Research Article

Psychological Science
1–11

DOI: 10.1177/09567976211034502
www.psychologicalscience.org/PS

Exact Number Concepts Are Limited to the Verbal Count Range

Benjamin Pitt[1], Edward Gibson[2], and Steven T. Piantadosi[1]
[1]Department of Psychology, University of California, Berkeley, and [2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Abstract
Previous findings suggest that mentally representing exact numbers larger than four depends on a verbal count routine (e.g., "one, two, three . . ."). However, these findings are controversial because they rely on comparisons across radically different languages and cultures. We tested the role of language in number concepts within a single population—the Tsimane' of Bolivia—in which knowledge of number words varies across individual adults. We used a novel data-analysis model to quantify the point at which participants ($N = 30$) switched from exact to approximate number representations during a simple numerical matching task. The results show that these behavioral switch points were bounded by participants' verbal count ranges; their representations of exact cardinalities were limited to the number words they knew. Beyond that range, they resorted to numerical approximation. These results resolve competing accounts of previous findings and provide unambiguous evidence that large exact number concepts are enabled by language.

Keywords
cognition, cross-cultural differences, language, number comprehension, open data

Received 2/16/21; Revision accepted 6/25/21

Language gives humans extraordinary cognitive abilities, but its role in numerical cognition remains unresolved. Studies of human infants and nonhuman animals have shown that at least some numerical abilities do not depend on language. Babies, monkeys, and even invertebrates can make precise distinctions between small quantities without counting (up to about four; Feigenson et al., 2004; Pahl et al., 2013) and can rapidly distinguish the numerosities of larger sets, although only roughly (Cheyette & Piantadosi, 2020; Dehaene, 1997; Dehaene et al., 1998). Whereas the ability to represent small exact and large approximate numbers is consistent across species, the ability to represent larger numbers exactly (e.g., exactly seven) appears to be unique to humans (Dehaene, 1997; cf. Brannon, 2005) and is often attributed to language (Bloom, 1994; Carey & Barner, 2019; Chomsky, 1988). Specifically, predominant accounts posit that the structure of the verbal count list (e.g., "one, two, three . . ."), which children learn to recite long before they understand the meanings of the number words (Davidson et al., 2012; Sarnecka et al., 2015; Wynn, 1992), allows them to discover the logic of numbers by induction (Bloom, 1994; Carey, 2004, 2009; Piantadosi et al., 2012; Spelke, 2003; cf. Butterworth et al., 2008; R. Gelman & Gallistel, 2004; Leslie et al., 2008).

This account draws support from studies of isolated groups with few or no words for exact quantities. Specifically, two indigenous groups in the Brazilian Amazon—the Pirahã and the Mundurukú—have no words denoting large exact quantities (and in the case of the Pirahã, no words for any exact quantity, not even "one"; Frank et al., 2008; Pica et al., 2004). To test large exact number concepts in such groups without using number words, researchers have used simple numerical

Corresponding Author:
Benjamin Pitt, University of California, Berkeley, Department of Psychology
Email: pitt@berkeley.edu

tasks that require only behavioral responses, often on sets of physical objects (e.g., seven pebbles; Fig. 1). Pirahã and Mundurukú adults perform well on these tasks only up to about four objects; for larger cardinalities, they are unable to reproduce the number of objects in a set exactly, relying instead on approximation (Frank et al., 2008; Gordon, 2004; Pica et al., 2004). A similar pattern has been found in Nicaraguan Homesigners, a group of congenitally deaf adults whose language lacks a count routine (Spaepen et al., 2011). In sum, people without words for large exact numbers seem unable to represent cardinalities larger than four, leading some scholars to conclude that the verbal count list "enable[s] exact enumeration" (Gordon, 2004, p. 498).

However, these findings are difficult to interpret (e.g., Butterworth et al., 2008; Gleitman & Papafragou, 2013), in part because they rely on comparisons across languages and cultures. Groups without exact number words (such as the Pirahã) are compared, if only implicitly, with groups that have productive counting systems (such as U.S. Americans). Of course, isolated groups differ radically from Western, educated, industrialized, rich, and democratic (WEIRD; Henrich et al., 2010) groups in many ways besides their knowledge of number words (e.g., Everett, 2009), and any of these differences could account for the observed difference in numerical cognition (R. Gelman & Butterworth, 2005; Spaepen et al., 2011). For example, some scholars suggest that the Pirahã failed to make exact numerical matches of large sets because they were simply "indifferent to exact numerical equality" (R. Gelman & Gallistel, 2004, p. 442; also see Laurence & Margolis, 2007; Spaepen et al., 2011), perhaps because "keeping track of large exact quantities is not critical for getting along in Pirahã society" (Casasanto, 2005, p. 1721). Indeed, whereas quantification is prized in WEIRD cultures, unindustrialized groups such as the Pirahã do not track chronological age, use currency, or have units of measurement (Cooperrider & Gentner, 2019; Diekmann et al., 2017; Everett, 2009). Cross-cultural comparisons cannot in principle distinguish whether large exact number concepts depend on a verbal count routine or on other aspects of language and culture.

Even if these studies had clearly established a causal role for language in large exact number concepts, it would remain unclear what role that is. Some accounts posit that the verbal count list is instrumental both for inducing the principles of number (e.g., Hume's principle; Carey & Barner, 2019; Schneider & Barner, 2020) and for using those principles to construct representations of specific cardinalities (e.g., exactly seven; Carey, 2004). Alternatively, the verbal count list may be necessary for inducing the logic of number only, which people could then use to enumerate large sets whether or not the corresponding verbal symbols were available to them. Previous cross-cultural studies were not able to distinguish between these possibilities because they tested numerical abilities only at the extremes; the numerical abilities of the Pirahã (or any group without large exact number words) could reflect a lack of the requisite number principles, number words, or both.

