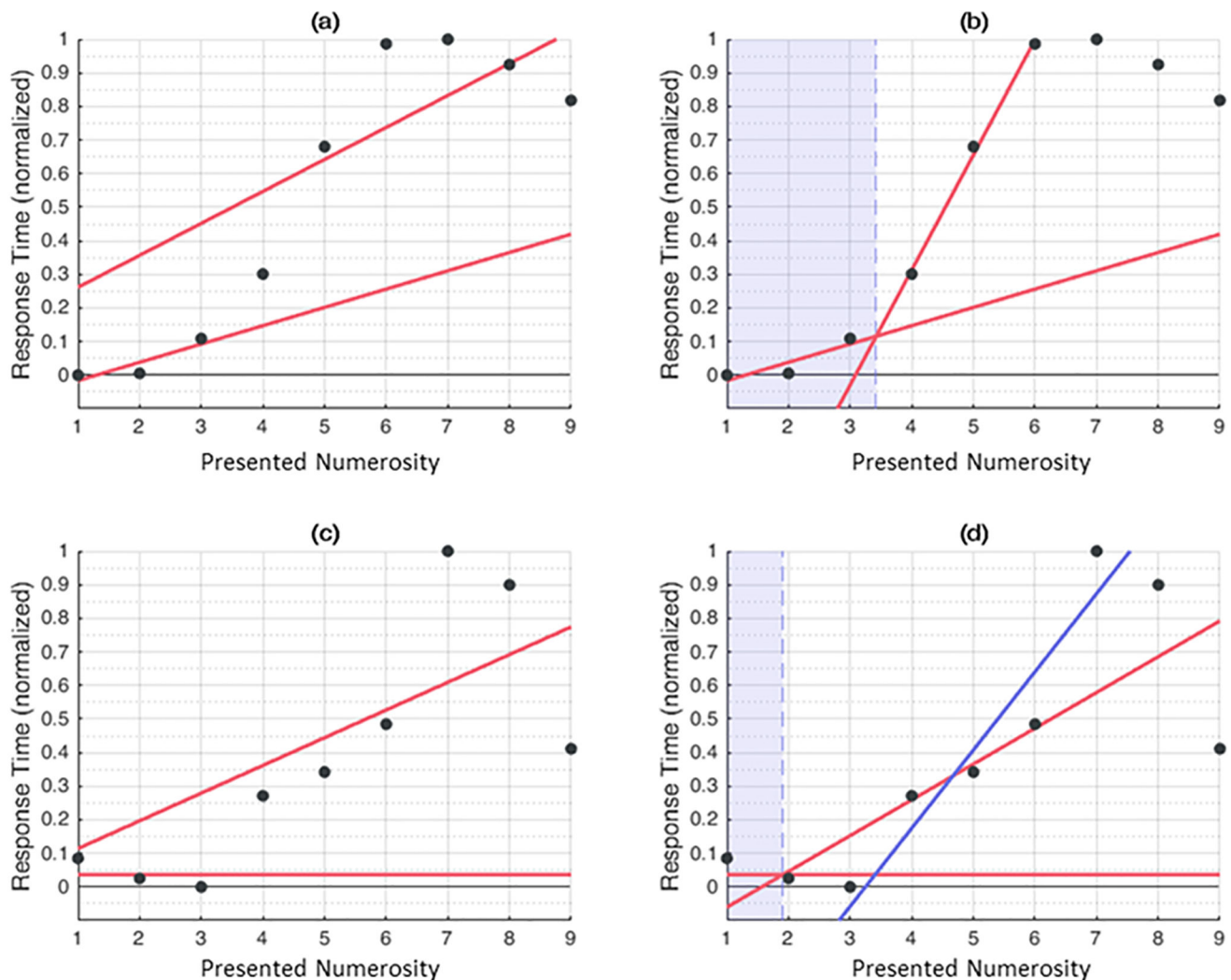


Figure 1. Comparison of bilinear fit on the same data (a-b are from one participant, and c-d are from another with RT normalized) to illustrate the effect of different assumptions and fitting constraints.

Note. (a) A basic bilinear fit with the following assumptions: positive slopes, and with the first line fit to response times from 1-3 and the second line fit to response times of 4 or higher, as is common in the literature. Due to the existence of a plateau in response time for this participant, this results in a very poor fit and no intersection point within the data range. (b) A truncating assumption can be made, removing data higher than 6 from the fit of the second line; this results in a better fit and the intersection point now suggests a subitizing range of ~3.46. Note that this type of assumption requires manual curation of the data, and cannot be assumed to apply to all participants without biasing the results – see (d). (c) A fit for a different participant using the same assumptions as (a) but forcing the first line to a slope of zero to ignore the negative RT drift in the subitizing range. Again, there is no intersection within the data range. (d) Uses the same truncating assumption as (b) but it does not produce the right intersection point (red lines). A new truncating assumption, removing data higher than 7 from the fit to produce the blue line, rather than 6, does find the correct intersection point. This simply illustrates that the a-priori assumptions and restrictions necessary to produce adequate bilinear fits are not generalizable across participants and experiments, and that bilinear fitting methods can be strongly affected by RT plateaus which result from experiment design (e.g. time limited response) or participant strategy.

Sigmoid Fit

As an alternative to the bilinear approach, a sigmoid fit is a very common way to determine the subitizing range (Piazza, Fumarola, Chinello, & Melcher, 2011; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). Sigmoid functions are a family of functions shaped like the letter S (see Figure 2): they start with a slope of zero or close to zero, continue with a steep positive slope and then turn back to zero.

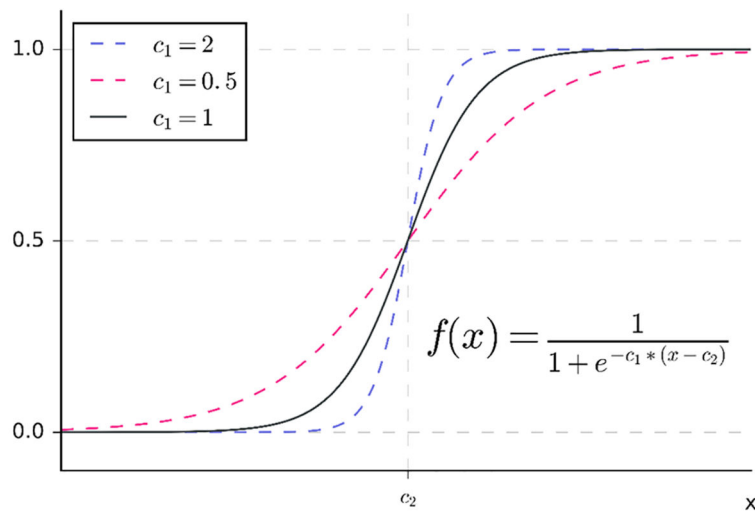


Figure 2. Basic sigmoid function.

