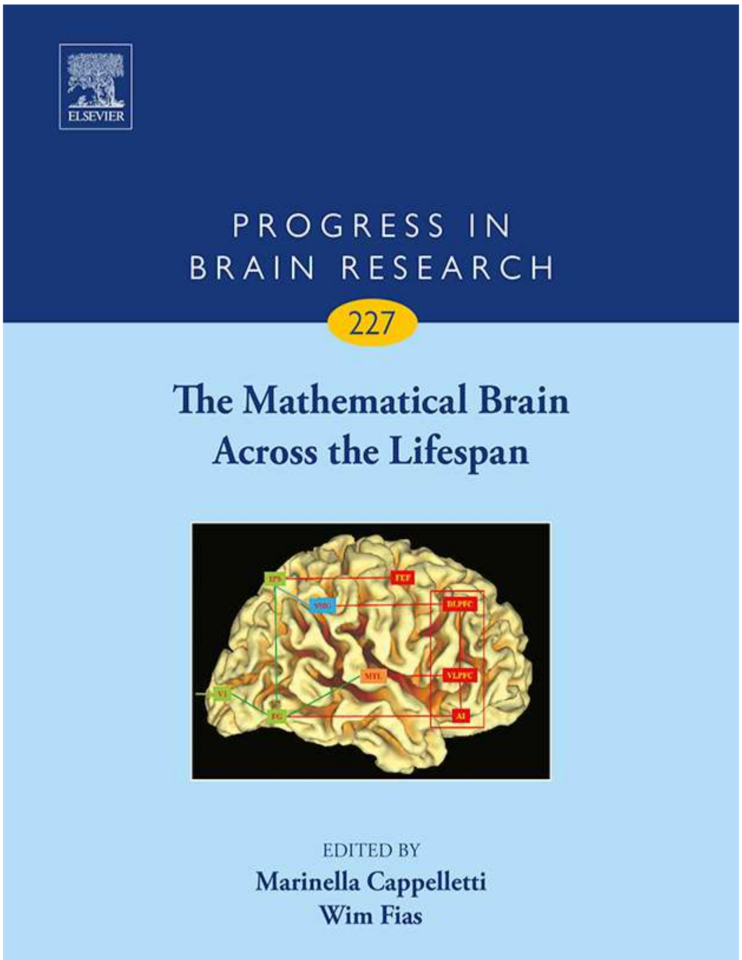

From V. Karolis and B. Butterworth, What counts in estimation? The nature of the preverbal system. In: Marinella Cappelletti and Wim Fias, editors, *Progress in Brain Research, Vol. 227*, Amsterdam: Elsevier, 2016, pp. 29-51.
ISBN: 978-0-444-63698-0

CHAPTER

What counts in estimation? The nature of the preverbal system

2

V. Karolis*[,1], B. Butterworth[†,‡,§]

*Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[†]Institute of Cognitive Neuroscience, University College London, London, United Kingdom

[‡]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Vic, Australia

[§]Research Center for Mind, Brain, and Learning, National Chengchi University, Taipei, Taiwan

[1]Corresponding author: Tel.: +020-7679-1150; Fax: +020-7813-2835, e-mail address: slava.karolis@kcl.ac.uk

Abstract

It has been proposed that the development of verbal counting is supported by a more ancient preverbal system of estimation, the most widely canvassed candidates being the accumulator originally proposed by Gibbon and colleagues and the analogue magnitude system proposed by Dehaene and colleagues. The aim of this chapter is to assess the strengths and weaknesses of these models in terms of their capacity to emulate the statistical properties of verbal counting. The emphasis is put on the emergence of exact representations, autoscaling, and commensurability of noise characteristics. We also outline the modified architectures that may help improve models' power to meet these criteria. We propose that architectures considered in this chapter can be used to generate predictions for experimental testing and provide an example where we test the hypothesis whether the visual sense of number, ie, ability to discriminate numerosity without counting, entails enumeration of objects.

Keywords

Preverbal system, Analogue magnitude system, Numerosity, Stochastic process, Poisson distribution

1 THE PREVERBAL SYSTEM

Any consideration of the mathematical brain across the lifespan should start with the moment of conception and the nature of the genotype that will build the brain. Does the genotype encode for brain systems that are specific to mathematics? If it does, what is the neural mechanism and what can it do?

Progress in Brain Research, Volume 227, ISSN 0079-6123, http://dx.doi.org/10.1016/bs.pbr.2016.04.025

Here, we approach this question by trying to model what this mechanism can and cannot be. It is now widely accepted that we are born with such mathematics, more specifically a number, mechanism, but exactly what it is and how it works is disputed. We will refer to this putative mechanism as "preverbal" since there is abundant evidence that it is available to the child before it learns to count with words. The aim of this chapter is to review statistical properties of models of preverbal magnitude processing that could provide a basis for the subsequent acquisition of verbal counting. Three proposals should be mentioned in this respect.

1.1 ACCUMULATOR MODEL

The first proposal originates in the work by Gelman and Gallistel who postulated that human infants possess a system of "numerons," which are "any distinct and arbitrary tags that a mind (human or nonhuman) uses in enumerating a set of objects" (p. 77), and they are to be distinguished from "numerlogs," the "traditional count words." Verbal counting depends on a developmental process by which a fixed sequence of independently acquired numerlogs is linked to numerons (Gelman and Gallistel, 1978).

They develop this idea in a later theoretical paper (Leslie et al., 2008). First, they specify numerons as "the brain's integer symbols" (p. 217) rather than any arbitrary tags. Second, they explicate numerons in terms of a mechanism derived from work with other species, usually rats and pigeons, and which they assume is innate in human infants. This is the "accumulator," a mechanism for enumeration first proposed by Gibbon and his colleagues (Gibbon et al., 1984; Meck et al., 1985). The basic design of the accumulator (Fig. 1) is essentially like a thermometer: the higher the mercury, the higher the temperature—the greater the accumulation, the larger the numerosity it represents. The key subcomponents are a "pacemaker" that generates energy, a "gate" (sometimes called a "switch") that opens for each object experienced, and a "storage" which stores a quantum of energy for each "gate" opening.

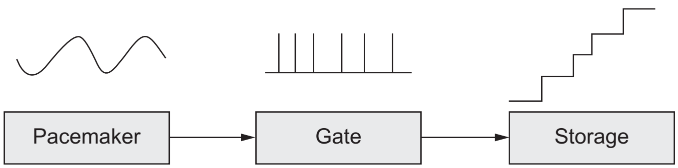

FIG. 1

The accumulator. The brain contains a "pacemaker" that generates quanta of energy and a "gate" that lets through a quantum for each object to be enumerated. The quanta are stored in the "storage" and the quantity of quanta represents the number of objects enumerated.

Adapted from Gibbon, J., Church, R.M., Meck, W.H., 1984. Scalar timing in memory. Ann. N. Y. Acad. Sci. 423, 52–77. Meck, W.H., Church, R.M., 1983. A mode control model of counting and timing processes. J. Exp. Psychol. Anim. Behav. Process. 9(4), 320–334.

We should note that quite often the term "accumulator" is used to denominate what we have called "storage." However, for the considerations of clarity, we will reserve the term "accumulator" to refer to the whole architecture rather than to its particular subcomponent.

One further feature of the model is that the mechanism can handle both continuous and discrete quantities. In the case of the accumulator, Gibbon et al. have used it to explore timing (Gibbon et al., 1984; Meck and Church, 1983)—rats, instead of responding to a particular number of sounds, are rewarded for responding to a particular stimulus duration. The mechanism measures the duration by holding the "gate" open for the duration of the stimulus, and this is then stored in the accumulator.

