Research Article

Does Subitizing Reflect Numerical Estimation?

Susannah K. Revkin,[1,2,3] Manuela Piazza,[1,2,3,4] Véronique Izard,[1,2,3,5] Laurent Cohen,[1,2,3,6,7] and Stanislas Dehaene[1,2,3,8]

[1]INSERM, U562, Cognitive Neuroimaging Unit, Gif/Yvette, France; [2]CEA, DSV/I2BM, NeuroSpin Center, Gif/Yvette, France; [3]Université Paris-Sud, IFR 49; [4]Center for Mind/Brain Sciences, University of Trento; [5]Department of Psychology, Harvard University; [6]AP-HP, Hôpital de la Salpêtrière, Department of Neurology, Paris, France; [7]Université Paris VI, IFR 70, Faculté de Médecine Pitié-Salpêtrière; and [8]Collège de France, Paris

ABSTRACT—*Subitizing is the rapid and accurate enumeration of small sets (up to 3–4 items). Although subitizing has been studied extensively since its first description about 100 years ago, its underlying mechanisms remain debated. One hypothesis proposes that subitizing results from numerical estimation mechanisms that, according to Weber's law, operate with high precision for small numbers. Alternatively, subitizing might rely on a distinct process dedicated to small numerosities. In this study, we tested the hypothesis that there is a shared estimation system for small and large quantities in human adults, using a masked forced-choice paradigm in which participants named the numerosity of displays taken from sets matched for discrimination difficulty; one set ranged from 1 through 8 items, and the other ranged from 10 through 80 items. Results showed a clear violation of Weber's law (much higher precision over numerosities 1–4 than over numerosities 10–40), thus refuting the single-estimation-system hypothesis and supporting the notion of a dedicated mechanism for apprehending small numerosities.*

For about 100 years, the fast, accurate, and seemingly effortless enumeration of up to three or four items has presented an enigma to psychologists (for a first account, see Bourdon, 1908). Indeed, adults' enumeration of a visual set of items shows a discontinuity between three or four items and larger numbers. Numerosity naming is fast and accurate for sets of up to three or four items, but suddenly becomes slow and error prone beyond this range, showing a linear increase of about 200 to 400 ms/item (e.g., Oyama, Kikuchi, & Ichihara, 1981; Trick & Pylyshyn, 1994). This dissociation is held to reflect two separate processes in exact enumeration, *subitizing* for small numerosities and counting for larger ones.

Address correspondence to Susannah K. Revkin, INSERM, U562, Cognitive Neuroimaging Unit, CEA/SAC/DSV/I2BM/NeuroSpin, Bât 145, Point Courrier 156, F-91191 Gif/Yvette, France, e-mail: susannahrevkin@gmail.com.

How subitizing operates remains debated. One view proposes that subitizing reflects the use of a numerical estimation procedure shared for small and large numbers (Dehaene & Changeux, 1993; Gallistel & Gelman, 1991). It is now well demonstrated that participants can quickly estimate the approximate quantity of a large array of dots, without counting. This estimation is subject to Weber's law: Judgments become increasingly less precise as numerosity increases, and their variability increases proportionally to the mean response, such that numerosity discrimination is determined by the ratio between numbers (Izard & Dehaene, 2008; Whalen, Gallistel, & Gelman, 1999). Weber's law can be accounted for by a logarithmic internal number line with fixed Gaussian noise (Dehaene, 2007)—a hypothesis that we adopt here for simplicity of exposition, although a similar account can be obtained with the *scalar-variability* hypothesis (according to which noise is proportional to the mean on a linear scale; Gallistel & Gelman, 1992).

Because Weber's law implies that the variability in the representation of small numbers is low, it has been suggested that this law may suffice to explain the transition from subitizing to counting. According to this hypothesis, in an exact-enumeration task with umlimited duration of stimulus presentation, participants would first generate a quick estimation, which would suffice to discriminate a numerosity n from its neighbors $n + 1$ and $n - 1$ when n is small, but would then have to switch to exact counting when n is larger than 3 or 4 and the estimation process becomes too imprecise to generate a reliable answer (Dehaene & Cohen, 1994).

An alternative account postulates a cognitive mechanism dedicated to small sets of objects. Studies of numerosity discrimination in young infants and animals have suggested the existence of two different systems for small and large numerosities (for a review, see Feigenson, Dehaene, & Spelke, 2004). Although babies and animals show a ratio effect for the discrimination of large numerosities, under some circumstances their performance with small numerosities (1–4) escapes Weber's law: They perform well when the quantities to be compared are smaller than 3 (or 4 for monkeys), but performance falls down to chance level when one of the numbers is larger than this limit, even if the ratio is one at which subjects succeed when both quantities are large.