Note. A Basic sigmoid function with two parameters (c_1 and c_2) as commonly used for subitizing calculations (Revkin et al., 2008).

Sigmoid function can describe the relationship between RT and the to-be-estimated number in some cases: in the subitizing range, the slope should be zero (or close to zero). Then, when counting is required, the slope becomes positive. This is the relationship described by bilinear fitting. However, when counting is impossible (due to the conditions of the experiment, like very short exposure to the stimuli or a requirement to answer quickly), RTs to different numbers become very similar, resulting in a slope close to zero again. The same is often true for the relationship between error rates and the to-be-estimated numbers. In the subitizing range, there are virtually no errors. During counting, the error rate increases with number, and when estimating, the error rate is similar across quantities, resulting in a slope of zero.

The data for RT or error rates are used to fit a sigmoid function with unknown parameters (see Figure 2) using any of a variety of non-linear fitting algorithms (e.g. the Levenberg-Marquardt algorithm; Marquardt, 1963). The subitizing range is then typically reported as the inflection point of the sigmoid curve (Revkin et al., 2008). This method is more robust than bilinear fitting as it requires no assumptions about the data. However, as mentioned by others (Pagano et al., 2014; Piazza et al., 2011), this method over-estimates the subitizing range. The inflection point (taken as the upper bound of the subitizing range) can be very far from where the slope actually begins to change for the first time (e.g. in Figure 2 compare $c_1 = 2$ with $c_1 = 0.5$ and note that for $c_1 = 0.5$ the inflection point lies much further from the initial region of shallow slope).

While both the basic sigmoid fit and the new combined method will be highly correlated, because a sigmoid curve is fitted in both methods in an identical manner, the basic sigmoid fit will always overestimate the range

due to the inflection point of the curve being taken as the upper ISR limit. This will naturally always lie above the point of first deviation from base RT. Depending on the particular pattern of RT, this gap may be large or small. This will not be a perfectly consistent shift, but because it is always positive, it is still useful for inter-group comparisons and has been used as such, because the delta will still be there. This kind of fit can be useful when comparing individual differences, as long as all the participants are tested using the same method. However, it holds little theoretical meaning on its own, since the reported inflection point does not necessarily accurately reflects the subitizing range.

A New Method for Calculating the Subitizing Range of Individuals

In this paper, we present and empirically test a novel method for calculating individual subitizing range (ISR). This method combines the bilinear fit and the sigmoid fit methods, requiring no a-priori assumptions about the data and correcting for the sigmoid overestimation of subitizing range, thereby managing to overcome the previously discussed limitations of these methods when used separately.

The ISR Algorithm

In this “combination” method, the RT data is first fit with a sigmoid function (a logistic function) as with the basic sigmoid method using a non-linear error minimization algorithm (we used `nlinfit` in MATLAB, which employs the Levenberg–Marquardt algorithm). Instead of using the inflection point of the curve as a proxy for the subitizing range, two lines are fit to the sigmoid curve; one with a slope of zero and an intercept where the sigmoid curve crosses the y-axis at $x = 0$ (i.e., the subitizing line), and a second line which is equivalent to the tangent at the inflection point (see Figure 3).

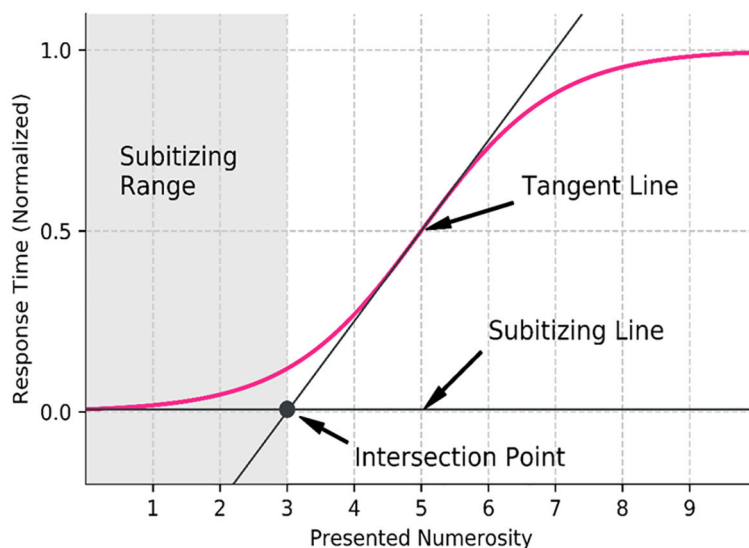


Figure 3. Illustration of the combined sigmoid and bilinear methodology.

Note. The red sigmoid curve is fit to the normalized response time data. A tangent line is then fit to the inflection point of the curve at $x = c_2$ (see Figure 2, $f(x) = 1/(1 + e^{-c_1(x-c_2)})$) and a subitizing line of slope zero intersecting the y-axis at the same point as the sigmoid curve. The intersection point of these two lines defines the upper bound of the subitizing range.

This finds the “elbow point”, or the point at which the slope begins to appreciably change. The intersection point of these two lines is thus a more accurate estimate of the upper bound of the subitizing range; this method corrects for the sigmoid overestimation of subitizing range and avoids the a-priori assumptions necessary to ensure consistent bilinear fitting in a purely bilinear model.

This approach requires no assumptions of data structure and can accurately fit both bilinear data and sigmoid data, thus eliminating the need for a multitude of bilinear approaches with varying methodologies and assumptions, and overcoming the limitations of previous approaches that used a sigmoid function (i.e. using only the midpoint of the function).

The Current Study

In the current study, we aimed to empirically test the suggested method. For this aim, we have conducted several enumeration studies in different conditions and with different populations. In Experiment 1, adult participants were briefly exposed to groups of items arranged in random or canonical patterns (familiar arrangements of items in space, like dots on a game die) – see [Figure 4b](#). In Experiment 2, 4-9-year-old children were briefly exposed to groups of randomly arranged items. In Experiment 3, Adults were exposed to groups of randomly arranged items until they responded. Experiment 3 contained more trials than the other experiments so we can evaluate the stability of the ISR measurement in a half-split reliability analysis.

Experiment 1: Limited Stimuli Presentation – Adults

Method

Preparing the Data

Before calculating ISR, we trimmed outlier trials and prepared the data (Using the ‘Prep_RT.m’ script). The user defines the range beyond which outliers are excluded. We elected to treat every trial that deviated more than 2.5 standard deviations as an outlier. In addition, the ‘Prep_RT.m’ script averages RT across trials for a presented number per participant, preparing the data for the script that computes ISR.

Participants

Thirty-eight participants (25 females), students at the University of Western Ontario, participated in the experiment for course credit. The mean age was 18.26 years ($SD = 0.9$ years). Three participants were left-handed. All participants had normal or corrected-to-normal vision, and no reported learning disabilities or attention deficits. The study was approved by Western University’s Non-Medical Research Ethics Board (NMREB).