Our thermometers come with calibration marks so that we can read off the temperature from a continuum: a particular level represents 30°, and so on. A decent thermometer is also expected to produce similar readings if the actual temperature remains the same. This is not the case for the accumulator because its "storage" comes without "marks" and representations of magnitudes that are quite noisy. The latter feature reflects the fact that nonverbal number representations obeys a scalar variability criterion (Izard and Dehaene, 2008); that is, the standard deviation of the error in responses grows proportionally to number magnitude. As Leslie et al. put it "What an analogue system will not support is the notion of exact equality, or perfect substitutability, because the mental magnitudes are noisy; they represent quantity plus or minus some percent uncertainty." (p. 213). In order to use the accumulator to count, the human learner needs to learn three things, according to Leslie et al. (see Fig. 2):

(a) To calibrate the accumulator so that each level corresponds to the mean activity for each numerosity;
(b) To link the calibrations with the independently acquired sequence of counting words ("compact notation"), which are interpreted as integers "because children are disposed to entertain integer-valued hypotheses when learning the meanings of those words" (Leslie et al, 2008, p. 213); and
(c) To learn that the calibration and the counting sequence can continue indefinitely. For this last characteristic, a simple recursive rule is needed to generate the next term in sequence to add to the previous term, plus the assumption of oneness (ONE in Fig. 2).

1.2 ANALOGUE MAGNITUDE SYSTEM

The second proposed model originates in network simulation studies (Dehaene and Changeux, 1993) and builds on the first, in the sense that it contains an accumulator as a part of its architecture. The model is shown in Fig. 3. One key feature of this model, as in the accumulator, is a stage in which the inputs—the objects to be enumerated—are "normalized." That is, whatever the object size, the object gets the same representation on the "location map." The output of the location map is

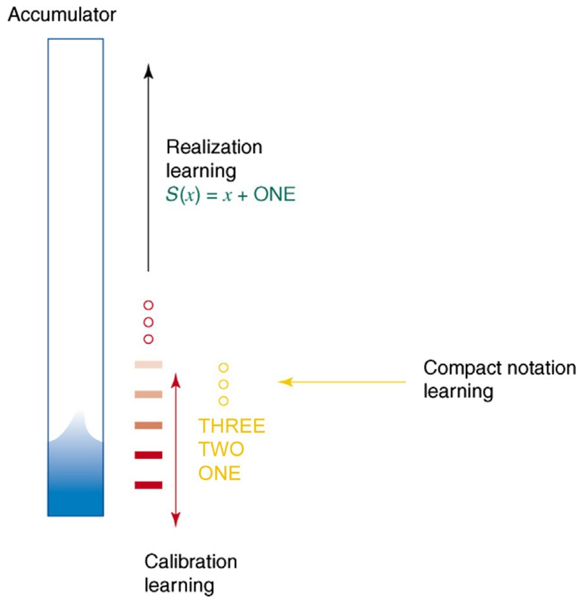

FIG. 2

A model of number learning and representation. "The model combines continuous magnitude (accumulator) and integer representations with three types of learning. … Accumulator magnitudes are depicted as inherently noisy … [The] model adds discrete representations of exact integer values that are generated by a successor function *S*. These are depicted as a series of hash marks in a grid that can be calibrated against accumulator magnitudes and associated with unique identifier symbols (such as "ONE"). The model identifies three types of learning directed by these systems of representation. First, integer values can be recursively realized by computing the function *S* (realization learning). Second, realized integer values (stored in memory) can be calibrated against continuous magnitudes by stretching or compressing the length of the grid relative to accumulator magnitudes (calibration learning). Third, realized integer values can be mapped to a compact notation (compact notation learning). A compact notation can be learned through a natural language that has count words. At least three variants of this model are possible, in which only the symbol ONE is innate; or ONE and TWO; or ONE, TWO, and THREE are innate."

From Leslie, A.M., Gelman, R., Gallistel, C.R., 2008. The generative basis of natural number concepts. Trends Cogn. Sci. 12(6), p. 214. doi:10.1016/j.tics.2008.03.004.

fed to a layer, here called "summation," where activation and noise are linearly proportional to the number of objects. The additional feature of this model is that the read-out for a given level of activation in the summation layer is represented in another layer as a Gaussian on a number line. Notably, in the study by Verguts and Fias (2004) both summation coding and Gaussian coding emerged as a result of training, whereas in Dehaene and Changeux (1993) these network properties had been handwired.

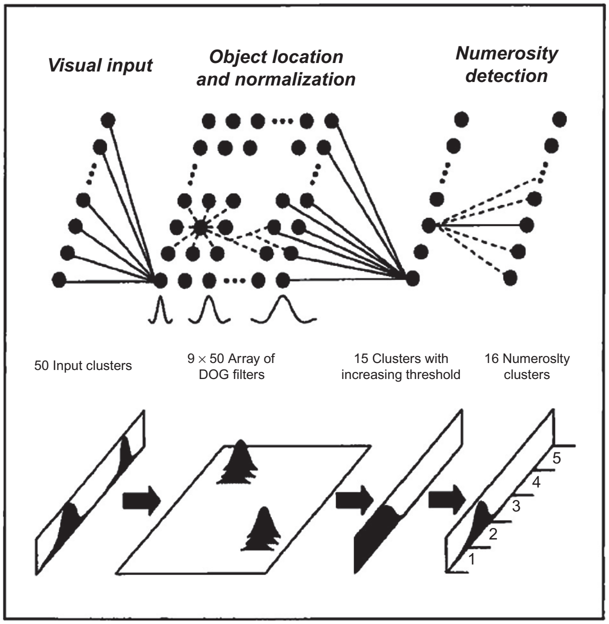

FIG. 3

A model of the numerosity detection system "Objects of different sizes at the input are first normalized to a size-independent code. Activations are then summed to yield an estimate of numerosity." The summed estimate is then read-out as Gaussian on an internal linear number line (numerosity clusters).

From Dehaene, S., Changeux, J.-P., 1993. Development of elementary numerical abilities: a neuronal model. J. Cogn. Neurosci. 5, p. 395.

From the above description, it can be seen that the model does not contradict the accumulator; it rather complements it with Gaussian read-outs. The point of deviation is, however, the view on the scaling of the Gaussians. In the original version of the model, as can be seen from Fig. 3, the magnitude "clusters" were mapped onto linear scale. The possibility of a compressive scale was briefly discussed (p. 393) but not implemented in the model. In the subsequent research, Dehaene and colleagues assume log compression of the magnitude representations (Dehaene, 2003; Izard and Dehaene, 2008; Piazza et al., 2004). Moreover, the emerging Gaussians in the study by Verguts and Fias (2004) showed a positive skew in the activity distribution, which is consistent with an idea that a "read-out" of the magnitude is represented on a compressed number line. In the remainder of the chapter, we will refer to a model that assumes a logarithmic spacing between Gaussian magnitude representations on a number line as the analogue magnitude system (AMS).

There is a clear theoretical rationale for proposing log compression for preverbal magnitude representations. First, unlike verbal counting system, preverbal system does not use positional notation in order to represent arbitrarily large magnitudes. This creates problem of numerical overload. (We will refer to an ability to handle this problem as autoscaling; Gallistel, 2011). Log scaling represents an efficient way of addressing this issue: the greater the value is the smaller system's gain is. Second, one can also assume an independence of noise characteristics from a magnitude value to satisfy scalar variability criterion—it automatically follows from first property (ie, a smaller gain for large numbers implies a greater overlap of Gaussians).