These studies suggest that infants have a distinct system for small numerosities, and that for larger numerosities, this system is supplemented by an estimation system similar to that found in adults. Trick and Pylyshyn (1994) have proposed a similar distinction for adults: a dedicated mechanism of visual indexing that operates over small sets of up to three or four objects and a separate estimation system for larger numbers. According to this proposal, the parallel tagging process that operates over small sets is preattentive, occurring at an early stage of visual analysis during which objects are segregated as individual entities. Because this process is limited to three or four items, a serial deployment of attention is required to enumerate larger quantities, as reflected by the onset of counting in an enumeration task.

In summary, two prominent accounts of subitizing have been proposed: (a) the hypothesis that a single numerical estimation system is used for both small and large sets and (b) the hypothesis that there is a tracking system dedicated to small sets. The present experiment was designed to test these two possibilities. We reasoned that if subitizing relies on numerical estimation, performance should be similar in a naming task with numerosities 1 through 8 and in the same task with decade quantities 10 through 80. If Weber's law is all that matters, these numerosities should be strictly matched for discrimination difficulty (same ratio between 1 and 2 as between 10 and 20, etc.; see Fig. 1). Therefore, once participants are trained with using only decade numbers, the disproportionately higher precision expected over the range from 1 through 4 should also be seen in the range from 10 through 40: One should see "subitizing" even for large numbers as long as they are sufficiently discriminable. If this were not the case, it would clearly indicate that Weber's law does not suffice to account for subitizing, and that a distinct process must be at play with numerosities 1 through 4.

We further reasoned that if subitizing arises from approximate estimation, its range should be determined by participants' capacity for numerosity discrimination (as measured in a large-number comparison task). Specifically, participants with better discrimination capacities should be more precise in both naming tasks, and in particular should have a larger subitizing range.

Our paradigm was designed so that conditions were identical for the two naming tasks. To prevent counting, subgrouping, or arithmetic-based strategies, we masked stimuli and required participants to respond within a short delay. It is important to note that we calibrated participants, as participants spontaneously underestimate large quantities, but can be trained to label them accurately (Izard & Dehaene, 2008). To reinforce this calibration process, we gave feedback at the end of each trial. Finally, because naming small quantities is a much more familiar task than naming decades, participants were trained intensively.

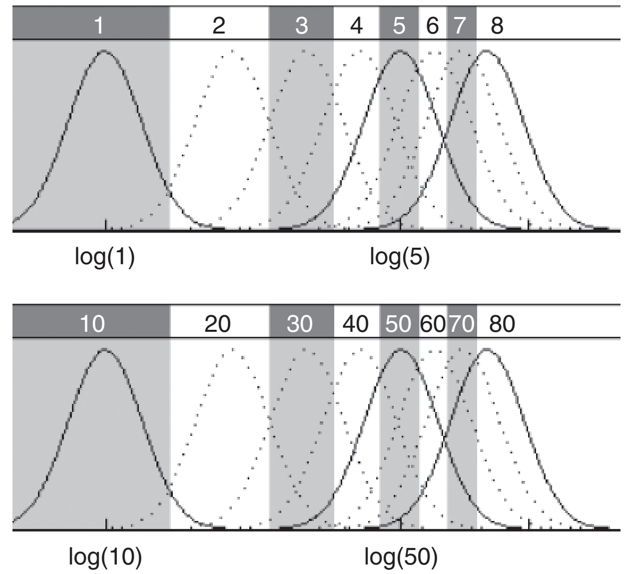

Fig. 1. The log number-line model of performance in the naming tasks. Numerosities are first encoded on an internal continuum, the mental number line, which, in this model, is compressed (so that numerosity 1 is represented by log(1), etc.; see x-axis). The model assumes that each numerosity is encoded by a fluctuating representation, that is, a random variable that is distributed according to a Gaussian distribution with fixed variability. In the figure, solid lines highlight the distributions corresponding to numerosities 1, 5, and 8 (top panel) and 10, 50, and 80 (bottom panel). The distributions for other numerosities are represented with dotted lines. The internal representation of numerosity is translated into a verbal number word by means of a response grid: The number line is divided into segments, each corresponding to a different verbal label (see the labels above the distributions). This figure depicts an optimal response grid, in which the response criterion marking the boundary between two adjacent response labels is optimally placed where the two underlying distribution curves meet. According to this model, numerosities 1 through 8 (top panel) do not differ from decades 10 through 80 (bottom panel) in discrimination difficulty, so naming performance should be equivalent for these two sets. Specifically, naming should be almost flawless over the first numerosities in each set (because there is little underlying representation overlap) and become progressively less precise for larger numerosities (because of the increase in overlap).