Apparatus

The study was designed and executed in OpenSesame ([Mathôt, Schreij, & Theeuwes, 2012](#)) version 3.00, on Windows 7, with a 15” screen, resolution 1024x786 pixels. Vocal responses were recorded using a USB Dynamic Recording Microphone. The same apparatus was used for all experiments.

Stimuli

Groups of 1-9 cartoon frogs (11.43 x 12.7 mm) appeared on a white background. We used images of frogs and a frame with a “beach” theme to make this design suitable for children as well as adults (see [Figure 4](#) for an

example). Each numerosity appeared in two different canonical arrangements and two different random arrangements, creating 36 different trials per block (9 numbers x 2 different canonical arrangements x 2 different random arrangements).

Procedure

Participants were asked to say aloud how many frogs they saw on the screen, as quickly and as accurately as possible. No feedback was given. In order to practice vocal responses, participants were given a practice in which they had to read aloud simple words. This was also used to adjust the sensitivity of the microphone for every participant. The practice discontinued once the experimenter was satisfied with the performance of the participant. In the experimental task, a trial started with a red fixation point presented at the center of the screen for 1,000 ms. Five hundred ms after the elimination of the fixation point, a target appeared for 350 ms, followed by a 100 ms mask. Then, a slide asking “how many?” was presented, and remained until a participant provided a vocal response. Participants were instructed to respond as soon as they can, either when the stimulus was presented or when the stimulus was replaced with the mask or the “how many” slide. RT was measured from the onset of the target stimuli until a vocal response was detected. The “how many?” slide remained until participants responded. Then, the experimenter typed the response on a keyboard. The next trial started 1,000 ms after the response was typed (see Figure 4a). Each participant completed six blocks with short breaks in between. The duration of the task ranged between 15-20 minutes.

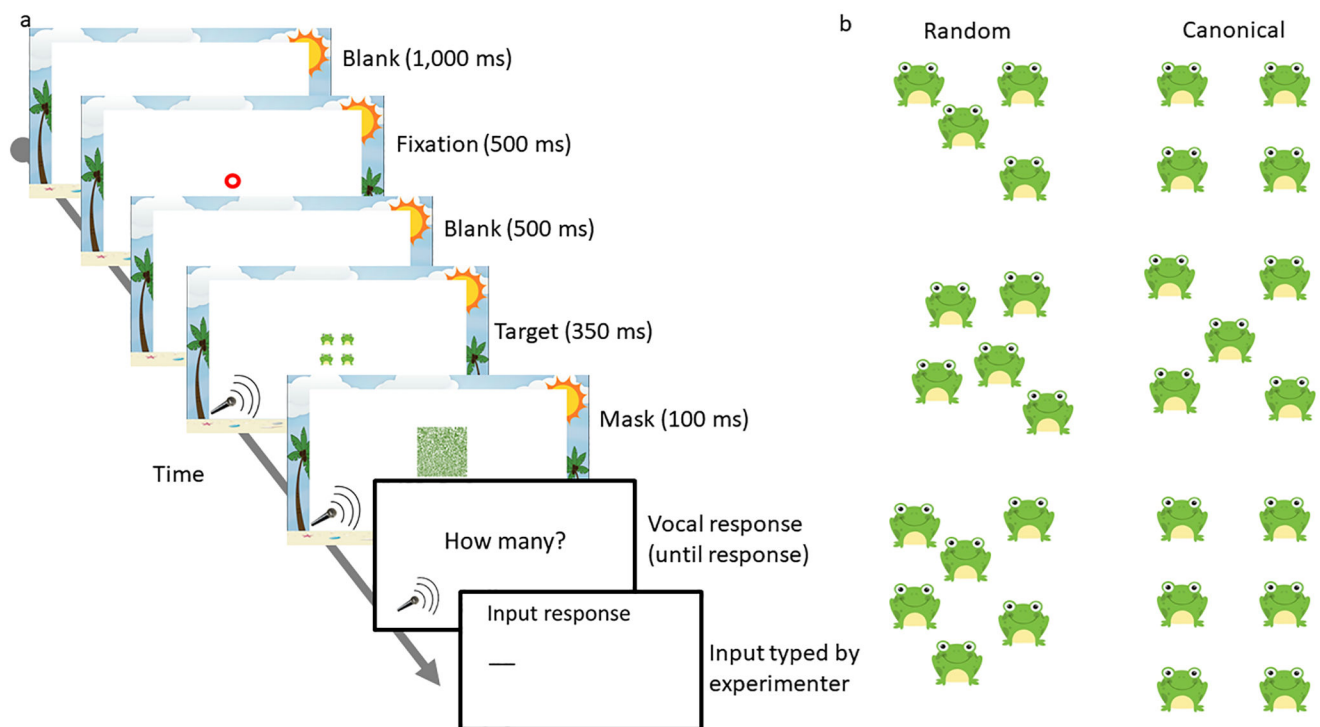


Figure 4. Procedure and stimuli.

Note. a) Experimental procedure. The microphone icon represents the slides in which participants were able to provide vocal response. b) Examples for patterns of spatial arrangements: random or canonical (a familiar arrangement, like on a game die).

Results

We will first describe our preprocessing procedure. Then, we will describe an analysis confirming that our study replicates previous findings. Finally, we will demonstrate how the ISR predicted by our method is related to differences between RT to canonical and random patterns.

Preprocessing

Trials in which participants made unrelated sounds that were registered as a response, (started to say one number and changed it to another, etc.), were excluded. This procedure resulted in the exclusion of 3.1% of the trials across conditions and participants. Only correct trials were used in the analysis. RT to correct trials were trimmed. Trials that deviated more than 2.5 standard deviations from the mean of a specific participant at a specific number and arrangement were removed, resulted in trimming 1.07% of the correct trials of the random condition and 1.49% of the correct trials of the canonical condition across all participants. ISR was calculated separately for the random and the canonical conditions.

Replicating Previous Results: Performance Is Faster and More Accurate in the Canonical Arrangement and in the Subitizing Range

RT and accuracy rates in random vs. canonical arrangements were analyzed to verify that our study replicates previous findings. In this analysis, we included all numbers, including 1 and 9, to get an overview of the pattern. [Figure 5a](#) provide a descriptive overview of the pattern of performance. Namely, RT and error rates increases with number, and responses to canonical arrangements are faster and more accurate than responses to random arrangements. To verify this pattern statistically, we performed two-way analyses of variance (ANOVAs) with range (subitizing: 1-4, counting: 5-9) and pattern (random, canonical) as independent variables and RT or accuracy rates as dependent variables in two separate ANOVAs. These analyses revealed that performance in the subitizing range was faster and more accurate than in the counting range $F(1, 37) = 225.139, p < .001, \eta_p^2 = .86$; $F(1, 37) = 374.38, p < .001, \eta_p^2 = .91$ for RT and accuracy, respectively. Performance was also faster and more accurate for canonical arrangements in comparison to random arrangements $F(1, 37) = 113.02, p < .001, \eta_p^2 = .75$; $F(1, 37) = 760.36, p < .001, \eta_p^2 = .95$ for RT and accuracy, respectively. Finally, we found an interaction between range and arrangement. Namely, the effect of arrangement, although significant across ranges, was stronger in the counting range than in the subitizing range $F(1, 37) = 66.77, p < .001, \eta_p^2 = .64$; $F(1, 37) = 396.91, p < .001, \eta_p^2 = .91$ for RT and accuracy, respectively.