One alternative to the model proposed by Leslie et al., requires that verbal counting itself reconfigures the preverbal mechanism. The best-known and most elaborated account has been proposed by Carey and colleagues and involves a "bootstrapping"—inductive generalization—from familiarity with small sets of objects experienced in the context of counting words to exact enumeration for larger sets (Carey, 2009; Le Corre and Carey, 2007). One problem with this proposal is that the mechanism by which generalizing from Gaussian representations of magnitude, which overlap and are at best only approximate representations of numerosity, is still unclear. Because this model remains underspecified, it will not be discussed further.

2 NEURAL IMPLEMENTATION OF A PREVERBAL SYSTEM AND VERBAL COUNTING SERIES

There is now substantial evidence concerning the neural underpinnings of the preverbal and verbal magnitude representations. Substantial evidence has been accumulated to date to suggest that parietal cortex is a critical area for implementing preverbal magnitudes. For instance, individual neurons in monkey intraparietal sulcus (IPS) demonstrate behavior that can be characterized as numerically selective filtering, such that their response rates form a Gaussian-like tuning curve around a preferred magnitude (Nieder and Miller, 2003; Nieder et al., 2002). Indirect evidence for this coding schema has also been demonstrated in fMRI of human subjects, indicating that IPS provides a "read-out" of numerosity that is independent of visual cues or presentational modality (Castelli et al., 2006; Harvey et al., 2013; Piazza et al, 2004). The IPS is also likely to be a region where magnitude representations become associated with symbolic numerals as its activity is modulated by magnitudes presented in a symbolic format as much as in a nonsymbolic format (eg, Piazza et al, 2004). Evidence for summation (accumulator) coding remains rather sparse, but a few studies suggest that the activity of SPL and its homolog in the monkey's brain, the lateral intraparietal region (Sereno et al., 2001), exhibit features of a linear accumulator (Roitman et al., 2007; Santens et al., 2010).

Even though, to the best of our knowledge, formation of verbal counting sequences in the brain has never been investigated, neural mechanisms implicated in related processes has been described. Hippocampal formation has been shown to play an important role in the formation of temporal and spatial sequences (Foster and Wilson, 2006; Schendan et al., 2003), and creating semantic associations between events (Henke et al., 1999). Its neurons in area CA1 form neuronal ensembles, or cliques (Lin et al., 2005), that respond similarly to a particular attribute of an event—ranging from more generic to very specific. Their coordinated activity makes them a robust coding unit that, when considered in the context of the activity of other cliques, can be viewed as a neural implementation of a binary code (Lin et al., 2006). For example, different cliques can, respectively, encode (A) unusual events in general, (B) disturbing motion (C) shaking, and (D) dropping. The earthquake would be encoded as an increased activity of cliques A, B, and C but not D (binary code: 1110), whereas elevator drop as an activity of cliques A, B, and D but not C (binary code: 1101). These findings can be extrapolated to the symbolic numerals, which hierarchical relations in counting sequences are explicated by the use of positional notation.

3 OUR AIM

We preempt further discussion by noting certain gaps in the account of Leslie et al. that to date represents the most explicit attempt of relating a preverbal mechanism to verbal counting. To reiterate, two features—the concept of "oneness" and the recursive rule "$S(x)=x+\mathrm{ONE}$"—underlie the transition from preverbal magnitude representations to the counting series according to this account. It assumes, without argument, that these faculties originate somewhere in the brain and are readily available for use. This is equivalent to saying that verbal counting *is not* an emergent property of a preverbal mechanism; the former emerges as a result of linking two independent processes—preverbal magnitude system and a sequence-processing algorithm of an unknown origin.

Our goal here is to explore an alternative possibility that verbal counting is an emergent property of preverbal mechanism, though, of course, the counting words themselves have to be separately acquired. The remainder of this chapter will be spent in assessing the strengths and weaknesses of these models in terms of their capacity to emulate the statistical properties of verbal counting. The level of our analysis is purely quantitative. That is, we are looking for models for preverbal estimation ability that can generate outputs that are *statistically commensurate with the verbal counting process*, with the focus on whether the models can satisfy three criteria set by verbal counting process—transition to exact representations, autoscaling, and commensurability of noise characteristics. We will also outline the modified architectures that may help improving models' capacity to meet these criteria.

4 BINOMIAL ACCUMULATOR

In order to analyze the candidate models, it would be convenient to parameterize them as architectures in which building blocks are artificial neurons. A simplified version of accumulator architecture could be a circuit consisting of a "gate" neuron that signals an occurrence of a new item to be enumerated by a discharge of activity and a "storage" neuron—an accumulator—that responds to a new arrival by increasing its discharge rate. Because we are interested only in a cognitive aspect of the model, this architecture does not include a pacemaker, which characterizes physiological state of the organism. In all simulations, we assume that the external signal is not mediated by the pacemaker and enters the "gate" directly.

We first consider models that assume a linear summation of the inputs in the storage neuron and therefore they necessarily generate a linear scale of the magnitude representations. We start with an architecture that will serve as a model of verbal counting and then, through a series of adjustments to this model, we will attempt to come up with an architecture that is both statistically related to the original model and, at the same time, characterizes behavioral patterns of nonverbal magnitude processing.

We will call this target model as a binomial accumulator (see Fig. 4 and its captions for parameters of the simulation). Unlike other accumulator models considered

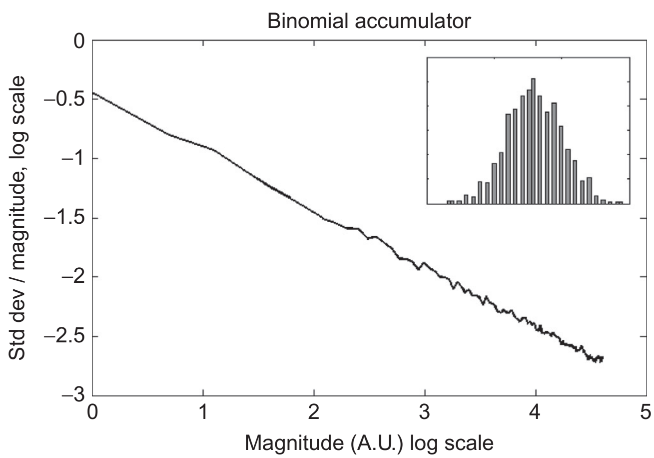

FIG. 4

The noise characteristics of binomial accumulator was simulated as a system that accumulates objects presented sequentially in an exact manner, with every additional object represented as a unit magnitude, but subject to the possibility that some objects will be not be registered (probability of omission for this plot = 0.3). The sum of objects registered by the system constitutes a magnitude accumulated over time. Here and below, the error has been estimated over 1000 simulated trials. The mechanism follows binomial distribution that has been implicated in verbal counting (Cordes et al., 2001). To satisfy scalar variability criterion, the line representing the error should be horizontal. *Inset*: the characteristic distribution of the accumulated magnitude at a particular time point.

below, the gate neuron of the binomial accumulator transmits a sequence 1's, and the accumulator neuron also precisely logs new arrivals. Omissions can nevertheless occur when some of 1's may fail to be registered; for example, when a transmission sequence is rapid. This would produce variability with respect to actual number of 1's transmitted by the gate neuron.

This model, despite its simplicity, captures main characteristics of the performance in the verbal counting paradigm as described in the study by Cordes et al. (2001). Here, participants see a number and are required to match magnitude of the number with a number of key presses. In the condition when counting of presses was allowed, the behavioral signatures of the participants' performance indicated a linear scaling and binomial variability. The latter implies that the standard deviation of noise is proportional to a square root of the mean. On a log–log plot, showing a ratio of a standard deviation to a mean as a function of a mean, this relation would be reflected by a line with a negative slope. It can be seen that the binomial accumulator replicates this pattern (Fig. 4). The additional observation is that the distribution is roughly symmetrical around mean (Fig. 4, inset).