Some studies have tested the role of number words in large exact number concepts without comparing language groups but have found mixed results. In a group of undergraduate students, verbal interference impaired performance on some numerical tasks more than on a spatial control task, suggesting a functional role for language in representing large exact numbers (Frank et al., 2012). However, despite verbal interference, participants performed well on two other tests of large exact number representations, including the orthogonal-matching task, complicating interpretation of the results. In another study, U.S. children overwhelmingly failed to make exact numerical matches of large sets, but they were also imprecise in a task that required only one-to-one matching of objects (Schneider & Barner, 2020). In sum, findings of previous studies do not clearly establish whether or how large exact number concepts are shaped by language.

Here, we addressed these inferential challenges by testing the relationship between number words and

Statement of Relevance

Animal species share some basic numerical abilities, but only humans can reason about exact numbers such as seven or 42. What allows us to accomplish this extraordinary cognitive feat? Here, we tested the role of language in number concepts among the Tsimane', an indigenous group whose adults vary in their knowledge of number words. We found that this linguistic knowledge placed an upper bound on participants' ability to mentally represent exact quantities; participants correctly matched the number of objects in (unlabeled) sets only when that number was within their highest verbal count. For larger sets, they had no way of representing exactly how many objects they saw. This finding provides the clearest evidence to date that number words play a functional role in people's ability to represent exact quantities larger than four and supports the broader claim that language can enable new conceptual abilities.

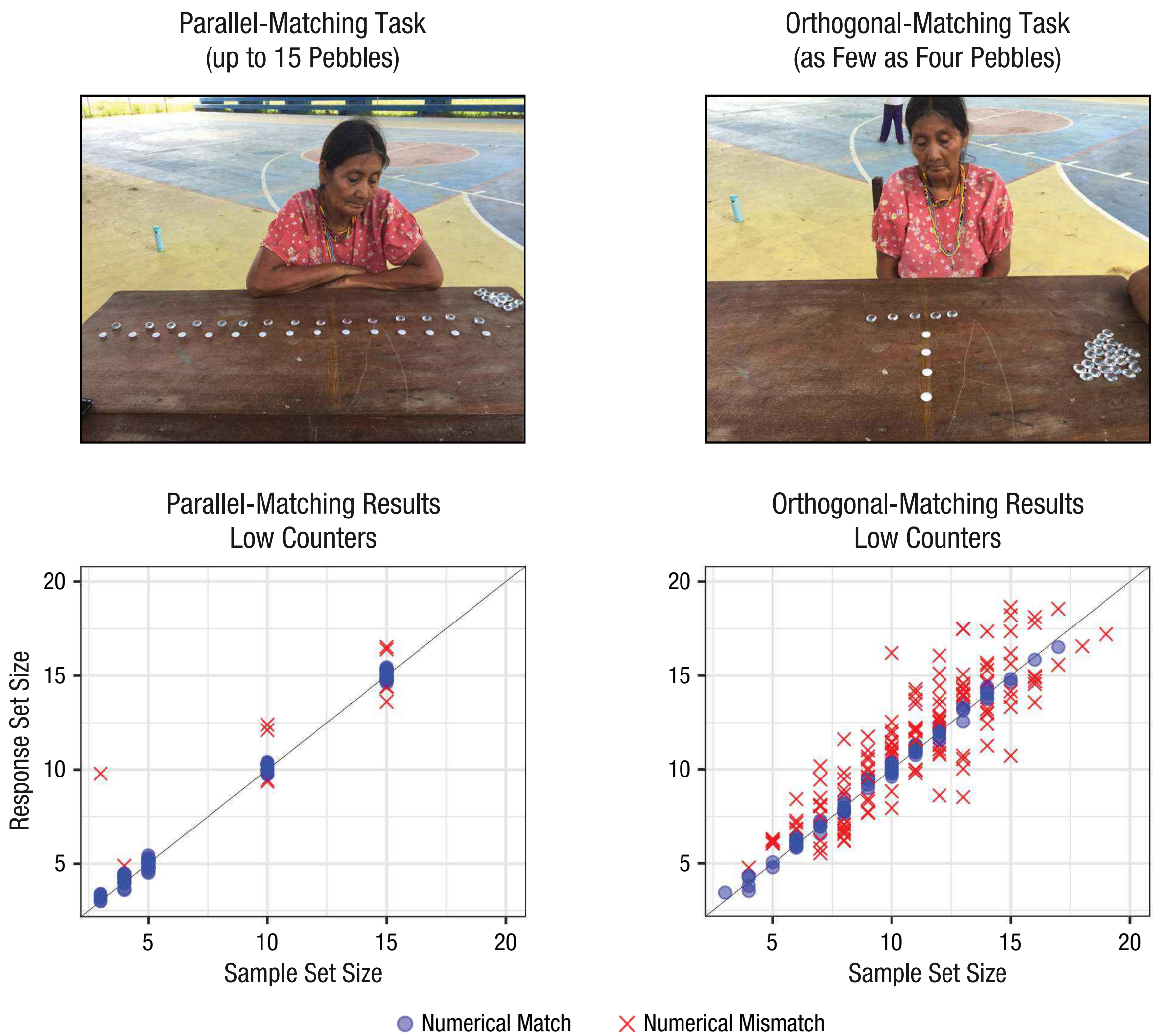

Fig. 1. Paradigm and results from the matching tasks. In the parallel-matching task (top left), the experimenter presented a sample array of objects (sets of 3, 4, 5, 10, and 15 white pebbles) in a lateral line, and participants arranged response items (glass pebbles) parallel to each item in the sample array. This task required participants to use one-to-one correspondence to numerically match the number of objects in the sample array. In the orthogonal-matching task (top right), the sample array was arranged in a line extending away from the participant, and participants arranged their response arrays in a line that was orthogonal to the sample array. Correct matching required participants to represent the cardinality of the sets (of 4 to 25 white pebbles). The graphs show the number of numerical matches and mismatches made by low counters to each sample array in each task.