ISR in the Different Spatial Arrangements

ISR was calculated for each participant separately for canonical and random arrangements. For the calculation of ISR, we used quantities 1-8 since in our other Experiments the maximal quantity was 8. In the canonical condition, the mean R^2 value for the fit was 0.83. Out of 38 participants, 4 had an R^2 value of 0.7 or less, which we considered unreliable. The correlation with the sigmoid fit method ([Revkin et al., 2008](#)) was 0.8. In the random condition, the mean R^2 value was 0.89. Two out-of-38 participants had an R^2 value of 0.7 or less. The correlation with sigmoid fit was 0.69. For random patterns, 56% of the participants has ISR of between 3 and 4 (including 4). However, 30% had ISR of 3 or even lower, and 14% had ISR greater than 4. For canonical patterns, the proportions were different. Namely, only 6% of the participants had ISR of 2 or lower, 37% had ISR of between 4-5 and 37% had ISR greater than 5. The actual distribution of ISR per pattern is depicted in [Figure 5b](#).

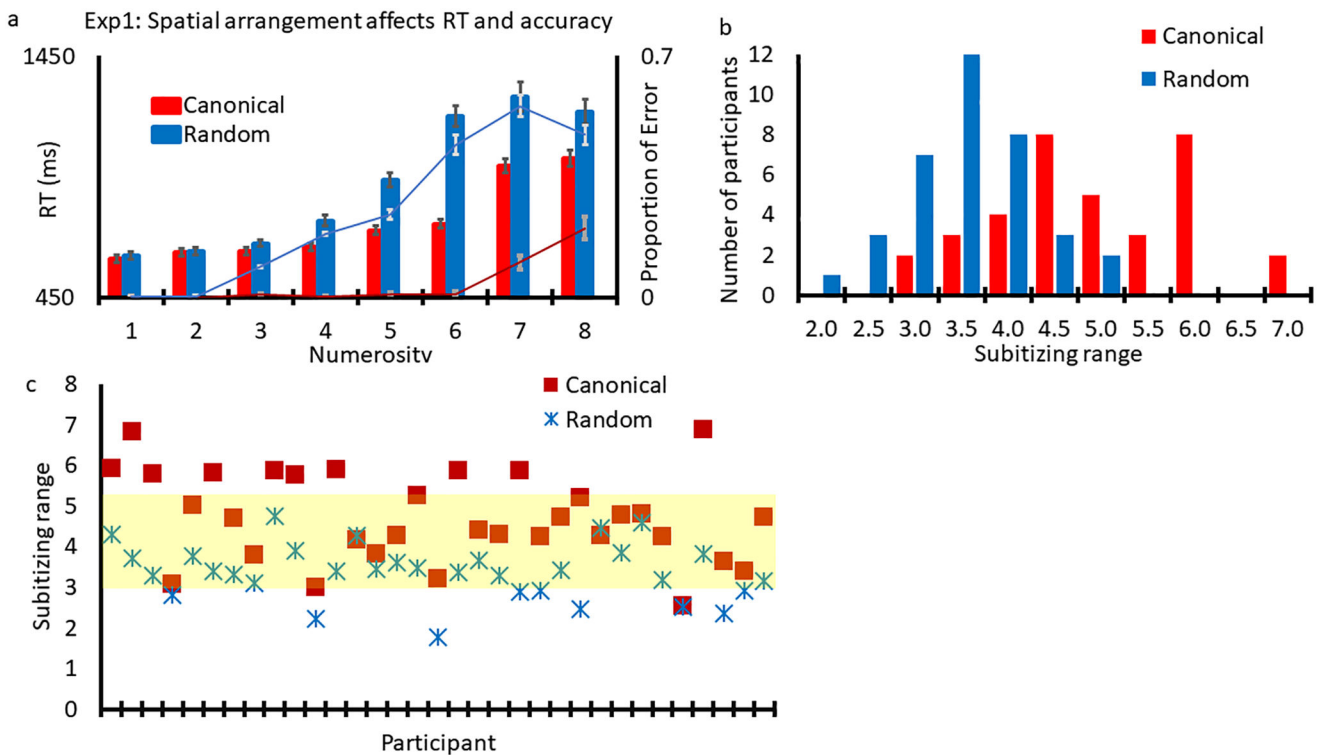


Figure 5. Experiment 1 – Results.

Note. (a) Error rates (in lines) and RT (in bars) as a function of the to-be-enumerated number and spatial arrangement. The white and gray bars represent standard error. (b) Distribution of ISR as a function of spatial arrangement. As expected, the ISR is larger on average for canonical arrangements. (c) Individual differences in ISR in canonical and random arrangements. The yellow area represents the range between 3 and 5.

As expected according to the literature, ISR was higher for canonical arrangements (mean = 4.73) compared with random arrangements (mean = 3.38). t -test confirmed that this difference is significant $t(32) = 8.24$, $p < 0.001$, Cohen's $d = 1.44$. Figure 5c describes the ISR of every participant in the different patterns. For all the participants except for two, ISR increased in the canonical condition. Namely, calculating the ISR using our new method replicated the expected pattern according to the literature (i.e., higher ISR for canonical arrangements).

Experiment 2: Limited Exposure Time – Children

Testing children poses unique challenges: their relatively short attention span limits the number of trials one can use. The data usually contains more noise. In addition, a design requires vocal response is even more problematic: responding vocally is exhausting even for adults, and with children, fatigue means even more noise and a need for even shorter experiment. Since many studies of subitizing are done with typically or atypically developed children and special populations (e.g., Ashkenazi et al., 2013; O'Hearn et al., 2011; Revkin et al., 2008), it was important to test whether the suggested method is suitable for such instances.

Method

Participants

Twenty-four typically developed children (13 females) between the ages of 4 and 9, were recruited from the psychology department's developmental participant pool. The mean age was 7 years ($SD = 1.6$ years). All participants had normal or corrected-to-normal vision, and no reported learning disabilities or attention deficits. Western University's Non-Medical Research Ethics Board (NMREB) approved the study. Three participants (4-5 year-old) were excluded due to low accuracy even in Numbers 1 and 2.

