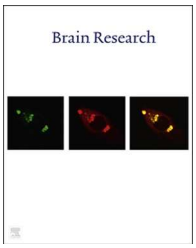
Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres



Research Report

A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network

Rakesh Sengupta[a], *Bapi Raju Surampudi*[a,c,*], *David Melcher*[b]

[a]*Center for Neural and Cognitive Sciences (CNCS), University of Hyderabad, Hyderabad 500046, India*
[b]*Center for Mind/Brain Sciences (CiMEC), University of Trento, Trento 38068, Italy*
[c]*School of Computer and Information Sciences, University of Hyderabad, Gachibowli, Hyderabad 500046, India*

ARTICLE INFO

Article history:
Accepted 9 March 2014
Available online 6 August 2014

Keywords:
Visual sense of numbers
Computational model
On-center off-surround
Neural network
Enumeration
Spatial attention
Individuation
Numerical cognition

ABSTRACT

It has been proposed that the ability of humans to quickly perceive numerosity involves a visual sense of number. Different paradigms of enumeration and numerosity comparison have produced a gamut of behavioral and neuroimaging data, but there has been no unified conceptual framework that can explain results across the entire range of numerosity. The current work tries to address the ongoing debate concerning whether the same mechanism operates for enumeration of small and large numbers, through a computational approach. We describe the workings of a single-layered, fully connected network characterized by self-excitation and recurrent inhibition that operates at both subitizing and estimation ranges. We show that such a network can account for classic numerical cognition effects (the distance effect, Fechner's law, Weber fraction for numerosity comparison) through the network steady state activation response across different recurrent inhibition values. The model also accounts for fMRI data previously reported for different enumeration related tasks. The model also allows us to generate an estimate of the pattern of reaction times in enumeration tasks. Overall, these findings suggest that a single network architecture can account for both small and large number processing.

1. Introduction

According to the theory of a visual sense of number (Burr and Ross, 2008a, 2008b), the ability to rapidly estimate the numerosity of a set of items reflects a basic, perceptual process. Developmental studies have shown that infants show an ability to distinguish between different numerosities at a young age (Xu and Spelke, 2000). Studies of non-human animals, such as cotton-top tamarins (Hauser et al., 2003), point towards the possible evolutionary origins of the visual

*Correspondence to: Current address: Cognitive Science Lab, International Institute of Information Technology (IIIT), Gachibowli, Hyderabad 500032, India.
E-mail addresses: qg.rakesh@gmail.com (R. Sengupta), bapiks@yahoo.co.in (B.R. Surampudi), david.melcher@unitn.it (D. Melcher).

http://dx.doi.org/10.1016/j.brainres.2014.03.014
0006-8993/

number sense. It has also been shown that making decisions based on numerosity is possible even in societies where language does not have words for distinguishing larger numbers (Gordon, 2004). The visual number sense also seems to distinguish between a range of smaller numbers (the subitizing range: Kaufman et al., 1949) and the range of larger numbers. Numbers in the subitizing range are characterized by their rapid and confident enumeration with very high degree of accuracy. For larger numbers humans use one of the two strategies – (1) precise sequential counting, or (2) estimation, a rapid but inexact process of enumeration. Counting is precise but is much slower than either subitizing or estimation. Estimation, although faster than counting, is much less precise (Whalen et al., 1999). An important question, then, is whether subitizing and estimation share a common mechanism and thus possibly a common substrate.

To even begin to understand how to disambiguate the answer we have to address certain complexities. In the visual domain, it has been argued that enumeration can be linked to object individuation, a visuo-spatial mechanism that allows us to locate and track a limited set size of objects (Piazza et al., 2011; Melcher and Piazza, 2011). In contrast, another approach has been to characterize subitizing as estimation mechanism operating at small numbers (Gallistel and Gelman, 1992; Dehaene and Changeux, 1993). However, the different Weber fractions[1] over the two ranges suggest a fundamental distinction based on small or large numerosities (Revkin et al., 2008). Trick and Pylyshyn (1994) have suggested subitizing, unlike estimation, might employ pre-attentive mechanisms that index potential objects. Burr et al. (2010) have systematically manipulated both spatial and temporal attention to show that such manipulations indeed affect subitizing performance but not estimation, which led them to suggest that in addition to a possible pre-attentive mechanism that is active across all numerosities there is an additional attentive mechanisms necessary for enumeration within subitizing range.