number concepts in the Tsimane', a group of unindustrialized farmer-foragers indigenous to the Bolivian Amazon (Huanca, 2008). Unlike other isolated populations, the Tsimane' have a fully productive system of number words in their language. Yet unlike adults in WEIRD cultures, Tsimane' adults exhibit considerable variation in their knowledge of the verbal count list; many Tsimane' adults can count indefinitely, but some do not know words above 10, others falter at 12, and so on. This variability allowed us (a) to compare verbal and numerical abilities across individuals, rather than across groups, and (b) to do so at many intermediate levels, not just at the extremes. To determine which large numbers participants could represent exactly, and which numbers they could only approximate, we used a novel statistical analysis to model participants' behavioral responses in a numerical matching task. This model used the known psychophysical properties of numerical estimation to determine the set size at which participants switched from exact to approximate number representations. By comparing this switch point with participants' highest verbal counts, we tested whether people need a system of number symbols (such as those in the verbal count list) in order to represent large exact numbers. If they do (Carey, 2009; Spelke, 2003), then we should find not only that these abilities are correlated but also that one systematically exceeds the other; participants' highest verbal counts should place an upper bound on their numerical representations. Alternatively, if number words are necessary for discovering the logic of number but not for deploying it (or not at all; e.g., Butterworth et al., 2008; Leslie et al., 2008), then participants' numerical representations should sometimes exceed their verbal count ranges. Unlike in previous studies, here the relationship between verbal counting and numerical reproduction cannot be attributed to broad cultural or linguistic differences because our participants shared the same culture and language and, in many cases, lived in the same small community.

Method

Participants

All participants gave verbal informed consent before participating in the study, which was approved by the institutional review board at University of California, Berkeley. As part of an initial questionnaire, participants were asked to count aloud as high as they were able, starting at one, in whatever language they preferred (i.e., Tsimane' or Spanish). Participants who faltered in their count routine for numbers below 20 were selected for the low-counter group (n = 15; mean age = 48.73 years, SEM = 4.34, mean schooling = 0.2 years, SEM = 0.11), and those who counted past 40 were selected for the control group of high counters (n = 15; mean age = 32.87 years, SEM = 4.52, mean schooling = 4.00 years, SEM = 0.65). The sample size was determined by the number of participants we encountered during our limited time in the field who qualified as low counters (with a maximum of 20). After this initial selection, we tested participants' count ranges using a pebble-counting task, described below.

Pebble-counting task

Participants were presented with a pile of glass pebbles (30 for low counters, 40 for high counters) on the testing table. Starting with the pebbles on their left side, participants moved them one at a time to the right while counting each one aloud in the language of their choosing.[1] After they stopped counting, participants were asked how many pebbles there were in the counted set. The experimenters and translator noted any counting errors. With three exceptions, participants performed this task twice, once before and once after completing the matching tasks. We used the higher of the two counts as participants' highest verbal count.

Numerical matching tasks

Participants then performed two nonverbal number tasks in which they were asked to make arrays with the same number of objects that they saw in a sample array. In the parallel-matching task, the experimenter presented a sample array of objects (in a lateral line) for each trial, and participants arranged their response items parallel to each item in the sample array (see Fig. 1, top left; see the Supplemental Material available online for details). Because the sample and response arrays were parallel, participants could use one-to-one correspondence to perform the match in this task, spatially aligning each object in their response array with an object in the sample array without representing the cardinality of either set. For this reason, the parallel-matching task does not test representations of large exact numbers. Rather, success on this task suggests understanding of exact numerical equivalence: For two sets to be equal in number, every element in each set must correspond to an element in the other set (Hume, 1739/1978; Jara-Ettinger et al., 2017; Schneider & Barner, 2020). This task also functioned as a comprehension check, ensuring that participants understood the mechanics of these numerical reproduction tasks. Participants correctly produced a parallel match for each of five sample arrays (N = 3, 4, 5, 10, and 15 pebbles) and then advanced to the orthogonal-matching task.

In the orthogonal-matching task, the sample arrays were arranged sagittally in a line extending away from the participant. Participants arranged their response arrays laterally (as in the parallel-matching task), in a line that was orthogonal to the sample array (Fig. 1, top right; see the Supplemental Material for details). Unlike the parallel-matching task, this task precluded spatially aligning sample and response arrays, and therefore required participants to represent the cardinality of each set. However, it is minimally demanding on participants' numerical abilities: Participants were not asked to perform any arithmetic operations, and because sample and response arrays remained visible throughout each trial, they could inspect them indefinitely before finalizing their responses (which were unspeeded). In a series of practice trials, all participants correctly performed orthogonal matches for sets of size 3, 4, and 5 (with feedback) before advancing to the critical trials.