5 POISSON ACCUMULATOR

Next, we can consider the behavioral results from the articulatory suppression condition in which participants have to repeat "the" rapidly in order to prevent verbal counting (Cordes et al., 2001). This restriction had a distinctive effect on the behavioral profile. The scaling of magnitudes remained linear, but the noise distribution obeyed scalar variability, ie, the performance becomes noisier with error standard deviation being proportional to the mean.

As we noted previously the latter feature represents an indispensible property of approximate number processing (Izard and Dehaene, 2008), and the design of a candidate model for preverbal magnitude mechanism should be able to mirror this sort of behavior. On a log–log plot such relation would be reflected by a line with a zero slope. In order to approximate this result, we are making the following adjustments to the binomial accumulator model. We keep a "gate → storage" architecture of the accumulator unaltered but allow the "gate" neuron to emit parcels of activity drawn from a Poisson distribution instead of a binomial distribution. We will refer to this architecture as a Poisson accumulator (see Fig. 5 and its captions for the parameters of the simulation). This adjustment is motivated by the fact that neurons in the brain do not use the binary code; the signal is propagated using Poisson-like discharge (Ma et al., 2006; Shadlen et al., 1996; Softky and Koch, 1993). As the results of the simulation show, the modification has no effect on the error pattern: the relation between error and mean continues to obey a square root rule. This is not particularly surprising considering that the binomial accumulator represents a special case of the Poisson accumulator (ie, the latter processes a wider range of values, not just 1's).

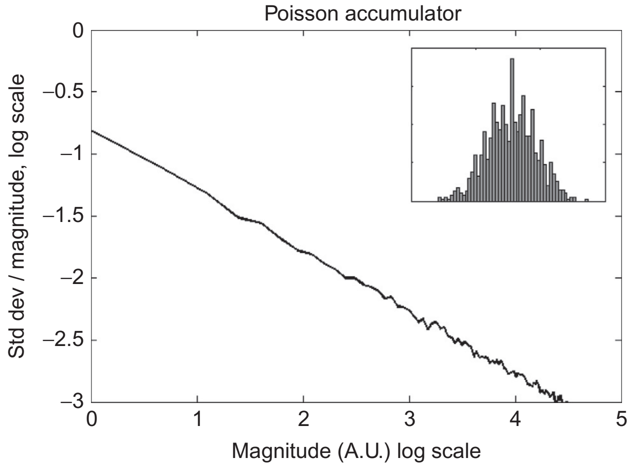

FIG. 5

Noise characteristics of Poisson accumulator. Unlike binomial accumulator that processes a sequence of 1's, the "gate" neuron of Poisson accumulator emits particles of energy, which magnitude is drawn from a Poisson distribution. To generate the plot, Poisson rate λ was set to 5. *Inset*: characteristic distribution of the accumulated magnitude at a particular time point.

6 DOUBLY STOCHASTIC PROCESS

The above demonstration shows that the linear accumulator model does not provide an unambiguous solution to account for the variability pattern observed in nonverbal magnitude processing. The architecture is actually better suited for modeling the variability pattern in verbal counting. As this has been well realized by earlier theorists (Gibbon et al., 1984), in order to meet the criterion of scalar variability, accumulator model must implicate additional source of variability. We will refer to this as a doubly stochastic process (Churchland et al., 2011).

Several proposals for the second source have been discussed in literature, for example, a multiplicative effect of noise in memory, whereby the variability of the accumulator is accentuated by the volatility of memory traces (Gibbon, 1992). Here, we emulate the process by assuming that, in addition to noise expected given a particular accumulator rate (which is Poisson), the other source of variance comes from trial-to-trial fluctuations of accumulation rate per se. The choice of gate rate as a source of variability is motivated by neurophysiological evidence from studies of neural basis of perceptual decision making (Churchland et al., 2011), but not necessary the only scenario that would give rise to scalar variability (see Gibbon et al., 1984; Killeen and Taylor, 2000). It can be seen (Fig. 6) that the error pattern of this process is close to matching the scalar variability criterion (the line is approximately

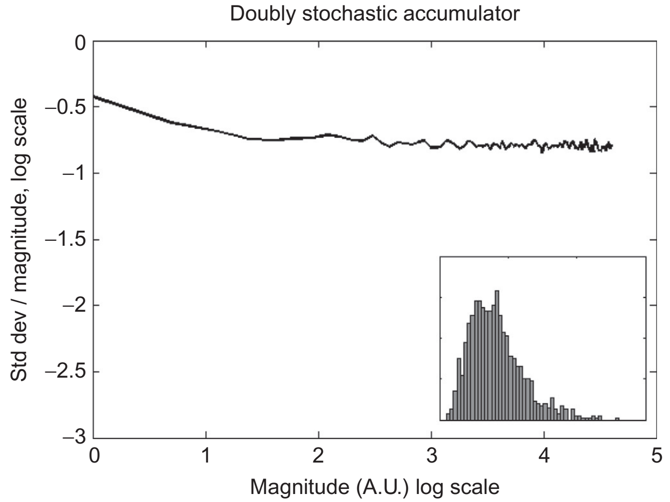

FIG. 6

Noise characteristics of a doubly stochastic accumulator was simulated by drawing a sample from the gamma distribution (shape parameter = 5, scale parameter = 1) and using this value to initialize gate rate (Poisson parameter λ) for each simulated trial. *Inset*: the characteristic distribution of the accumulated magnitude at a particular time point.

horizontal). There is an additional feature that emerges in this process—the distribution of the error is not symmetrical but is positively skewed (see Fig. 6, inset).

7 IMPLICATIONS OF LINEAR ACCUMULATOR MODELS

One can learn several things about the workings of the linear accumulator model from the above simulations. First, binomial and Poisson (including doubly stochastic) accumulators are statistically related constructs in terms of scaling and noise characteristics and verbal counting can be characterized as a special case of the accumulator model, the one that uses a binary code. Second, the accumulator model that meets scalar variability criterion requires more than one source of error, which generically can be characterized as the fluctuations in the states of the generative model (as opposed to variability of model output given a particular state); the digression to the binomial-like pattern of variance in the accumulator amounts to "switching off" the variability in the secondary source of error. Finally, it turns out that a linear doubly stochastic process is able to generate positively skewed distributions. This observation is of a particular importance given that on more than one occasion the asymmetry of a distribution was utilized as a decisive argument in favor of log-scale representations (Nieder and Miller, 2003; Piazza et al., 2004; Verguts and Fias, 2004). Here, this feature emerges as result of a linear process (also see Ratcliff, 1978).

One can, however, note substantial shortcomings of the linear accumulator models if considers the information processing mechanisms that they are supposed to model. First, the variability in the verbal counting stems from input omissions, ie, from the failures to map an item to a subsequent member of the symbolic number sequence. This is not the case for a Poisson accumulator. Its imprecision is not (or at least not only) a result of the failure to register the input; the primary source of the error is a noisy code. Second, a linear Poisson accumulator is unable to represent orders of magnitude, which means it is unable to represent arbitrary large magnitudes without an overload. Consequently, the utility of the model in serving as a precursor for verbal counting is contingent on its ability to autoscale (Gallistel, 2011).

8 NUMERICAL CONSEQUENCES OF THE AMS HYPOTHESIS

An efficient way of implementing the autoscaling principle is of course log compression. However, before we consider architectures that are consistent with the AMS hypothesis of approximate number, we would like to highlight several numerical issues that are associated with accumulator-to-(log) AMS architecture, proposed in computational studies by Dehaene and Changeux (1993) and Verguts and Fias (2004).