Stimuli

Groups of 1-8 cartoon frogs (11.43 x 12.7 mm) appeared on a white background. We used images of frogs and a frame with a "beach" theme to make this design suitable for children as well as adults (see [Figure 4](#) for an example). The stimuli set included six variation of random patterns for every numerosity. The specific variation was randomly selected during the experiment. The number of times each number appeared was design to gain the most trials in the critical range (i.e., between 2-6, since the subitizing range in children with random patterns was not expected to be greater than 6), and without limiting the range to 6. We divided the experiment into 5 blocks in order to allow frequent breaks. The Numerosity 1 appeared 10 times (twice in each block); Numerosities 2-6 each appeared 30 times (six times per block); Numerosity 7 or 8 appeared in different blocks, intermittently. Namely, in blocks 1,3 and 5 the Number 7 appeared twice and Number 8 did not appear at all, and in Blocks 2 and 4, the Number 8 appeared twice and the Number 7 did not appear at all.

Procedure

The procedure was similar to that of Experiment 1. For 4-5-year-old children the practice included naming pictures instead of reading words.

Results

Preprocessing

5.45% of the trials across participants were excluded for technical reasons, such as the microphone did not pick up the sound of the participant at the first try, a sound in the room was picked up as a response, or the participant started to say one number and continued to another (e.g., fou...five). Outliers (above and below 2.5 standard deviations) were also removed, resulting in exclusion of 2.86% of all correct trials across participants and numerosities.

Descriptive Statistics

To demonstrate the expected pattern of enumeration task, we plotted error rate and RT as a function of numerosity ([Figure 6a](#)). Errors were very rare between 1-3 and then the error rate steeply increased with number. RT gradually increased with a slope that became steeper with number.

Calculating ISR

We calculated ISR according to RT for correct responses only. Four participants had R^2 value lower than 0.7 and they were excluded from further analysis. Mean R^2 value for the rest of the participants was 0.92. The correlation with a sigmoid fit was 0.78. The distribution of ISR values across participants is depicted in [Figure 6b](#). The most common ISR was 4. However, some participants had ISR as low as 2 and 2.5 or high as 5. It can

also be seen in Figure 6b, that 63% of the children had ISR of between 3–4 (including 4), 12% had ISR lower than 3, and 25% had ISR greater than 4. While we do not have enough participants to draw strong conclusions, we can say that in our limited sample, the proportions of ISR range of children looks different from that of adults (Figure 5b).

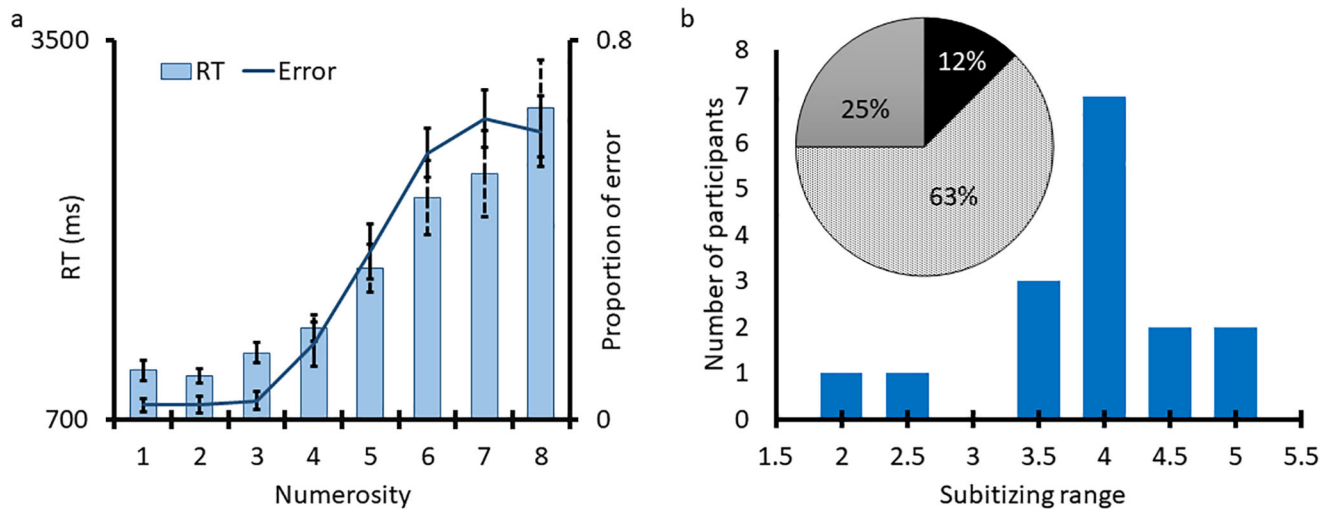


Figure 6. Experiment 2 – Results.

Note. In Experiment 2, the participants were 4–9 year old children. (a) RT and proportion of error as a function of the to-be-enumerated quantity. (b) Distribution of ISR in the sample (bars) and proportion of ISR (pie chart: black: ISR < 3; stripes: 3 <= ISR <= 4; gray: ISR > 4).

Experiment 3: Unlimited Exposure Time – Adults

One critical difference between enumeration tasks in the literature is the time to which the stimulus is presented. In some studies, the stimuli are presented for a limited time. For example, Ashkenazi et al. (2013) limited stimuli presentation time to 200 ms, Green and Bavelier (2003): 50 ms, Wender and Rothkegel (2000): 64 ms. In other studies, stimuli are presented until response is provided (Krajcsi, Szabó, & Mórocz, 2013; Logan & Zbrodoff, 2003; Starkey & McCandliss, 2014). Experiments 1 and 2 used limited presentation time. However, as mentioned in the introduction, both bilinear fit and sigmoid fit are sensitive to differences in presentation times. Therefore, Experiment 3 was conducted with unlimited presentation time to test whether our method will be validated even under this condition. In addition, Experiment 3 contained more trials than Experiments 1 and 2, because we wanted to have enough trials to test for the stability of our method. In other words, we wanted to be able to split the data into two data sets (a half-split test) and check how consistent the ISR scores will be.

Method

Participants

Thirty-one participants (26 females), students at the University of Western Ontario, participated in the experiment for course credit. The mean age was 18.19 years ($SD = 1.49$ years). Two participants were left-handed. All participants had normal or corrected-to-normal vision, and no reported learning disabilities or attention deficits. Western University's Non-Medical Research Ethics Board (NMREB) approved the study.

Stimuli

The stimuli that were used for Experiment 2 were used also for this experiment.

Procedure

Participants were asked to say aloud how many frogs they saw on the screen, as quickly and as accurately as possible. We emphasized that being fast is equally important to being accurate. No feedback was given. As described for the previous studies, a word-reading practice was given before the experimental task. In the experimental task, a trial started with a red fixation point presented at the center of the screen for 1,000 ms. Five hundred ms after the elimination of the fixation point, a target appeared until participant made a response. Then, the experimenter typed the response on a keyboard. RT was measured from the onset of the target stimuli until a vocal response was detected. The next trial started 1,000 ms after the response was typed (see Figure 7). Each participant completed six blocks (204 trials in total) with short breaks in between. The duration of the task ranged between 15-20 minutes.