In critical trials, participants received no feedback about their performance. For high counters, the first critical trial was a sample array of 10 objects. For low counters, the first critical trial was a sample array with two fewer objects than the participant's highest verbal count. From this starting point, we followed a predefined staircasing procedure (i.e., +2 for correct, −1 for incorrect) to determine the size of each sample array until participants (a) produced three incorrect response arrays for sample arrays of the same number (e.g., samples with 15 objects), (b) correctly matched three arrays numbering 20 or more, or (c) completed 20 critical trials.

Psychophysical model of numerical abilities

To evaluate the limits of participants' exact numerical representations, we analyzed their distribution of responses using a generative Bayesian data analysis (A. Gelman et al., 2014). This model formalized a process in which participants use an exact system (with constant error) for smaller sets and an approximate system (with scalar variability) for larger sets. The number at which participants switched from exact to approximate representations is the participant's *switch point*, our dependent measure.

Formally, for the exact system (i.e., numbers below the switch point) we assumed that participants responded from a Cauchy(μ_{low} + n, σ_{low}) distribution, where n is the number of objects in the sample set, and μ_{low} and σ_{low} are location and scale parameters (so that $\mu_{low} \approx 0$ means that responses are centered on the true value n, and $\sigma_{low} \approx 0$ means that responses cluster tightly around the mode $\mu_{low} + n$). A Cauchy distribution was used because errors in the exact system likely reflect inattention or confusion, and estimation of this distribution is robust to outliers. For the approximate system, we assumed a standard model of approximate-number psychophysics (Dehaene, 2011) where participants respond according to the distribution normal(n, $w_i \times n$), where w_i is a Weber ratio parameter that varies by individual. Putting these together, the model assumes that when shown a sample of n objects, participant responses r follow

$$P\left(r \mid n, w_i, \mu_{\text{low}}, \sigma_{\text{low}}\right) \sim \begin{cases} \text{Cauchy}\left(\mu_{\text{low}} + n, \sigma_{\text{low}}\right) & \text{if } n \leq s_i \\ \text{Normal}\left(n, w_i \times n\right) & \text{if } n > s_i \end{cases},$$

where s_i is the switch point of the ith participant. In addition, we included a hierarchical model for participant Weber ratios w_i, in which w_i ~ normal(μ_W and σ_W) was constrained to be positive, meaning that we partially pooled participant estimates of Weber fractions. We put a uniform prior on s_i between 1 and 40,[2] a standard normal prior on μ_{low}, and exponential(1) priors on σ_{low}, μ_W, and σ_W (see the Supplemental Material for details).

This model allowed us to infer the likely distribution of switch-point values s_i from participants' pattern of behavioral responses while accounting for the uncertainty inherent both in exact enumeration (i.e., a noise parameter for low numbers, shared across participants) and in numerical approximation (i.e., a Weber ratio fitted to each participant).

Results

Whereas high counters counted to 40 without error on both trials, low counters' highest verbal counts ranged from 6 to 20 (M = 12.6) and often differed across the two trials (mean absolute difference = 2.0).

Parallel matching

In the parallel-matching task, high counters performed at ceiling, correctly matching each of the sample sets (N = 3, 4, 5, 10, 15 pebbles) on their first attempt. Low counters were 85% accurate on their first attempts and 70% accurate on sets larger than five (i.e., 10 and 15). With one exception, their incorrect responses were within two of the correct number (see Fig. 1, bottom left), and no participant made more than two errors. When participants did make an error, they then showed 100% accuracy on their second attempt, fully reconstructing the response set without feedback about the magnitude or direction of their error.

Orthogonal matching

Participants were less accurate in the orthogonal-matching task (M = 51% correct) than in the parallel-matching task (M = 93% correct), even for the same cardinalities (56% correct for sets of 3, 4, 5, 10, or 15; see Fig. 1, bottom right).

The model estimated a mean Weber ratio of 0.13, consistent with Weber ratios found in studies of numerical estimation in adults (Piazza et al., 2004; Pica et al., 2004), including Tsimane' adults (Gibson et al, 2017). The noise for low numbers was estimated to have a mean (μ_{low}) of −0.14 and a standard deviation (σ_{low}) of 0.14.

The critical question is how participants' switch points were related to their verbal counting abilities.

Figure 2 (left) shows estimated switch points as a function of participants' highest verbal counts, and Figure 3 shows the data for each individual low counter. (Because high counters' switch-point data were less informative on their own, owing to potential ceiling effects, their individual plots can be seen in the Supplemental Material.) Although analysis of the response data was conducted blind to participants' counting abilities, it inferred markedly different switch points for low and high counters solely on the basis of their numerical matching responses. Whereas switch points among the low counters averaged less than seven and never exceeded 11, the average switch point among the high counters was more than 28, $t(17.47) = 11.01$, $p < .0001$.[3] The highest verbal count reliably predicted the switch point above and beyond any effect of formal education: Higher counters had higher switch points, $\beta = 0.55$, $SEM = 0.01$, $t(26) = 5.48$, $p < .0001$. This relationship also held for participants with no education: The highest verbal count reliably predicted the switch point even when we analyzed only those participants with no formal education (i.e., 12 low counters and two high counters), $\beta = 0.40$, $SEM = 0.15$, $t(12) = 2.67$, $p = .02$ (all tests were two-sided).