Even neural network theorists would have to acknowledge that AMS functions as a map in their networks—a summary statistic for the inputs acquired through the work of a linear accumulator. There are no signs that the summation layer in their networks responds to a stimulus number in any but a perfectly linear way. That creates an internal contradiction for a claim (eg, Verguts and Fias, 2004) that an accumulator-to-AMS architecture produces a logarithmically scaled numerical code.

One way to circumvent this issue is to propose that summation is performed in a log space. This is, perhaps, a bad solution. If an accumulator has to summate N individual objects in a log space, then it would need to be able to summate N logs of 1, that is, summate N zeros. The output of summation would always be a zero, irrespective how big N is. More generally, the sum of logs is not equal to a log of a sum, ie, the final result for an accumulated activity in the log space would not comply with a log hypothesis.

The other possibility is that a conversion from a linear to log scheme may be due to a transfer function from the accumulator to the AMS. In other words, even though the accumulator operates linearly, there may be reasons that the map or read-out represents the results of summation as their log. This possibility is perfectly valid, but its ramifications are less attractive: such model would loose the ability to autoscale because it relies on the processing layer that linearly increments its activity. From a computational perspective, there is no an added value in assuming a log scaling for AMS Gaussian. Such system still requires a solution how to autoscale a linear accumulator.

9 UTILITY OF AMS HYPOTHESIS

Despite the complications outlined earlier, there are two facts that motivate further exploration of the AMS hypothesis.

First, the description of variance structure is somewhat more parsimonious for the model. The scalar variability emerges naturally from the design of the system—only an assumption of a log scale is required. In other words, if one could measure AMS noise directly, one would find that system's internal noise at any time point scales well with scalar variability criterion. The performance measured at a particular trial would in essence represent an independent drawing from a noise distribution with constant parameters. Unlike linear accumulator model it does not require a secondary source of variability to account for behavioral observations.

The second fact is that in a typical paradigm studying approximate number a subject would either (a) be presented briefly with an array of dots, that varies in its low-level features, such as dot size, making then difficult to use for numerosity estimation, or (b) be required to estimate without counting a sequence of visual or auditory signals. This brings forth a crucial distinction between two processes, which is not always acknowledged (but see, for example, Castelli et al., 2006; Nieder et al., 2006). In the first case a subject is required to pool numerical information over space, whereas in the second case information is pooled over time. The comparison of findings from two paradigms (Cordes et al, 2001; Izard and Dehaene, 2008; Whalen et al., 1999) suggests that even though there is a strong case for sequential number being represented on a linear scale, there are also reasons to believe that simultaneous numerosities, or "a visual sense of number," as it is referred to in Burr and Ross (2008), is represented logarithmically. These two factors, model parsimony and the distinction between simultaneous and sequential number, motivate attempts to build a case for AMS.

10 AMS INTEGRATOR

Before we attempt to build what we call an AMS accumulator, a question can be asked whether the process of extracting simultaneous numerosity should necessarily be characterized as a numerical process; does the visual sense of number involve enumeration of the dots at all? The studies by Gebuis and Reynvoet demonstrating that simultaneous numerosity judgments are biased by a continuous quantity, such as overall dot area (Gebuis and Reynvoet, 2012a,b) suggest that it may not necessarily be the case—simultaneous number may represent a second-order statistic for continuous quantities that correlate with number.

To highlight consequences of this view, one can consider a simple case. Let us suppose that Visual Number = Density*Area. Given that the subjective scale for both density and area can be assumed to be log, the numerosity computation becomes formally equivalent to a weighted integration problem, ie, $N = w_d D + w_a A$, where D, A, and N are estimates of, respectively, density, area, and number, all in log space, and w_d and w_a are respective weights. Keeping log scale of number is an important part of these computations, because it allows for summation of relevant stimulus features. In other words, AMS would act here as an integrator, at which a number representation becomes realized for the first time.

11 BUILDING AN AMS ACCUMULATOR

An alternative to the above could be to construct AMS as an accumulator where a log-like scaling would emerge in natural way from its dynamics. One probable solution is motivated by recurrent networks and, more specifically, by a recent simulation study of Wang and colleagues, showing the emergence of different timescales from a hierarchy of structural connections (Chaudhuri et al., 2015); hierarchically lower units demonstrated transient responses, whereas the higher association areas accumulate inputs over time and demonstrate a persistent activity. Furthermore, a computational study by Howard et al. (2014) proposes that neuronal ensembles comprising neurons with varying time constants can enable a reconstruction of temporal and positional sequences in memory, thereby linking this architecture to computations that may take place in hippocampus.

To emulate the behavior of an accumulator system with the above dynamics, we assume a two-unit hierarchy, with units differing in their timescales and connected sequentially such that the unit with a short time constant receives the direct input from the environment and provides an input to the unit with longer time constant. A simplest way to implement the differences in the timescale is to represent each unit as a "leaky" accumulator, with the (hierarchically) first unit leaking more rapidly than the second unit. The leaking rate here serves as a proxy for the differences in the timescales.

Fig. 7 shows the results. The activity rate of the first unit with a transient activity, following a short burning-in period oscillates around a certain value. In many ways its behavior is comparable to the behavior of a "gate" neuron in the conventional accumulator models. The second unit, which "leaks" at a slower rate, continues to grow for a longer period and its trajectory demonstrates a nonlinearity reminiscent of a log function. The architecture can clearly handle the overload issue. If magnitudes grow to extreme values, the system may no longer be able to register new values and therefore its response rate saturates. Such behavior has a good ecological relevance. If a display of dots (for simultaneous numerosity) contains a large numerosity, it is plausible that some of the items may fail to attract the attention or receive very little representational space. However, the noise of the system violates the criterion of scalar variability. Interestingly, the violation occurs even if one implements a secondary source of variability, similarly to what we have done in the case of the doubly stochastic accumulator (see inset in Fig. 7).

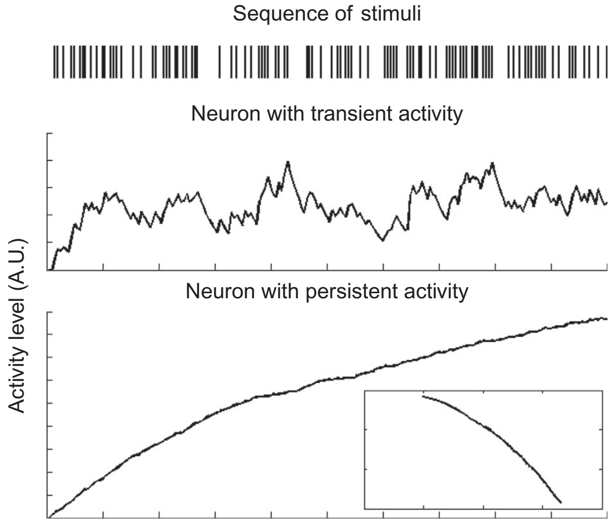

FIG. 7

"Leaky" accumulator comprising a circuit of two neurons with different activity timescales. Scales are linear. The neuron with a transient activity provides inputs to the neuron with a persistent activity. The gain in a "leaky" accumulator at a time t was simulated using dynamic equations $dX_t = I - kX_{t-1}$, where I is Poisson distributed ($\lambda = 5$). The "leak" rate k was equal to 0.1 and 0.01 for the neurons with short and long time scales, respectively. *Inset*: log–log plot of noise characteristics as a function of magnitude.

12 AMS ACCUMULATOR OR AMS INTEGRATOR?

The utility of the computational models, including the simple models considered in this chapter, is twofold: (a) they summarize, with a varying degree of complexity, intuitions about unobserved internal variables and (b) one can use them as a generative process in order to match actually observed behavioural patterns. In this sense, they are extremely useful tools for generating predictions for experimental testing. Unfortunately, in the domain of cognitive research on number, with an exception of the brilliant work by Dehaene and colleagues (eg, Piazza et al., 2004; Izard and Dehaene, 2008) and Gallistel and colleagues (eg, Cordes et al., 2001, 2007), the model-based approach to the experimental design has not been fully exploited.