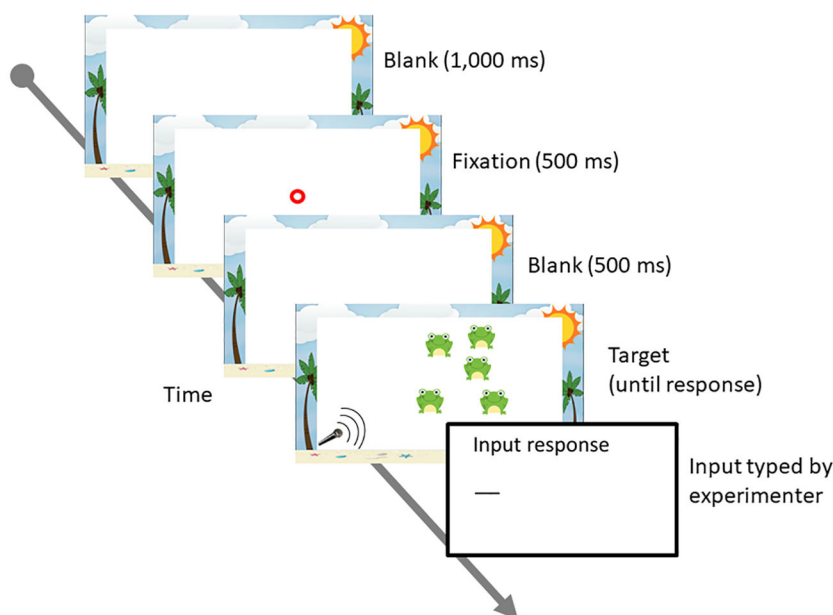


Figure 7. Procedure of Experiment 3: Unlimited presentation time.

Results

Preprocessing

Due to technical problems, 2.95% of the trials across participants were removed. Outliers (above and below 2.5 standard deviations) were also removed, resulting in exclusion of 2.26% of all correct trials across participants and conditions.

Descriptive Statistics

As seen in Figure 8a, the slope describing error rates or RT as a function of the to-be-enumerated numerosities were shallow until around 4, and became steeper for larger quantities.

ISR Calculation

ISR was calculated based on RT for correct responses only. All participants had R^2 value higher than 0.7. The mean R^2 value was 0.96. The correlation with sigmoid fit was 0.75. The distribution of ISR in our sample is described in [Figure 8b](#); while most participants range between 4-5, some participants had lower (3-3.5) or higher (5-5.6) ISR. 56% of the participants had ISR of between 4-5. 34% had ISR of 5 or higher, and 13% had ISR of 3 or lower. Again, with caution due to our relatively small sample size, the proportion of ISR looks different from the ISR with adults who viewed the stimuli for limited duration and encouraged to respond fast.

Reliability Analysis

The current Experiment had more trials than the other experiments. This enabled us to test how consistent the ISR scores obtained by our methods are. For this aim, we first divided the data from Experiment 3 into two halves, by separating odd and even trials. Namely, trials Number 1,3,5 etc. of each participant were part of the “odd trials” data set, and trials Number 2,4,6 etc. were part of the “even trials” data set. For each data set, we calculated ISR. Then we correlated the ISRs of the two data sets. This procedure yielded a Pearson value of 0.54 ($p = 0.002$). To compensate for the relatively short length of each half, we applied a Spearman-Brown correction for split-half analysis ([Cortina, 1993](#)). This yielded a reliability estimate of 0.69. [Figure 8c](#) describes the changes in ISR for each participant: The markers (in different shapes and colors) indicate the ISR that was calculated with all the trials. The black line represents the minimal and maximal ISR that was calculated on half the trials. The different marker colors and shapes represent the difference in ISR: the ISR of six participants (orange rhombus) differ in more than one units between the two halves of the experiment. For the rest of the participant (green circle marker) the change was of less than one unit.

Comparing the new method to a sigmoid fit method, our method applies the same sigmoid fit as reported in previous studies ([Piazza et al., 2011](#); [Revkin et al., 2008](#)). However, we have established the subitizing range using different criteria; the studies above set the subitizing range to be at the inflection point of the sigmoid curve. Our method sets the subitizing range to be in the intersect of two lines: the line where the sigmoid curve crosses the y-axis and the line that is the tangent at the inflection point intersect ([Figure 3](#)). Hence, the two methods are expected to yield different means that are highly correlated. Put differently, the ISR estimates are expected to correlate because both methods use the same sigmoid fit. However, we expect that the mean ISR estimates obtained by the inflection point method will be significantly higher than the ISR obtained by our method.

To test this expected pattern, we have calculated for every participant in Experiment 3 the ISR twice: once according to our method and once according to the inflection point. The results were as expected; the two ISR estimates highly correlated (Pearson's $r = 0.75$, $p < 0.001$); the mean ISR calculated by our method was 4.22, while the mean ISR calculated by the inflection point was 5.8. This difference was found to be significant in a t -test for paired samples $t(30) = 21.3$, $p < 0.001$. This confirmed that our method corrects the over-estimation occurs while calculating the inflection point ([Pagano et al., 2014](#); [Piazza et al., 2011](#)) and more closely reflects the theoretical definition of the subitizing range – the point where the slope first becomes steeper.