Importantly, participants' counting abilities and matching abilities were related beyond simple correlation: Low counters' switch points fell at or below their highest verbal count (i.e., below the diagonal dotted line) with only one exception, as shown in Figure 2 (left). According to Pearson χ^2 tests, this ratio (i.e., above:below) differed significantly from chance, $\chi^2(1, N = 15) = 9.60$, $p = 002$. Note that, in principle, data points could fall below the line simply because of a floor effect, in which participants showed low switch points independent of their counting abilities. However, this possibility is unlikely to explain our results, for two reasons. First, our switch-point estimates varied widely on the basis of participants' numerical matching performance alone, demonstrating the sensitivity of our tests. Second, in a further test of this possibility, we conducted a permutation test in which we randomized the pairings of participants' highest verbal counts and switch points. This procedure respects the marginal distribution of each variable and therefore allowed us to evaluate what proportion of data points we should expect to fall below the line by chance (i.e., if verbal counting performance and numerical matching performance were statistically independent). In 10,000 permuted samples, the number of participants whose switch points exceeded their highest counts was 7.72 on average and was never as small or smaller than the number we observed (i.e., one), indicating that the observed pattern is extremely unlikely to have occurred by chance ($p < .001$).

Figure 2 (right) shows the probability that low counters' switch points exceeded their highest verbal counts, calculated using each participant's distribution of switch-point estimates. With one exception, these switch points were below the 50% threshold (M = 11.96%), indicating that they were likely within participants' verbal count range. For numbers beyond their highest verbal count, low counters' responses were on average seven times more likely to reflect an approximate system than an exact system.

In addition to our generative model, we also used a simple behavioral criterion to evaluate participants' highest match: the number at which they failed to produce an exact match three times (i.e., three numerical mismatches, which also served as one of our stopping criteria during testing). Given the staircase procedure we used for testing, producing three numerical mismatches on sets of N required a combination of failing on sets of $N + 1$ and succeeding on sets of $N - 2$. We therefore defined the highest match as two less than the number at which participants produced three mismatches. This alternative measure was highly correlated with participants' switch points as estimated by the model, $R^2 = .67$, $t(11) = 2.97$, $p = .01$. Although these two measures were often identical (see Fig. 3), the highest match was on average higher than estimated switch points (mean difference = 1.77) and therefore provides a more conservative estimate of participants' numerical reproduction abilities. Nevertheless, this alternative measure showed the same relationship to the highest count as the switch-point estimates from our model: With one exception, participants' highest matches were at or below their highest verbal counts (see red circles in Fig. 2, left, and dashed red lines in Fig. 3), and this ratio differed significantly from chance, $\chi^2(1, N = 13) = 7.69$, $p = .006$. Low counters' verbal count range reliably predicted their highest match, $t(11) = 2.44$, $p = .03$. This alternative measure also revealed the same difference between groups: Whereas the highest match for low counters (by this criteria) was below 10 on average (and was always below 15), no high counter produced three mismatches on any number we tested; rather, they all succeeded to make exact numerical matches into the 20s.[4]

Discussion

In a group of Tsimane' adults, the ability to represent exact numbers was limited to the part of the verbal count list they had mastered. Using a generative model of their responses, we found that participants with a limited verbal count range could reliably match the numerosity of sets only within this verbal range; for larger numbers of objects, they overwhelmingly failed to make exact