Meanwhile, this approach allows addressing queries that otherwise would be difficult to answer. To provide an example, several studies have shown that precision of the visual sense of number predicts the level of mathematical achievements in school (Halberda et al., 2008). By contrast, a study by the author of this chapter suggests that this association is not unique—mathematical achievements are also associated with processing a variety of visual features (Tibber et al., 2013). Hypothetically, even showing that numerosity is a better predictor would not settle the argument, because it is not counterintuitive to argue that the proficiency in math is linked not so much to the ability to accurately discriminate continuous magnitudes as to the ability to integrate information from several continuous magnitude modalities (as alleged by AMS integrator model).

The problem can be addressed from a different angle. It is evident that the AMS accumulator and integrator presume rather different generative processes. The two hypotheses can be pitted against each other using a standard numerical discrimination task (Bahrami et al., 2013; Cappelletti et al., 2013; Halberda et al., 2008). AMS accumulator has an indispensible temporal dimension. The parallel processing of numerosity is restricted to the few items in the subitizing range; therefore, at least a partly serial process should be utilized in order to accumulate items over visual space. This is not the case for AMS integrator whose task is only to combine information from various sources. Consequently, if an accumulator-like mechanism is implicated in the processing of simultaneous number, a prediction would be that an increase of an overall number of dots in the display should lead to an increase of RT, as more time may be needed to pool a larger number of inputs together.

12.1 METHOD

12.1.1 Participants

Forty-eight adult individuals (mean age = 23.9 (3.2), 24 male) participated in the experiment. Most of them were the students in the National Chengchi University, Taiwan.

12.1.2 Stimuli

In every trial, two sets of blue and yellow dots were presented. The participants were asked to judge as fast as possible whether there were more blue dots or more yellow dots. The ratio of number of dots between blue and yellow were 2 (2:1), 1.33 (4:3), 1.2 (6:5), and 1.14 (8:7), with a total number of dots in a display varying between 11 and 30. Yellow dots were more numerous in a half of the trials. The cumulative area and dot size were controlled by equaling the total number of blue pixels to the total number of yellow pixels in half of the trials and by equaling the size of the average blue dots to the size of the average yellow dots in the other half. The total length of the experiment was 320 trials.

12.1.3 Analysis

Using the linear mixed-effect regression model, we assessed whether the reaction times would be affected by a total number of dots independently from the numerical ratio between two sets of dots. The model included a total number of dots, a ratio between numerosities of the two sets and their interaction as predictors and mean subject RT per each unique combination of the ratio and the total number of dots as a dependent variable. Prior to averaging, RTs were logarithmically transformed as they demonstrated considerable positive skewness. The interindividual variability was accounted for by using participants as a random factor in order to group model intercept.

12.2 RESULTS AND DISCUSSION

The analysis demonstrated that RT decreased with an increase in ratio (t (524) = 16.03, $p < 0.001$, $\beta = -0.53$, CI = [−0.59 −0.46]), consistent with expectations that stimuli that are easier to differentiate would require shorter decision times. Contrary to predictions for accumulator mechanism, RT *decreased* as the total number of dots increased (t (524) = 2.13, $p = 0.034$, $\beta = -0.0052$, CI = [−0.0088 −0.0003], Fig. 8). There was no significant interaction between two factors, $t < 1$. To verify that the effect was driven by the total number of dots in two sets as opposed to the numerosity of either smaller or larger sets or both, we compared the above model to the models that included numerosities of smaller set or larger set or both as predictors instead of the total of numbers of dots. These alternative models demonstrated a worse fit (on the basis of Bayesian Information Criterion) to the data than the original model.

To summarize, the results of the earlier analysis demonstrate that, even for a relatively small range of dots, the RT does not increase with a number of dots in the display. In fact, the reaction times decreased. This result fails to support the hypothesis that the process of dot accumulation takes place in this task. One might argue, though this requires further study, that the decrease in RT occurred because more dots mean more information is available to the perceptual mechanism, hence the easier it is to make the discrimination (also see Burr et al., 2010; Vetter et al., 2008, on the role of attention in numerosity comparison and estimation).

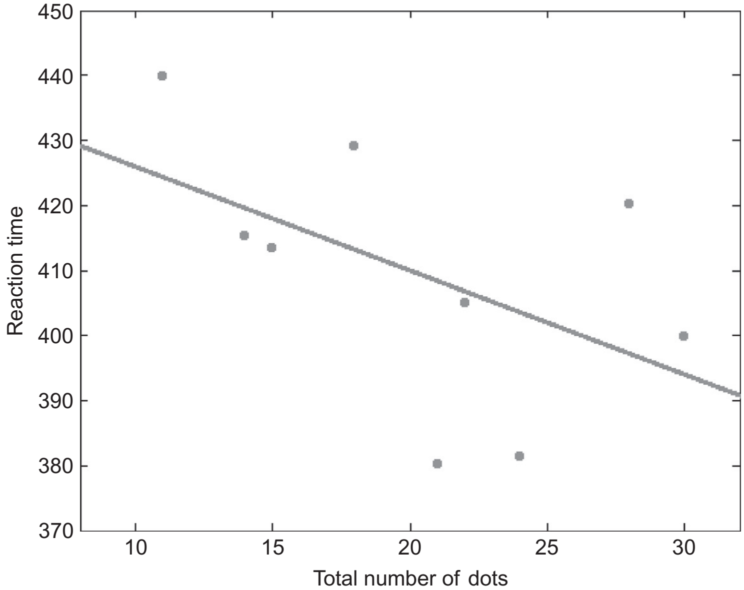

FIG. 8

The average response times as a function of the total of number dots residualized with respect to the ratio between two sets of dots.

13 REPRESENTATIONS OF MAGNITUDE ORDERS: STOCHASTIC CASCADES

The question that we will attempt to answer in the end is whether we can build a linear system that would also be able to address the issue of autoscaling. After all, the linear accumulator appears to have attractive statistical features, which bears similarity with the verbal counting process, most notably its linear scale. One solution to the autoscaling problem for orders of magnitude is to implement a "compact notation"—for example, the familiar positional notation or the hierarchical binary coding system exemplified by neuronal ensembles in the hippocampus. After all, "oneness," which is an important concept for Leslie et al. model, is just a special case of a magnitude order.

A simple schema for tackling the issue of autoscaling is to imagine a hierarchy of neurons, for which the labels "gate" or "storage" would depend on a reference point. If we consider a particular neuron, it can be seen as a gate for a hierarchically higher neuron in the circuit and it would be a storage neuron otherwise. (Even for the models considered above, the first unit in a sequence can be considered as a "storage" neuron with respect to external stimulation.)

The model works as follows. The parcels of energy that are emitted by the gate can be seen as a content temporarily stored in the neuron—up to the moment when it passes its content to the next neuron, whereupon it resets to zero (=the "gate" closes). As has been argued by Killeen and Taylor (2000), this system, which they brand a "stochastic cascade," is perfectly suited for the task of being a neural version of, say, a decimal number system. The work of such accumulator with three hierarchical levels is shown in Fig. 9. As it also can be seen (Fig. 9, inset) the variability pattern is similar to that observed for a Poisson accumulator. This is not surprising, considering that the latter is just a special case of a stochastic cascade. In other words, for a system to satisfy a scalar variability criterion, at least one of the units in this architecture should demonstrate doubly stochastic behavior.