METHOD

Participants

Eighteen right-handed participants (8 men, 10 women; mean age = 24.9 years, range = 18–38) with no history of neurological or psychiatric disease, and normal or corrected-to-normal vision, gave written informed consent.

Tasks and Procedure

Tasks were programmed using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) and administered on a portable computer at a viewing distance of 57 cm. Participants performed a comparison task and two naming tasks.

Dots Comparison Task

In this task, participants were presented with two dot arrays and judged as accurately and as quickly as possible which one contained more dots. On each trial, one array contained a fixed numerosity (16 for half the trials, 32 for the other half), and the other array (*varying numerosity*) contained a numerosity that was smaller or larger than the fixed numerosity by one of four possible ratios: 1.06, 1.13, 1.24, or 1.33. Thus, comparison difficulty was manipulated (the smaller the ratio, the harder the comparison). The two fixed numerosities and four ratios were presented in random order across blocks. Participants responded by pressing the mouse button on the same side as the larger array (using their left or right index fingers). The dots, present on the screen until participants responded, were black, and the two arrays appeared in two white discs on either side of a central white fixation spot (after a delay of 1,400 ms); the background of the display was black. On half the trials, dot size of the varying-numerosity array was held constant, and on the other half, the size of the area occupied by the varying-numerosity array was held constant; in the fixed-numerosity arrays, these parameters were varied simultaneously. This design was adopted to prevent participants from basing their performance on these nonnumerical parameters (see Dehaene, Izard, & Piazza, 2005, for a more detailed description of the logic behind such manipulations). Participants performed 16 training trials with accuracy feedback and then performed a total of 128 experimental trials (32 trials per ratio).

Numerosity Naming Tasks

Participants performed two naming tasks, one with numerosities 1 through 8 (the 1–8 task) and one with decade numerosities 10 through 80 (the 10–80 task). Both tasks were administered in each of two sessions. The tasks order was counterbalanced across sessions and participants.

The procedure was identical for the two naming tasks. Participants were explicitly informed which task was coming next and which quantities could appear during the task. They were instructed to name the number of dots as accurately and quickly as possible, within 1 s (otherwise the trial would be discarded). They were first calibrated by viewing 16 examples of the stimuli, which consisted of random patterns of dots. In order to make sure that participants' estimation was based on numerosity and not on other continuous parameters, we kept dot density constant in half the calibration and test stimuli and kept dot size constant in the other half. During calibration, examples and correct answers were presented for up to 10 s, according to the participant's need.

Test trials began with a central cross, which flashed twice to announce the arrival of the dots; the dots were followed by a flicker mask and finally a black screen (see Fig. 2). Participants