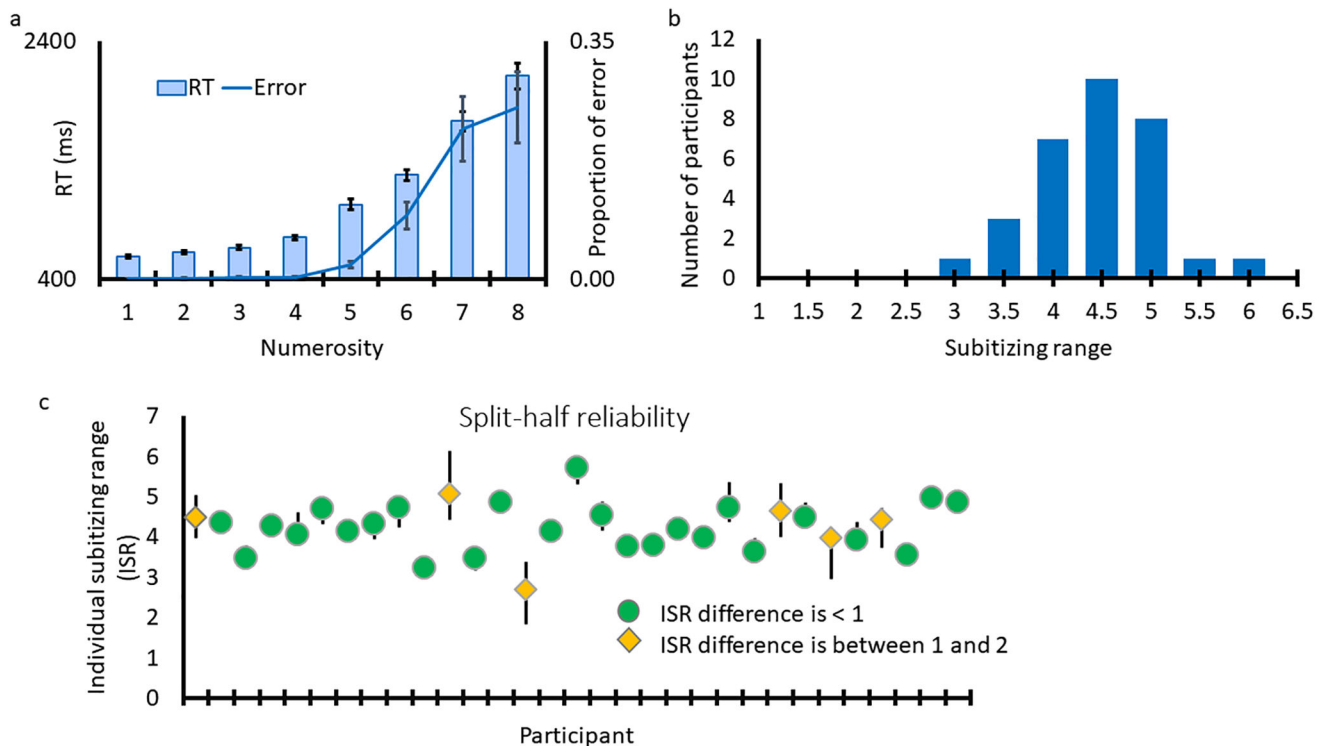


Figure 8. Experiment 3 – Results.

Note. (a) Proportion of error and RT as a function of the to-be-enumerated numerosities. (b) Distribution of ISR in the Experiment's sample (bars). (c) Stability of ISR in the two halves of the experiment. The position of the markers describe the ISR value calculated for all the trials. The black bar describes the change in ISR when calculated on half of the trials. The different colors and shapes of the markers represent the difference between the ISR calculated in the first and second half of the data set. Green circles represent a difference of less than 1 in the ISR between the two halves of the experiment, and an orange rhombus indicate ISR difference of between 1 to 2 between the two halves of the experiment.

General Discussion

In this paper, we present a novel method for calculating subitizing range at a single-subject level. The method combines sigmoid fit with bilinear fit, thereby overcoming some of the limitations these methods present separately. There is no need to assume an initial subitizing range or a normal distribution of data, no need to limit or truncate the range for fitting, the fit is less affected by outliers, and the ISR value does not over-estimate the subitizing range.

To validate this method, we used different common experimental designs and multiple populations. We replicated the effect of spatial arrangement: ISR was higher for canonical than for random arrangements. We also demonstrated that the correlation of ISR values using a traditional sigmoid fit and our new method is relatively high (between .7 and .83) in different populations (children and adults) and in different experimental designs (limited/unlimited stimuli presentation time and random/canonical arrangements).

Importantly, we used the exact same fit to sigmoid curve as previous studies (Piazza et al., 2011; Revkin et al., 2008). However, by using a different criterion for finding the subitizing range on that curve, our results reflect more accurately the theoretical definition of the subitizing range, namely, when the slope first increase.

Finally, we also conducted a reliability analysis to demonstrate that the ISR produces reliable results even with a procedure that is expected to take less than 30 minutes to complete. This is especially important when the ISR is tested with children or other populations with short attention span. The task used in this study is available online in a children-friendly format.

Variability of ISR in Typically Developed Population

The current method allows to evaluate the subitizing range of individuals and to compare proportions of ISRs amongst different populations. Usually the subitizing range is reported to be between 3 to 4 or between 3 to 5. Our sample of typically developed adults and children, although small, suggests that there is variability in the subitizing range even in the general population. This variability depends not only on the population but also on the conditions of the task. For example, both Experiments 1 and 3 tested typically developed adults (university students). However, in Experiment 1, the stimuli were presented for limited time and participants were encouraged to respond quickly. In Experiment 3, the stimuli remained on screen until participants responded, and speed was as important as accuracy. In these two experiments, the proportion of different ISR seems different: the proportion of ISR greater than 4 was higher when stimuli were presented until response. This finding needs to be taken with caution given our relatively small sample size. Future studies that will employ our method of analysis in different conditions with larger sample size will be able to better evaluate how much the ISR is affected by different experimental conditions.

In addition, there is also variability in how much spatial arrangement affects the subitizing range in an individual (Figure 5c). Such differences raise interesting questions regarding the variability of subitizing and its relation to math abilities, spatial abilities, visuospatial working memory, or any of the different cognitive mechanisms (Ashkenazi et al., 2013; Green & Bavelier, 2003; O'Hearn et al., 2013) in the general population. The current method, allowing us to estimate an individual's subitizing range can go a long way in helping to find connections between individual differences in subitizing and individual differences in a variety of cognitive abilities.

ISR Values Estimate the Likelihood of a Quantity to Be 'Subitized'

The ISR values obtained here and with other methods (sigmoid and bi-linear fits) are not necessary whole numbers. What does ISR values represent? Is it accurate to say that an individual with a subitizing range of 4 will always, under various conditions, subitize 1-4 items and count more than 4 items? If it so, then ISR values that are not whole numbers are meaningless. We, however, think that such account for the meaning of ISR is less likely, given the multiple factors influencing enumeration. One such factor, as discussed above, is the time one has to see the stimuli. For example, compare trying to enumerate the number of legs of a housefly, flying very fast, vs. enumerating the number of petals of a flower you are holding. Even if both quantities are 4, it could be that with the fly, you will subitize (and maybe fail to reach a correct answer), while with the flower you will count, just because you have enough time and you want to be accurate.

Another example for a factor that highly affect enumeration is the spatial arrangement of the to-be-enumerated items. A canonical pattern of 5 items will (most likely) be 'subitized' while a random pattern of 5 will be (most likely) counted. However, one must keep in mind that not all patterns are completely canonical or completely random. In other words, some random patterns may be less or more familiar than other random patterns. Some studies have demonstrated this point. For example, Krajcsi et al. (2013) used stimuli that were canonical but with different amount of 'noise' (e.g., a shift of one dot in a canonical pattern of 6 dots arranged in two vertical

lines of 3 dots) and demonstrated that performance was influenced by the different 'noise' level of the patterns. Similarly, in our experiment, as in everyday lives, participants have encountered patterns that, although we have considered random, could have been more familiar or less familiar to them. Therefore, it is possible that, for instance, some stimuli of Numerosity 5 were subitized while other stimuli containing the same numerosity were counted.