Why can a hierarchical accumulator serve as a key link to developing exact representations? This is because at some order of a magnitude, the noise associated with stimulus processing becomes too small to affect this order. For example, there maybe subjective uncertainty about how many dots are presented on a screen, say, 20 or 30. This uncertainty would make a strong impact on the state of the first-order accumulator that is tuned to minute fluctuations of stimulus values. However, these fluctuations would make a much less impact on the state of the second-order accumulator and virtually no impact on the state of the third-order accumulator. For this architecture, "exact" would mean "above the noise levels associated with processing stimulus parameters."

14 LOG VS LINEAR: IS THIS AN ISSUE (FOR LEARNING)?

It has been argued (eg, Dehaene, 2003) that the scale of the accumulator-to-AMS architecture can be considered automatically compressive if noise increases with an increase in numerosity. Behavioural consequences of log scale with constant error

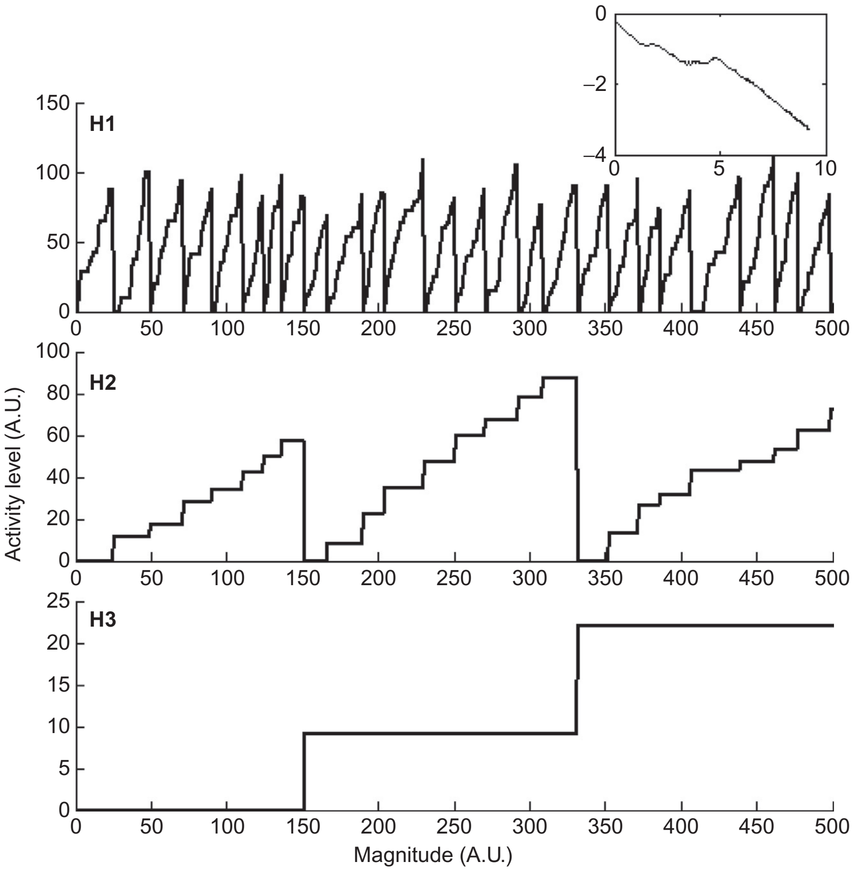

FIG. 9

Stochastic cascade. The process was simulated as a circuit of accumulators where each subsequent neuron in the chain increased its firing rate by a value drawn from Poisson distribution ($\lambda = 10$), subject to the condition that the input from a preceding neuron passes a threshold, at which case the activity of the preceding neuron is reset to zero. The threshold was implemented using a logistic function. H1, H2, and H3 stand for the 1st, 2nd, and 3rd hierarchical levels, respectively. Each level is a representation of a notional magnitude order, analogously to units, decades, and hundreds of the decimal number system. *Inset*: log–log plot of the error as a function of magnitude mean.

and linear scale with error obeying scalar variability are essentially identical; therefore, the selection of a model for the behavior is a matter of convenience rather than theoretical necessity. A few studies however have shown that the claim that two hypotheses generate identical predictions is not accurate (Gibbon and Church, 1981; Karolis et al., 2011). This becomes evident as soon as a response to stimuli requires performing (perhaps, implicitly) arithmetical operations (Karolis et al., 2011).

A transition from a log to linear scale is also not straightforward if viewed from the perspective of a learner. Learning to count from log AMS would effectively means learning to use exponentiation. Leaving aside doubts that a child can perform this operation, we are still left with the problem that an antilog of an integer is normally not an integer, ie, it is not countable.

The above intuition seems to contradict the experimentally observed data. It has been shown that the performance of young children on a task that maps number symbols onto a physical line with its end numerically defined—eg, 0–100—is best described by a subjective magnitude scale that, according to Siegler and colleagues, is log compressed (Siegler and Opfer, 2003). However, not all children show this, and those that do, quickly learn to mark the line as if their subjective scale of magnitudes is linear not log (Iuculano et al., 2008). In any case, it is not clear that one can infer from the performance on this task the nature of the subjective scale of magnitudes: there will be a cascade of cognitive processes between the subjective scale and the external performance (Barth and Paladino, 2011; Karolis et al., 2011).

15 CONCLUSIONS

In a lifespan perspective, we should consider that the infant comes into the world with a mechanism that provides crucial support for the long process of learning to count. Our review of statistical facts associated with various accumulator and AMS architectures indicate that linear accumulator architecture, taken as a model for preverbal magnitude mechanism, possesses distinct characteristics that can enable development of verbal counting. Specifically, its error and scaling characteristics are commensurate with a model of verbal counting, the one we called a binomial accumulator. A hierarchically organized circuit of accumulators is able to implement the discrete order of magnitudes, and hence, the ONE in the model of Leslie et al. (2008).

ACKNOWLEDGMENTS

We would like to thank Han Cheng-Ching and Yen Nai-Shing (National Chengchi University, Taiwan), and Daniele Didino (Tomsk Polytechnic University, Russia) for granting us a permission to analyze the data set in Section 12.

REFERENCES

Bahrami, B., Didino, D., Frith, C., Butterworth, B., Rees, G., 2013. Collective enumeration. J. Exp. Psychol. Hum. Percept. Perform. 39 (2), 338–347. http://dx.doi.org/10.1037/a0029717.

Barth, H.C., Paladino, A.M., 2011. The development of numerical estimation: evidence against a representational shift. Dev. Sci. 14, 125–135. http://dx.doi.org/10.1111/j.1467-7687.2010.00962.x.

Burr, D.C., Ross, J., 2008. A visual sense of number. Curr. Biol. 18, 425–428.

Burr, D.C., Turi, M., Anobile, G., 2010. Subitizing but not estimation of numerosity requires attentional resources. J. Vis. 10 (6). http://dx.doi.org/10.1167/10.6.20.

Cappelletti, M., Gessaroli, E., Hithersay, R., Mitolo, M., Didino, D., Kanai, R., et al., 2013. Transfer of cognitive training across magnitude dimensions achieved with concurrent brain stimulation of the parietal lobe. J. Neurosci. 33 (37), 14899–14907. http://dx.doi.org/10.1523/jneurosci.1692-13.2013.

Carey, S., 2009. Where our number concepts come from. J. Philos. 106 (4), 220–254.

Castelli, F., Glaser, D.E., Butterworth, B., 2006. Discrete and analogue quantity processing in the parietal lobe: a functional MRI study. Proc. Natl. Acad. Sci. U.S.A. 103 (12), 4693–4698.

Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H., Wang, X.-J., 2015. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. Neuron 88 (2), 419–431. http://dx.doi.org/10.1016/j.neuron.2015.09.008.

Churchland, A., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., Shadlen, M.N., 2011. Variance as signature of neural computations during decision-making. Neuron 69, 818–831.