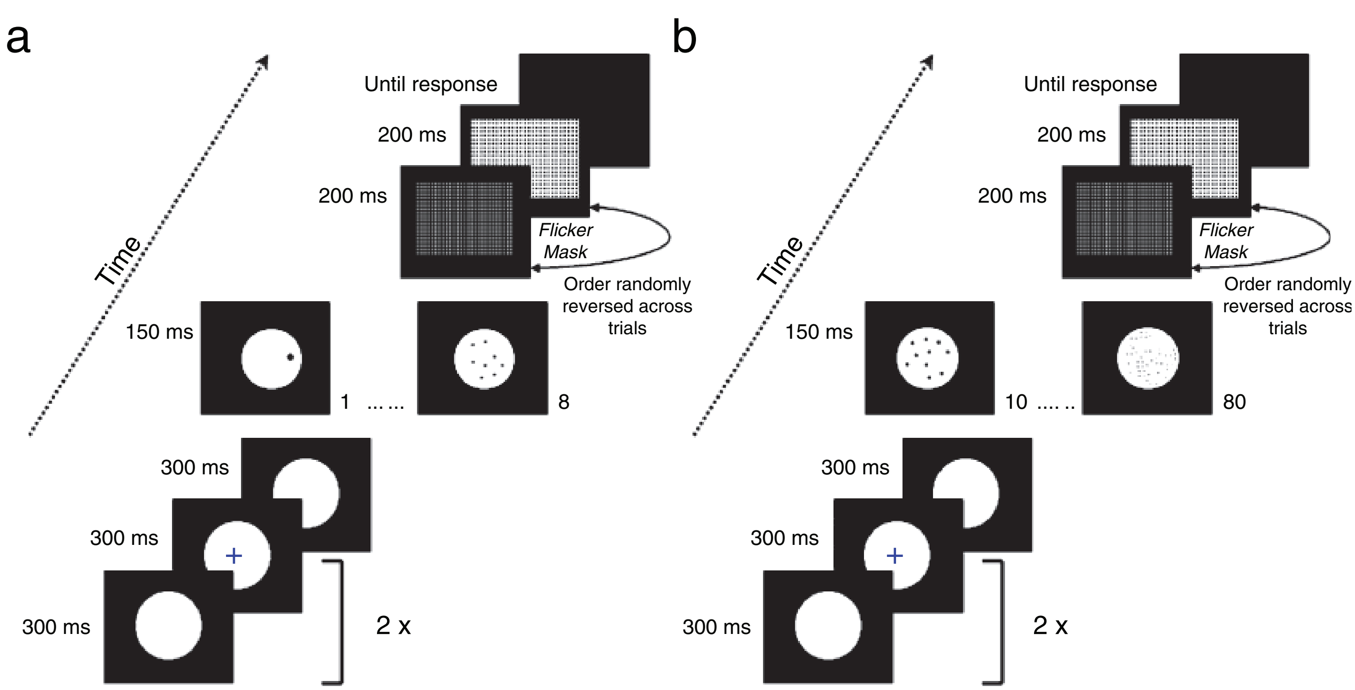

Fig. 2. The naming tasks. In the 1–8 naming task (a), after a fixation cross flashed twice, participants were shown a group of 1 to 8 dots, followed by a mask; the task was to name the presented numerosity as quickly as possible using the labels "1" through "8." In the 10–80 naming task (b), the procedure was identical except that only numerosities 10 through 80 were presented, and participants used only decade names (i.e., "10"–"80") as labels.

responded using a microphone. Responses given within 1 s were entered by the experimenter using the keyboard, and participants then received feedback (the correct response was displayed if the response had been incorrect). If response time exceeded 1 s, a slide that encouraged faster responses and showed the correct answer was displayed. Each block consisted of 16 calibration trials followed by 40 experimental trials presenting each numerosity five times, in random order. Participants performed four blocks of each test in each session, for a total of eight blocks (320 trials, 40 presentations of each numerosity) over the two sessions. The first two blocks of each test in each session were discarded as training, and analysis was therefore limited to a maximum of 160 trials per test (20 trials of each numerosity per test, or less if the participant responded too slowly on some trials).

For analysis, the error rate, mean response time (RT), mean response, and variation coefficient (VC; standard deviation of response divided by mean response) were calculated for each numerosity and each participant. Scalar variability and Weber's law are reflected by a stable VC across numerosities (Izard & Dehaene, 2008; Whalen et al., 1999), and the VC thus gives an indication of the overall precision of the underlying numerical representation (Izard & Dehaene, 2008).

RESULTS

Dots Comparison Task

Accuracy on the dots comparison task was used to estimate the internal Weber fraction (w), a measure of the precision of underlying numerical representation, for each participant, using a method previously described (the "maximum-likelihood decision model" described in the Supplemental Data from Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). This method basically estimates the standard deviation of the theoretical Gaussian distribution that determines the precision of the internal numerosity representation on a log scale (see Fig. 1). Mean w across participants was 0.18 ($SD = 0.06$, $Mdn = 0.16$). Participants were divided by median split into two discrimination-precision groups: low ($w > 0.16$; 7 participants) and high ($w \leq 0.16$; 11 participants). The two groups did not differ on overall RT, $t(16) = 1.50$, $p = .15$.

Numerosity Naming Tasks

Few numerosity trials were excluded because of excessive RTs (1–8 task: $M = 3.44$, $SD = 2.31$; 10–80 task: $M = 6.78$, $SD = 3.95$). For each task, preliminary analyses of variance (ANOVAs) showed that the order of the task, session number, and nonnumerical continuous parameter held constant did not have significant effects on error rates, RTs, and VCs; data were therefore collapsed across these factors. The data were then analyzed in a 2 (numerosity range: 1–8 vs. 10–80) × 2 (discrimination-precision group: low vs. high) × 8 (rank order of the numerosity: from 1, for 1 or 10, up through 8, for 8 or 80) ANOVA.