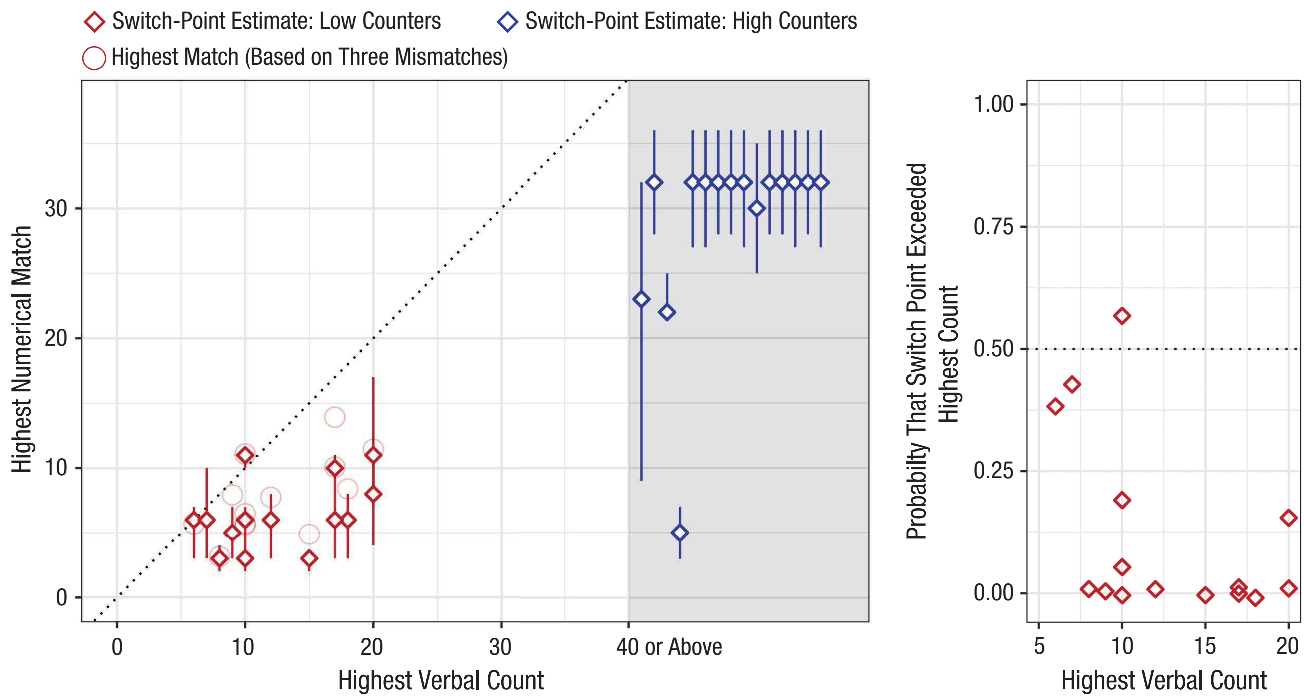

Fig. 2. Overview of results. The graph on the left shows participants' estimated switch points as a function of their highest verbal counts, separately for low counters (red shapes) and high counters (blue shapes). The switch point is defined as the number at which participants switched from exact to approximate number representations. Diamonds indicate median estimates, and error bars indicate 50% confidence intervals. Red circles show results from an alternative measure of highest numerical match that was based on the set size at which participants produced three mismatches. The graph on the right shows the probability that low counters' switch points exceeded their highest verbal counts. Above the dashed line, participants' switch points likely exceeded their verbal count range; below the dashed line, they likely did not.

matches (by two measures), relying instead on numerical approximation. Numerical matching abilities also improved with verbal counting abilities: Participants who showed no upper bound on their verbal count range also showed no upper bound on their numerical matching abilities.

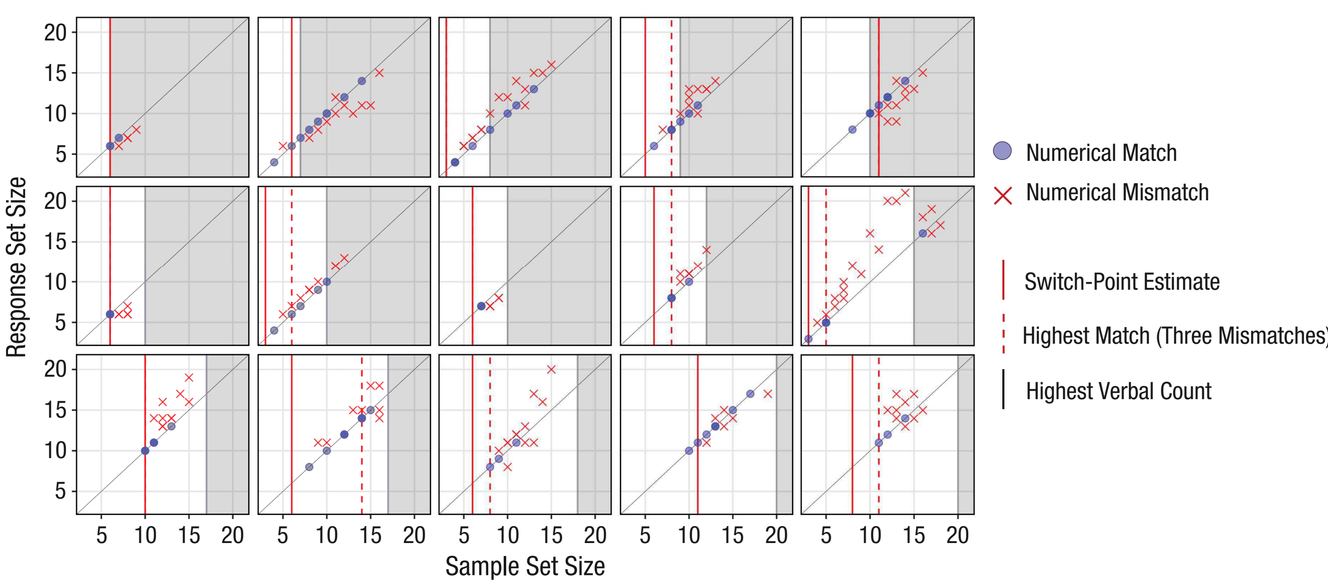

Fig. 3. Data from individual low counters ($N = 15$). Blue dots show correct numerical matches, and red Xs show mismatches. The solid red line indicates each participant's switch point (as estimated by the computational model), and the dashed red line indicates their highest match (based on the set size at which they produced three numerical mismatches). Shaded regions are beyond the participant's verbal count range.

These findings clarify the role of language in number concepts in three ways. First, unlike other isolated groups, our Tsimane' participants succeeded in representing some cardinalities above 4. This success shows that even low counters were not "indifferent to exact numerical equality" (R. Gelman & Gallistel, 2004, p. 442) but rather were attuned to it in both in the orthogonal and parallel-matching tasks (in which they succeeded for sets as large as 15). Yet despite this sensitivity to exact number (and the availability of alternative matching strategies), participants were unable to represent cardinalities beyond their verbal range, even for sets smaller than 10.

Second, the differences in conceptual abilities that we observed cannot reflect broad differences across groups, as our participants shared a common language and culture. In principle, the correlation between participants' numeric abilities could reflect differences in their formal education. However, the highest count reliably predicted the highest numerical match when we controlled for differences in education and when we analyzed only the participants with no formal schooling at all. Therefore, this relationship cannot easily be attributed to differences in language, culture, or formal education.