Cordes, S., Gelman, R., Gallistel, C.R., Whalen, J., 2001. Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. Psychon. Bull. Rev. 8 (4), 698–707.

Cordes, S., Gallistel, C.R., Gelman, R., Latham, P., 2007. Nonverbal arithmetic in humans: light from noise. Percept. Psychophys. 69, 1185–1203.

Dehaene, S., 2003. The neural basis of the Weber–Fechner law: a logarithmic mental number line. Trends Cogn. Sci. 7, 145–147. http://dx.doi.org/10.1016/S1364-6613(03)00055-X.

Dehaene, S., Changeux, J.-P., 1993. Development of elementary numerical abilities: a neuronal model. J. Cogn. Neurosci. 5, 390–407.

Foster, D.J., Wilson, M.A., 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440, 680–683.

Gallistel, C.R., 2011. Mental magnitudes. In: Dehaene, S., Brannon, E.M. (Eds.), Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought. Academic Press, London.

Gebuis, T., Reynvoet, B., 2012a. The interplay between nonsymbolic number and its continuous visual properties. J. Exp. Psychol. Gen. 141 (4), 642–648. http://dx.doi.org/10.1037/a0026218.

Gebuis, T., Reynvoet, B., 2012b. The role of visual information in numerosity estimation. PLoS One 7 (5), e37426. http://dx.doi.org/10.1371/journal.pone.0037426.

Gelman, R., Gallistel, C.R., 1978. The Child's Understanding of Number, 1986th ed. Harvard University Press, Cambridge, MA.

Gibbon, J., 1992. Ubiquity of scalar timing with a Poisson clock. J. Math. Psychol. 36, 283–293.

Gibbon, J., Church, R.M., 1981. Time left: linear versus logarithmic subjective time. J. Exp. Psychol. Anim. Behav. Process. 7, 87–108.

Gibbon, J., Church, R.M., Meck, W.H., 1984. Scalar timing in memory. Ann. N. Y. Acad. Sci. 423, 52–77.

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

Harvey, B.M., Klein, B.P., Petridou, N., Dumoulin, S.O., 2013. Topographic representation of numerosity in the human parietal cortex. Science 341, 1123–1126.

Henke, K., Weber, B., Kneifel, S., Wieser, H.G., Buck, A., 1999. Human hippocampus associates information in memory. Proc. Natl. Acad. Sci. U.S.A. 96 (10), 5884–5889. http://dx.doi.org/10.1073/pnas.96.10.5884.

Howard, M.W., MacDonald, C.J., Tiganj, Z., Shankar, K.H., Du, Q., Hasselmo, M.E., Eichenbaum, H., 2014. A unified mathematical framework for coding time, space, and sequences in the hippocampal region. J. Neurosci. 34 (13), 4692–4707. http://dx.doi.org/10.1523/jneurosci.5808-12.2014.

Iuculano, T., Tang, J., Hall, C., Butterworth, B., 2008. Core information processing deficits in developmental dyscalculia and low numeracy. Dev. Sci. 11 (5), 669–680.

Izard, V., Dehaene, S., 2008. Calibrating the mental number line. Cognition 106 (3), 1221–1247.

Karolis, V., Iuculano, T., Butterworth, B., 2011. Mapping numerical magnitudes along the right lines: differentiating between scale and bias. J. Exp. Psychol. Gen. 140 (4), 693–706. http://dx.doi.org/10.1037/a0024255.

Killeen, P., Taylor, T., 2000. How the propagation of error through stochastic counters affects time discrimination and other psychophysical judgments. Psychol. Rev. 107, 430–459.

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. http://dx.doi.org/10.1016/j.cognition.2006.10.005.

Leslie, A.M., Gelman, R., Gallistel, C.R., 2008. The generative basis of natural number concepts. Trends Cogn. Sci. 12 (6), 213–218.

Lin, L., Osan, R., Shoham, S., Jin, W., Zuo, W., Tsien, J.Z., 2005. Identification of network-level coding units for real-time representation of episodic experiences in the hippocampus. Proc. Natl. Acad. Sci. U.S.A. 102 (17), 6125–6130. http://dx.doi.org/10.1073/pnas.0408233102.

Lin, L., Osan, R., Tsien, J.Z., 2006. Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes. Trends Neurosci. 29, 48–57. http://dx.doi.org/10.1016/j.tins.2005.11.004.

Ma, W.-J., Beck, J.M., Latham, P.E., Pouget, A., 2006. Bayesian inference with probabilistic population codes. Nat. Neurosci. 9, 1432–1438.

Meck, W.H., Church, R.M., 1983. A mode control model of counting and timing processes. J. Exp. Psychol. Anim. Behav. Process. 9 (4), 320–334.

Meck, W.H., Church, R.M., Gibbon, J., 1985. Temporal integration in duration and number discrimination. J. Exp. Psychol. Anim. Behav. Process. 11, 591–597.

Nieder, A., Miller, E.K., 2003. Coding of cognitive magnitude. Compressed scaling of numerical information in the primate prefrontal cortex. Neuron 37, 149–157.

Nieder, A., Freedman, D.J., Miller, E.K., 2002. Representation of the quantity of visual items in the primate prefrontal cortex. Science 297, 1708–1711.

Nieder, A., Diester, I., Tudusciuc, O., 2006. Temporal and spatial enumeration processes in the primate parietal cortex. Science 313 (5792), 1431–1435.

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.

Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108.

Roitman, J.D., Brannon, E.M., Platt, M.L., 2007. Monotonic coding of numerosity in Macaque lateral intraparietal area. PLoS Biol. http://dx.doi.org/10.1371/journal.pbio.0050208.

Santens, S., Roggeman, C., Fias, W., Verguts, T., 2010. Number processing pathways in human parietal cortex. Cereb. Cortex 20 (1), 77–88. http://dx.doi.org/10.1093/cercor/bhp080.

Schendan, H.E., Searl, M.M., Melrose, R.J., Stern, C.E., 2003. An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. Neuron 37 (6), 1013–1025. http://dx.doi.org/10.1016/S0896-6273(03)00123-5.

Sereno, M.I., Pitzalis, S., Martinez, A., 2001. Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. Science 294, 1350–1354.

Shadlen, M.N., Britten, K.H., Newsome, W.T., Movshon, J.A., 1996. A computational analysis of the relationship between neuronal and behavioral analysis to visual motion. J. Neurosci. 16, 1486–1510.

Siegler, R.S., Opfer, J.E., 2003. The development of numerical estimation: evidence for multiple representations of numerical quantity. Psychol. Sci. 14, 237–243.

Softky, W.R., Koch, C., 1993. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J. Neurosci. 13, 334–350.

Tibber, M.S., Manasseh, G.S.L., Clarke, R.C., Gagin, G., Swanbeck, S.N., Butterworth, B., et al., 2013. Sensitivity to numerosity is not a unique visuospatial psychophysical predictor of mathematical ability. Vision Res. 89, 1–9. http://dx.doi.org/10.1016/j.visres.2013.06.006.

Verguts, T., Fias, W., 2004. Representation of number in animals and humans: a neural model. J. Cogn. Neurosci. 16 (9), 1493–1504.

Vetter, P., Butterworth, B., Bahrami, B., 2008. Modulating attentional load affects numerosity estimation: evidence against a pre-attentive subitizing mechanism. PLoS One 9 (e 3269), 1–6. http://dx.doi.org/10.1371/journal.pone.0003269.

Whalen, J., Gallistel, C.R., Gelman, R., 1999. Nonverbal counting in humans: the psychophysics of number representation. Psychol. Sci. 10 (2), 130–137. http://dx.doi.org/10.1111/1467-9280.00120.