Error Rate

The error rate was significantly lower in the range from 1 through 8 ($M = 21\%$, $SD = 7\%$) than in the range from 10 through 80 ($M = 51\%$, $SD = 6\%$), $F(1, 256) = 518.32$, $p < .01$, and was also lower for participants in the high-precision group ($M = 32\%$, $SD = 4\%$) compared with those in the low-precision group ($M = 39\%$, $SD = 2\%$), $F(1, 256) = 30.06$, $p < .01$; there was also a significant effect of rank order, $F(7, 256) = 104.49$, $p < .01$, the error rate being lower for smaller numerosities within each range.

Crucially, the interaction between range and rank order was highly significant, $F(7, 256) = 32.64$, $p < .01$, thus violating the prediction of a similar influence of rank order across the two ranges, as derived from Weber's law. In the range from 1 through 8, the error rate was essentially zero for numerosities 1 through 4, and began to rise steeply beginning with numerosity 5 (see Fig. 3a). By contrast, in the range from 10 through 80, errors were frequent even for numerosities 20 and 30 (see Fig. 3b).

The group factor interacted significantly with rank order, $F(7, 256) = 3.65$, $p < .01$, as the lower error rate for participants with high precision in numerical comparison was particularly evident for ranks 6 through 8. The triple interaction was also significant, $F(7, 256) = 4.17$, $p < .01$; participants with high precision made fewer errors than those with low precision for most numerosities in the large-numerosity task, but only for numerosities 5 through 7 in the small-numerosity task. The groups did not differ for numerosities 1 through 4.

Response Times

RTs revealed a main effect of range, $F(1, 256) = 517.40$, $p < .01$, being faster in the range from 1 through 8 ($M = 588$ ms, $SD = 32$ ms) than in the range from 10 through 80 ($M = 737$ ms, $SD = 44$ ms). Participants with high discrimination precision were slightly slower ($M = 672$ ms, $SD = 30$ ms) than those with low precision ($M = 655$ ms, $SD = 41$ ms), $F(1, 256) = 8.09$, $p < .01$. There was also a main effect of rank order, $F(7, 256) = 31.36$, $p < .01$; RTs increased from rank 1 through rank 5 and then stabilized. Crucially, an interaction of range and rank order, $F(7, 256) = 27.14$, $p < .01$, again showed differential processing of the small numbers 1 through 4, which were processed with much faster RTs than either 5 through 8 or 10 through 80 (see Figs. 3c and 3d). This result again shows a distinct pattern of processing within the subitizing range, contrary to predictions derived from Weber's law.

Finally, range also interacted with group, $F(1, 256) = 9.03$, $p < .01$, as participants with high precision were slightly slower than those with low precision in the 10–80 task only ($M = 751$ ms, $SD = 36$ ms, vs. $M = 715$, $SD = 48$ ms). All other effects were nonsignificant.