Finally, whereas previous studies have shown cross-cultural correlations between verbal counting abilities and numerical reproduction abilities, our inferences did not rely on correlation. Rather than simply asking whether one ability predicts the other ability, we also asked whether one ability systematically exceeds the other, allowing us to assess the causal relationship between them. In principle, people could represent "an unbounded set of discrete values" as needed (Leslie et al., 2007, p. 132) once they are equipped with the logic of large exact numbers, whether by experience or by a "preverbal counting mechanism" (Gallistel & Gelman, 1992, p. 43). If so, then low counters' numerical matching ranges should have systematically exceeded their verbal counting ranges. We found the opposite pattern, providing the strongest evidence to date that number words play a functional role in representing large exact numbers (Carey, 2004, 2009; Carey & Barner, 2019; Le Corre & Carey, 2007; Piantadosi et al., 2012; Spelke, 2003) and that this role is not all or nothing. Rather, the ability to represent large exact numbers depends critically on the availability of the corresponding (verbal) symbols. In this way, the verbal count list may serve not only in the induction of numerical principles but also in their use.[5]

In interpreting the findings in the Pirahã, Mundurukú, and other isolated groups, some researchers have characterized the verbal count list as a "cognitive technology" (Frank et al., 2008, p. 819; also see Frank & Barner, 2012), one of many "cultural tools" (Dehaene, 2011, p. 263) for representing number. Although these descriptions may be compelling, they do little to clarify whether a verbal count list is necessary for representing large exact numbers. Some scholars have argued that just as a bicycle is useful but not necessary for transportation, "using words to name exact numerosities is useful but not necessary" (Butterworth et al., 2008, p. 13182) for representing large exact numbers, providing an efficient way to encode numerical information that "complements, rather than alter[s] or replac[es], nonverbal representations" (Gleitman & Papafragou, 2013, p. 515). If such nonverbal representations of large exact numbers exist (e.g., Butterworth et al., 2008; Leslie et al., 2008), they had no effect on the numerical abilities of our participants (or of the Pirahã, Mundurukú, or Nicaraguan Homesigners), none of whom showed any sign of "alternative representational strategies" (Gleitman & Papafragou, 2013, p. 520). Rather, if the verbal count list is a cognitive technology, it is one that not only facilitates large exact number representations but also enables them, allowing people to maintain precise numerical information over time and space (Frank et al., 2008).

Beyond theories of numerical cognition, these findings also bear on a broader debate about the role of language in cognition (Fodor, 1975; Hume, 1748/2000; Bowerman & Levinson, 2001; Lupyan, 2016; Sapir, 1929; Whorf, 2012). Linguistic-relativity effects have been reported in a variety of domains, including color (Forder & Lupyan, 2019; Regier & Kay, 2009), time (Gijssels & Casasanto, 2017), musical pitch (Dolscheid et al., 2013), and spatial reasoning (Majid et al., 2004). However, the idea that language shapes thought remains controversial (Gleitman & Papafragou, 2013; McWhorter, 2014; Pinker, 1994), in part because there are many versions of the Whorfian hypothesis (Casasanto, 2016; Kay & Kempton, 1984). In a strong version of the hypothesis, language not only can change conceptual representations but also can enable new ones (Casasanto, 2016; R. Gelman & Butterworth, 2005). The present results reveal such an effect in the domain of number, where language appears to enable representations of exact cardinalities larger than four. To be clear, language may not be the only external symbol system that can enable large exact number concepts (and not all known number words are precisely represented; Landy et al., 2013). For example, finger counting (Bender & Beller, 2012), body-part counting (Saxe, 1981), and abacus use (Frank & Barner, 2012) may also support the development and elaboration of such concepts (Overmann, 2018; Wiese, 2003). Whatever set of symbols people use, their ability to represent large exact numbers extends only as far as their mastery of those symbols.

Transparency

Action Editor: Vladimir Sloutsky
Editor: Patricia J. Bauer
Author Contributions

The experiment was conceived and designed by B. Pitt. The data were collected by B. Pitt and E.Gibson and analyzed by B. Pitt and S. T. Piantadosi. The manuscript was written by B. Pitt with guidance from E. Gibson and S. T. Piantadosi. All authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by Grant No. 1760874 and No. 2000759 from the National Science Foundation's Division of Research on Learning (to S. T. Piantadosi), Award No. 1R01HD085996 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) at the National Institutes of Health (to S. T. Piantadosi and J. Cantlon), and a network grant from the James S. McDonnell Foundation (to B. Pitt). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Open Practices

Deidentified data and analysis scripts have been made publicly available via OSF and can be accessed at https://osf.io/me7w4/. The design and analysis plan for the study were not preregistered. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Benjamin Pitt https://orcid.org/0000-0002-4912-5627

Acknowledgments

We thank Tomás Huanca, Esther Conde, and Tania Cruz for coordinating our fieldwork; our interpreters Manuel Roca, Robin Nate, and Elías Hiza for their hard work; and Kensy Cooperrider for comments on the manuscript.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211034502

Notes

1. Both Tsimane' and Spanish have recursive, base-10 counting systems. Unlike in English and Spanish, which have irregular constructions such as "eleven" (11) or "trece" (13), Tsimane' number words are regular throughout the teens (e.g., ten-one, ten-two, ten-three). Some participants used a mixture of both languages (e.g., starting in Tsimane' for the lower numbers and then switching into Spanish for the higher numbers), whereas other participants used only one language to count.

2. Our model requires s to have an upper bound, which we set to the largest number of objects tested (i.e., 40 objects in our test of high counters' verbal count range). This provided a conservative estimate of high counters' switch points, which may be much higher (or unbounded), and allowed us to test the sensitivity of these estimates to differences in numerical matching performance.