Given the multiple factors that influence enumeration, we suggest that the ISR is not the actual subitizing range, but the *likelihood* that a quantity will be subitized. Considering this view, we would like to argue that ISR of 3.5, for example, suggests that theoretically, out of 100 experimental trials, Numerosity 3 will be subitized in 50/100 trials and counted in the other 50, so on average the ISR would be 3.5. Another way of looking at such values suggests that a person with ISR of 4.3 is more likely to subitize an array of 4 items than a person with ISR of 3.9. Such differences can be important when testing for individual differences in the subitizing range and correlating them with other abilities. Since math scores or spatial abilities scores will usually have more variability than ISR, having an ISR that is not only a whole number can make the variability of the to-be-compared measure more even, and reflect better the ISR distribution in the sample.

Strengths and Limitations

Our new approach combines the advantages of bilinear regression with a non-linear regression sigmoid fit to take advantage of the more robust nature of sigmoid fitting while correcting for the overestimation in subitizing range. There is now no need to make any prior assumptions about where the subitizing range may be, no data needs to be truncated prior to fitting, and the approach is more robust to experimental variations (e.g. limited presentation time or an emphasis on speed) and to participant strategies, as well as being less affected by outliers in RT. The approach has been codified into easy-to-use Matlab scripts that can be run on both Windows and Mac systems and is generalizable across participants and experiments. The approach has been validated in several experiments with different populations (see results), replicating classic results from literature.

A limitation of the method we report here is that it can only be used with response time but not accuracy data. Especially with adults, and in experiments where presentation time is not restricted or limited, some participants obtain consistently high accuracy measures, thus presenting no identifiable transition outside of the subitizing range. However, this limitation is the same for sigmoid fitting and bilinear fitting. We recommend using RT data for this reason, as it is more suitable to measure the difference between the subitizing and counting range across participants and experimental designs.

Conclusions

This study presented a new method and tools for calculating individual subitizing range and validated the method by demonstrating its correlation with sigmoid fit across different populations and experimental procedures. The experimental files, and scripts, as well as readme files are all available online at <https://osf.io/5cp6r/>. We hope that by using the ISR calculator and the suggested experiment, more insights into the factors influencing the subitizing range could be gained.

Funding

The authors have no funding to report.

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

The authors have no support to report.

Supplementary Materials

We make available the experiment file used to run the experiments (in the free software “Open Sesame”), a ready-to-use set of MATLAB scripts, with instructions and details about the process of the calculation.

Index of Supplementary Materials

Leibovich-Raveh, T., Lewis, D. J., & Ansari, D. (2018). *Individual Subitizing Range (ISR)* [Supplementary materials]. doi:[10.17605/osf.io/5cp6r](https://doi.org/10.17605/osf.io/5cp6r)

References

- Allen, R., & McGeorge, P. (2008). Enumeration: Shape information and expertise. *Acta Psychologica*, 129(1), 26-31. doi:[10.1016/j.actpsy.2008.04.003](https://doi.org/10.1016/j.actpsy.2008.04.003)
- Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2013). Do subitizing deficits in developmental dyscalculia involve pattern recognition weakness? *Developmental Science*, 16(1), 35-46. doi:[10.1111/j.1467-7687.2012.01190.x](https://doi.org/10.1111/j.1467-7687.2012.01190.x)
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. doi:[10.1037/0021-9010.78.1.98](https://doi.org/10.1037/0021-9010.78.1.98)
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534-537. doi:[10.1038/nature01647](https://doi.org/10.1038/nature01647)
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, 62(4), 498-525. doi:[10.2307/1418556](https://doi.org/10.2307/1418556)
- Krajcsi, A., Szabó, E., & Mórocz, I. Á. (2013). Subitizing is sensitive to the arrangement of objects. *Experimental Psychology*, 60(4), 227-234. doi:[10.1027/1618-3169/a000191](https://doi.org/10.1027/1618-3169/a000191)
- Logan, G. D., & Zbrodoff, N. J. (2003). Subitizing and similarity: Toward a pattern-matching theory of enumeration. *Psychonomic Bulletin & Review*, 10(3), 676-682. doi:[10.3758/BF03196531](https://doi.org/10.3758/BF03196531)
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431-441. doi:[10.1137/0111030](https://doi.org/10.1137/0111030)
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314-324. doi:[10.3758/s13428-011-0168-7](https://doi.org/10.3758/s13428-011-0168-7)
- O'Hearn, K., Franconeri, S., Wright, C., Minshew, N., & Luna, B. (2013). The development of individuation in autism. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 494-509. doi:[10.1037/a0029400](https://doi.org/10.1037/a0029400)

- O'Hearn, K., Hoffman, J. E., & Landau, B. (2011). Small subitizing range in people with Williams syndrome. *Visual Cognition*, 19(3), 289-312. doi:[10.1080/13506285.2010.535994](https://doi.org/10.1080/13506285.2010.535994)
- O'Hearn, K., Velanova, K., Lynn, A., Wright, C., Hallquist, M., Minshew, N., & Luna, B. (2016). Abnormalities in brain systems supporting individuation and enumeration in autism. *Autism Research: Official Journal of the International Society for Autism Research*, 9, 82-96. doi:[10.1002/aur.1498](https://doi.org/10.1002/aur.1498)
- Pagano, S., Lombardi, L., & Mazza, V. (2014). Brain dynamics of attention and working memory engagement in subitizing. *Brain Research*, 1543, 244-252. doi:[10.1016/j.brainres.2013.11.025](https://doi.org/10.1016/j.brainres.2013.11.025)
- Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, 121(1), 147-153. doi:[10.1016/j.cognition.2011.05.007](https://doi.org/10.1016/j.cognition.2011.05.007)
- Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, 19(6), 607-614. doi:[10.1111/j.1467-9280.2008.02130.x](https://doi.org/10.1111/j.1467-9280.2008.02130.x)
- Simons, D., & Langheinrich, D. (1982). What is magic about the magical number four? *Psychological Research*, 44(3), 283-294. doi:[10.1007/BF00308426](https://doi.org/10.1007/BF00308426)
- Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology*, 126, 120-137. doi:[10.1016/j.jecp.2014.03.006](https://doi.org/10.1016/j.jecp.2014.03.006)
- Toms, J. D., & Lesperance, M. L. (2003). Piecewise regression: A tool for identifying ecological thresholds. *Ecology*, 84(8), 2034-2041. doi:[10.1890/02-0472](https://doi.org/10.1890/02-0472)
- Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, 101(1), 80-102. doi:[10.1037/0033-295X.101.1.80](https://doi.org/10.1037/0033-295X.101.1.80)
- Wender, K. F., & Rothkegel, R. (2000). Subitizing and its subprocesses. *Psychological Research*, 64(2), 81-92. doi:[10.1007/s004260000021](https://doi.org/10.1007/s004260000021)