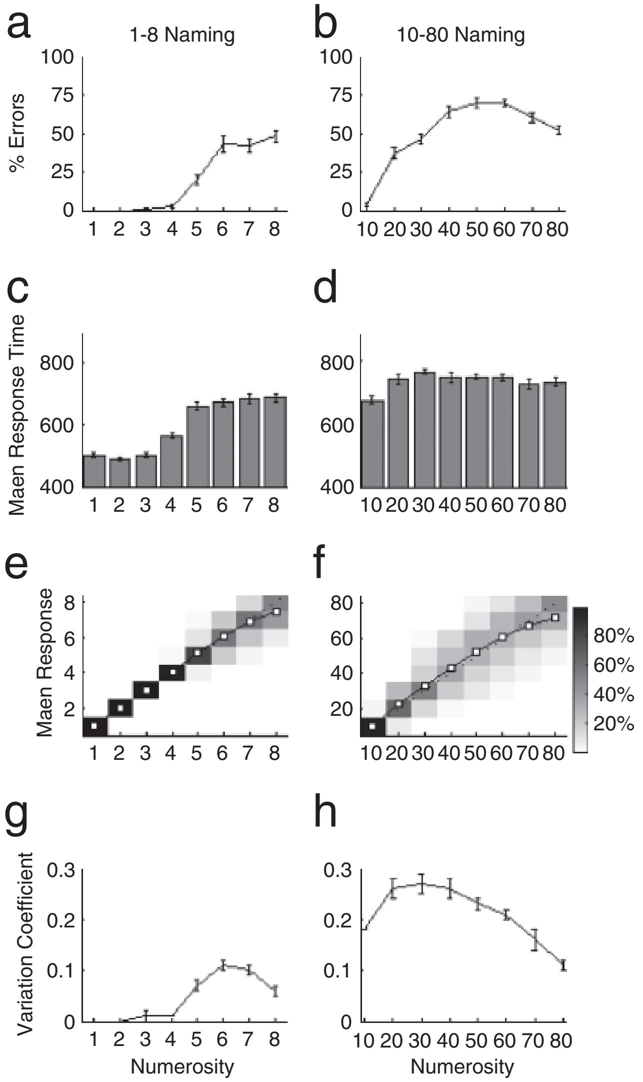

Fig. 3. Results of the two naming tasks: the 1–8 task (left) and the 10–80 task (right). From top to bottom, the graphs show the percentage of errors (a, b), response time (c, d), mean response (e, f), and variation coefficient (g, h) as a function of presented numerosity. Error bars represent ±1 *SE*. In (e) and (f), the dotted lines indicate ideal performance, and the gray shading indicates response frequency in relation to the total number of responses given for each numerosity (see the scale on the right).

In sum, we again found clear differences between the two ranges, participants being much faster in the 1–8 task than in the 10–80 task and showing a subitizing effect only over the numerosities 1 through 4. Also, discrimination precision influenced performance only in the 10–80 task, which suggests that variability in the range from 1 through 8 was governed by principles other than large-number estimation accuracy.

Mean Response and Variation Coefficient

In both number ranges, mean response was quite close to the correct one, and variability in responses increased as numerosity increased, a signature of estimation processes (see Figs. 3e and 3f). However, a clear broadening of the response range appeared already at numerosity 20 in the 10–80 task, whereas a comparable broadening did not appear until much later (numerosity 5) in the 1–8 task.

To validate these observations statistically, we estimated mean response and standard deviation of responses by fitting the cumulative response distribution for each numerosity and each participant with the cumulative of a Gaussian distribution function. Fit was overall excellent for both the 1–8 task[1] (R^2: $M = 1.00$, $SD = 0.00$) and the 10–80 task (R^2: $M = .99$, $SD = .006$), except in the case of extreme numerosities, which sometimes showed anchoring effects (very little response variability). Extreme numerosities were therefore excluded from the VC analyses for both ranges, and VCs were analyzed in a 2 (range) × 2 (group) × 6 (rank order of the numerosity) ANOVA.

There was a main effect of range, $F(1, 192) = 636.25$, $p < .01$, VC being much lower in the 1–8 task ($M = 0.05$, $SD = 0.02$) than in the 10–80 task ($M = 0.23$, $SD = 0.04$). There was a trend toward a main effect of rank order, $F(5, 192) = 2.31$, $p = .05$, VC being lower for the extreme numerosities, presumably because of a remaining anchoring effect. Crucially, an interaction between range and rank order was again observed, $F(5, 192) = 26.52$, $p < .01$; VC was dramatically lower in the range from 1 through 4 compared with the range from 5 through 8, but no such effect was seen for 10 through 40 versus 50 through 80 (Figs. 3g and 3h).

A main effect of group, $F(1, 192) = 25.45$, $p < .01$, indicated that participants with high precision had lower VCs ($M = 0.13$, $SD = 0.03$) than participants with low precision ($M = 0.16$, $SD = 0.02$). No group-by-range interaction or triple interaction was found; however, participants with higher precision had a lower mean VC over numerosities 20 through 70, $t(16) = -2.27$, $p < .05$, and over numerosities 5 through 7, $t(16) = -4.62$, $p < .01$, but not over numerosities 2 through 4, $t(16) = -0.74$, $p = .47$ (see Fig. 4).

In summary, responses showed an abrupt increase in variability between numerosities 4 and 5, a result not expected from a purely Weberian estimation process. No such discontinuity was found in the range from 10 through 80. Also, participants with higher discrimination precision had lower VCs, particularly in the 10–80 task and outside the subitizing range in the 1–8 task.

Predictors of Subitizing Range and Response Precision

We conducted correlation analyses to explore further the links between different measures of response precision. First, we determined a subitizing range for each participant using the data from the 1–8 task. The subitizing range was estimated by fitting the full RT curve (excluding numerosity 8) with a sigmoid

[1]Most participants showed no variability in response for numerosities 1 through 4, and the VC was null without fitting response distributions in these cases.