3. High counters' switch points were estimated on the basis of the range of tested values. Their true switch points (and verbal count ranges) are likely much higher and may be unbounded. Therefore, the difference in switch points we observed between groups is likely an underestimate of the true difference.

4. Two of the 15 low counters did not fail three times on the same number within 20 critical trials, and so their data do not appear in Figures 2 or 3.

5. The availability of a verbal count list may be necessary but not sufficient for representing large exact numbers, as evidenced by the large gap between some participants' highest count and highest match.

References

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition*, *124*(2), 156–182.

Bloom, P. (1994). Generativity within language and other cognitive domains. *Cognition*, *51*(2), 177–189. https://doi.org/10.1016/0010-0277(94)90014-0

Bowerman, M., & Levinson, S. C. (Eds.) (2001). *Language acquisition and conceptual development*. Cambridge University Press.

Brannon, E. M. (2005). What animals know about numbers. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 85–107). Psychology Press.

Butterworth, B., Reeve, R., Reynolds, F., & Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences, USA*, *105*(35), 13179–13184.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59–68.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, *23*(10), 823–835.

Casasanto, D. (2005). Crying "Whorf." *Science*, *307*(5716), 1721–1722.

Casasanto, D. (2016). Linguistic relativity. In N. Riemer (Ed.), *The Routledge handbook of semantics* (pp. 158–174). Routledge.

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour*, *4*(12), 1265–1272.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. MIT Press.

Cooperrider, K., & Gentner, D. (2019). The career of measurement. *Cognition*, *191*, Article 103942. https://doi.org/10.1016/j.cognition.2019.04.011

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (2nd edition). Oxford University Press.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.

Diekmann, Y., Smith, D., Gerbault, P., Dyble, M., Page, A. E., Chaudhary, N., Migliano, A. B., & Thomas, M. G. (2017). Accurate age estimation in small-scale societies. *Proceedings of the National Academy of Sciences, USA*, *114*(31), 8205–8210.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613–621.

Everett, D. L. (2009). *Don't sleep, there are snakes: Life and language in the Amazonian jungle*. Random House.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105–1123.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*(1), 134–149.

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.

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, *64*(1–2), 74–92.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1–2), 43–74.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Taylor & Francis.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences*, *9*(1), 6–10.

Gelman, R., & Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science*, *306*(5695), 441–443.

Gibson, E., Jara-Ettinger, J., Levy, R., & Piantadosi, S. (2017). The use of a computer display exaggerates the connection between education and approximate number ability in remote populations. *Open Mind*, *2*(1), 37–46.

Gijssels, T., & Casasanto, D. (2017). Conceptualizing time in terms of space: Experimental evidence. In B. Dancygier (Ed.), *The Cambridge handbook of cognitive linguistics* (pp. 651–668). Cambridge University Press.

Gleitman, L., & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 504–523). Oxford University Press.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695), 496–499.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*(2–3), 61–83.

Huanca, T. (2008). *Tsimane oral tradition, landscape, and identity in tropical forest*. South-South Exchange Programme for Research on the History of Development.

Hume, D. (1978). A treatise of human nature. In D. D. Raphael (Ed.), *British moralists: 1650–1800*. (Original work published 1739)

Hume, D. (2000). *An enquiry concerning human understanding*. Prentice Hall. (Original work published 1748)

Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R., & Gibson, E. (2017). Mastery of the logic of natural numbers is not the result of mastery of counting: Evidence from late counters. *Developmental Science*, *20*(6), Article e12459. https://doi.org/10.1111/desc.12459

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*(1), 65–79.

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science*, *37*(5), 775–799.

Laurence, S., & Margolis, E. (2007). Linguistic determinism and the innate basis of number. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Vol. 3. Foundations and the future* (pp. 139–169). Oxford University Press.

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

Leslie, A. M., Gallistel, C., & Gelman, R. (2007). Where integers come from. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Vol. 3. Foundations and the future* (pp. 109–149). Oxford University Press.

Leslie, A. M., Gelman, R., & Gallistel, C. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, *12*(6), 213–218.

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning*, *66*(3), 516–553.

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114.

McWhorter, J. H. (2014). *The language hoax: Why the world looks the same in any language*. Oxford University Press.

Overmann, K. A. (2018). Constructing a concept of number. *Journal of Numerical Cognition*, *4*(2), 464–493.

Pahl, M., Si, A., & Zhang, S. (2013). Numerical cognition in bees and other insects. *Frontiers in Psychology*, *4*, Article 162. https://doi.org/10.3389/fpsyg.2013.00162

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*(2), 199–217.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503.

Pinker, S. (1994). *The language instinct*. William Morrow.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446.

Sapir, E. (1929). The status of linguistics as a science. *Language, 5*(4), 207–214.

Sarnecka, B. W., Goldman, M. C., & Slusser, E. B. (2015). How counting leads to children's first representations of exact, large numbers. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 291–309). Oxford University Press.

Saxe, G. B. (1981). Body parts as numerals: A developmental analysis of numeration among the Oksapmin in Papua New Guinea. *Child Development*, *52*, 306–316.

Schneider, R. M., & Barner, D. (2020, July). Children use one-to-one correspondence to establish equality after learning to count. In *CogSci 2020 Proceedings* (pp. 515–521). Cognitive Science Society.

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences, USA*, *108*(8), 3163–3168.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 277–311). MIT Press.

Whorf, B. L. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Wiese, H. (2003). Iconic and non-iconic stages in number development: The role of language. *Trends in Cognitive Sciences*, *7*(9), 385–390.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251.