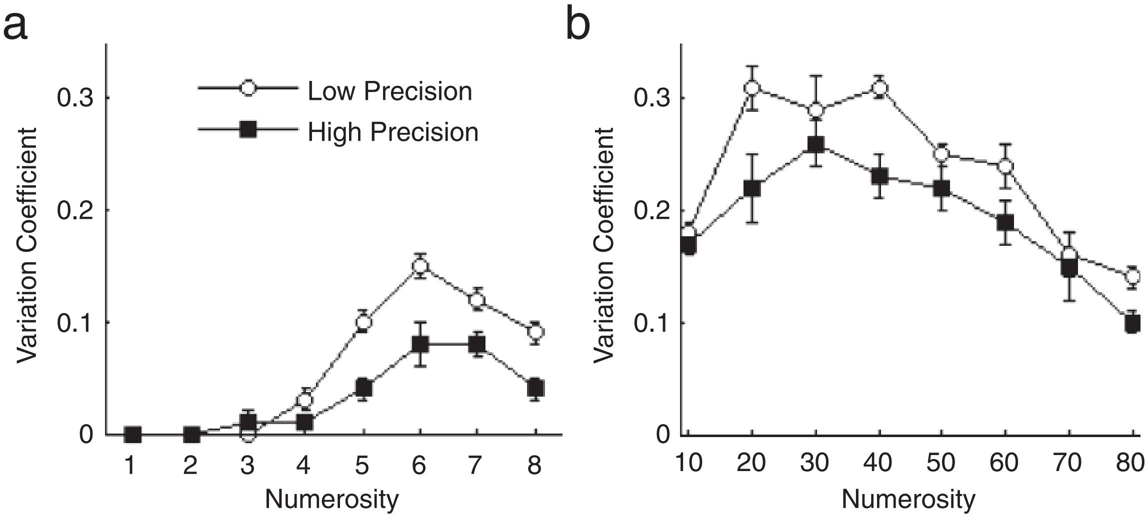

Fig. 4. Variation coefficient as a function of presented numerosity and discrimination-precision group (low precision or high precision) in the 1–8 task (a) and the 10–80 task (b). Error bars represent ±1 *SE*.

function of numerosity and taking the inflexion point of that curve (1 outlier participant was excluded). Data fit was excellent (mean $R^2 = .91$, $SD = .12$). This analysis yielded a mean subitizing range of 4.38 ($SD = 0.25$). If subitizing is due to a single process of estimation used for small and large numbers, subitizing range, Weber fraction in numerosity comparison, and precision of numerosity naming should be tightly correlated. Contrary to this prediction, subitizing range was unrelated to w, the discrimination parameter ($r = -.03$), and was unrelated to VC for numerosities 2 through 7 (r = −.08) and 20 through 70 ($r = -.31$). However, VC for numerosities 20 through 70 was more strongly correlated with VC for numerosities 5 through 7 than with VC for numerosities 2 through 4 (rs = .78 and .52, respectively).

DISCUSSION AND CONCLUSIONS

In conflict with Weber's law, but in agreement with the hypothesis that there is a mechanism dedicated to processing small numbers, various measures revealed a disproportionate precision in the range of numerosities 1 through 4. The VC approached zero for these numerosities, indicating null or very little variability in response, errors being exceedingly rare in this range. In contrast, there was no clear advantage for numerosities 10 through 40 in the 10–80 task. In particular, the VC for 10 through 40 was high, as errors were frequent and response variability was large.

Analyses of interindividual variability confirmed the special status of the subitizing range. Compared with participants with low discrimination precision for large numerosities, those with high precision made fewer errors over most numerosities in the 10–80 task and outside the subitizing range in the 1–8 task and were overall more precise. However, the subitizing range did not correlate either with discrimination precision or with naming precision.

In sum, the clear performance differences between the two naming ranges, the unique advantage for numerosities in the subitizing range, and the absence of a correlation between subitizing and large-number performance strongly suggest that there is a separate system dedicated to small numerosities (1–4), and go against the hypothesis that subitizing is estimation at a high level of precision (Dehaene & Changeux, 1993; Gallistel & Gelman, 1991). Our results are in line with those of studies of young infants and animals, which provide evidence for a separate mechanism of apprehension of small quantities in these populations (for a review, see Feigenson et al., 2004).

Our study also allowed us to investigate the link between numerosity comparison and numerosity naming. According to the log number-line model (Dehaene & Changeux, 1993; Izard & Dehaene, 2008; see Fig. 1), a single parameter, the internal Weber fraction, should directly predict both abilities. Our data support this hypothesis, as participants with higher discrimination precision were also more precise in naming. Those results are in line with a recent mathematical theory that shows how performance and RT curves in those classical numerical tasks can be derived from first principles based on the log number-line hypothesis (Dehaene, 2007).

Our data contrast with those of Cordes, Gelman, Gallistel, and Whalen (2001), who found no difference between the VCs within and outside the subitizing range and therefore argued for continuous representation of small and large numerosities. Although our data suggest the existence of a distinct, exact system for small numerosities, it is possible that such a system coexists with an approximate system for small numerosities, but that the system used in a given case depends on task conditions. In the study by Cordes et al. (2001), stimuli were Arabic numerals, and responses were fast nonverbal tapping. Perhaps there is a separate system for the apprehension of small numerosities that predominates in cases in which nonsymbolic visual stimuli, such as dots, are presented simultaneously as a set.

It is important to note that for both naming tasks, participants were intensively trained and received regular feedback to counter a possible effect of familiarity with naming smaller numerosities. Although one could object that this training was still insufficient, the clear discontinuity between numerosities 1 through 4 and 5 through 8 would still need to be explained. Such a discontinuity is perhaps not surprising in RTs in a classical subitizing task (unlimited duration of stimulus presentation), because participants are thought to switch strategies and start counting at about 4 or 5 items (Piazza, Giacomini, Le Bihan, & Dehaene, 2003). However, in our study, the masking and narrow window for response prevented participants from counting, and, indeed, RTs showed no serial increase in either the subitizing range (1–4) or the counting range (5–8). Because counting was prevented, proponents of the subitizing-as-estimation hypothesis would have to argue that the entire curve over the range from 1 through 8 was due to numerosity estimation—yet the results clearly indicate that estimation was dramatically more precise from 1 through 4 than from 5 through 8, in disagreement with a system obeying Weber's law. Simulations by current models of numerosity estimation, such as Dehaene and Changeux's (1993) or Verguts and Fias's (2004) model, produce performance that obeys Weber's law even in the small-number range, and are thus unable to account for the present data with a single process.

Our results also showed slightly better performance over numerosities 5 through 8 than over numerosities 10 through 80 (see Fig. 3), which is not expected if subjects used only estimation over both these ranges. However, it is possible that this pattern reflects use of a strategy of coupling subitizing with estimation in the range from 5 through 8. Subjects might have subitized subgroups of the stimuli in addition to estimating the remaining quantity of dots, a strategy that was not applicable in the 10–80 task.

Our results clearly suggest that subitizing is not linked to discrimination difficulty as determined by Weber's law. However, one could argue that a single process of discrimination underlies subitizing and estimation, but that Weber's law does not hold for small numbers up to 3 or 4, for which the Weber fraction would be much smaller. For instance, it could be suggested that in the small-number range, variables other than the ratio between stimuli (possibly variables relating to spatial arrangement or to other perceptual factors) boost number discriminability. Such a postulate may salvage the hypothesis that a single process of discrimination underlies subitizing and estimation, but it essentially amounts to postulating that small numerosities receive a special treatment within the estimation system—a proposal that seems hard to separate from the original subitizing hypothesis.

In conclusion, although our study provides evidence against estimation as the underlying mechanism of subitizing, the question remains open as to whether subitizing relies on a domain-specific numerical process or on a domain-general cognitive process. One hundred years after the discovery of subitizing, its mechanisms remain as mysterious as ever—but it is now clear that they are not based on a Weberian estimation process.

Acknowledgments—This study was supported by INSERM, CEA, two Marie Curie fellowships of the European Community (MRTN-CT-2003-504927, QLK6-CT-2002-51635), and a McDonnell Foundation centennial fellowship.

REFERENCES

Bourdon, B. (1908). Sur le temps nécessaire pour nommer les nombres. *Revue philosophique de la France et de l'étranger*, *65*, 426–431.

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

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Attention & performance XXII: Sensori-motor foundations of higher cognition* (pp. 527–574). Cambridge, MA: Harvard University Press.

Dehaene, S., & Changeux, J.P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*, 390–407.

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 958–975.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*. Available from http://www.unicog.org/main/pages.php?page=Documentation

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

Gallistel, C.R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In F. Craik, W. Kessen, & A. Ortony (Eds.), *Essays in honor of George Mandler* (pp. 65–81). Hillsdale, NJ: Erlbaum.

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

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221–1247.

Oyama, T., Kikuchi, T., & Ichihara, S. (1981). Span of attention, backward masking and reaction time. *Perception & Psychophysics*, *29*, 106–112.

Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 1237–1245.

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime users guide*. Pittsburgh, PA: Psychology Software Tools.

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*, 80–102.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*, 1493–1504.

Whalen, J., Gallistel, C.R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*, 130–137.

(Received 7/12/07; Revision accepted 11/26/07)