In order to explore possible computational strategies for enumeration Dehaene and Changeux (1993) developed a model using a reinforcement-based supervised learning approach (with a proposed extension towards self-organization) to explain possible learning mechanisms in infants. However this model restricts enumeration only up to five items. As another possible solution Stoianov and Zorzi (2012) have trained 'deep' networks to use pixel by pixel information of images through unsupervised learning. They show that numerosity detectors can emerge in the highest level of the generative network. Their model could account satisfactorily for numerosity comparison task data in monkeys and human adults in the larger number (estimation, not subitizing) range. These numerosity detectors can be compared to the 'number neurons' reported by Nieder et al. (2002) and Roitman et al. (2007). Both the models mentioned above have complex structures that allow for learning of numerosity detectors. Another interesting model was developed by Grossberg and Repin (2003), based on an on-center off-surround architecture. This model involved an interaction between a spatial number map and semantic categories to explain error rates and reaction times in human numerosity comparison data. In the present work we wanted to investigate what properties a network of numerosity detectors should have in order to account for enumeration performance across both small and large numbers of items.

Building on work from Roggeman et al. (2010), we constructed a recurrent on-center off-surround network that receives a normalized pre-processed input and the output is the mean steady state activity of the network. This kind of network has been used to describe different kinds of phenomena in the domain of vision and working memory (Grossberg, 1973; Usher and Cohen, 1999). However any computational account of numerosity has to pass a few crucial tests: (a) the model simulations should qualitatively demonstrate how the different regimes of numerosity (subitization and estimation) emerge through internal network dynamics; (b) the model should be able to account for empirical findings like numerosity comparison data in human adults; and (c) it should be possible to make testable predictions from the model. In the following sections we will try to show how the proposed model fares under these conditions.

One advantage of our approach is that it does not start with dedicated number neurons that activate to specific number estimates, but attempts instead to see whether such capacities might emerge from a recurrent on-center off-surround network used more generally in perception of objects and scenes. A second aspect of our approach is that, rather than trying to fit the data with a model based on a large number of parameters, the proposed network has fewer number of parameters and the same network has been used to fit disparate data sets. Finally, we are able to generate novel testable predictions that emerged from the model itself regarding the link between sensory processing and numerical cognition.

2. Results

2.1. Model

We began by modeling critical features of a recurrent on-center off-surround network reported in Usher and Cohen (1999) and used in Roggeman et al. (2010). It is essentially a saliency map model based on the nonlinear leaky competing accumulation models (LCA) that have been used to account for performance in multiple-alternative choice paradigms (Bogacz et al., 2007). These models capture the recurrent on-center and off-surround nature of activity observed in neural systems (Grossberg, 1973; von der Malsburg and Buhmann, 1992) particularly in the visual modality.

The network consists of a single layer of completely interconnected nodes (Fig. 1). Each node corresponds to a neuronal assembly encoding an object or location of an object (or particular features) depending upon the cognitive phenomenon being modeled. The three main parameters that define the type of network are α (strength of self-excitation

[1]Weber fraction or Weber ratio refers to the minimum relative change in stimulus intensity in order for the stimulus level to be perceived as different from a reference stimulus intensity.

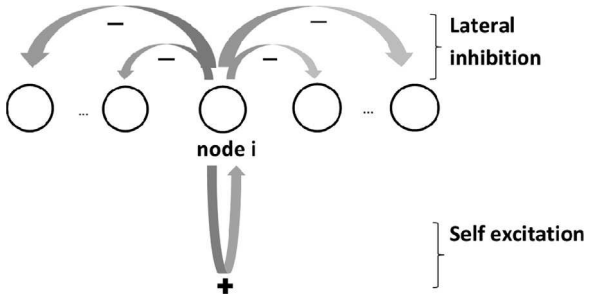


Fig. 1 – Illustration of the on-center off-surround model, showing nodes with excitatory and inhibitory connections.

Table 1 – Simulation parameters.

Parameter	Value
N	70
α	2.2
β Range	0.01–0.15
Duration of stimulus presentation (no. of time steps)	5
Total duration of simulation in time steps	50
Number of simulations run to compute the average	100

for each node), β (strength of lateral inhibition between nodes) and λ (decay constant for the passive decay term).

The differential equation governing the time-evolution of the network of N nodes is given by

$$\frac{dx_i}{dt} = -\lambda x_i + \alpha F(x_i) - \beta \sum_{j=1, j\neq i}^{N} F(x_j) + I_i + noise \quad (1)$$

$x_i(t)$ is the activation of node i at time t. I_i represents the intensity of external input ($\forall i,\ 0 \le I_i \le 1$). In our simulation I_i is a unit step function, i.e., it has the value 1 for certain number of time steps for the particular node i and 0 for rest of the time steps. Input is only presented for a finite amount of time, typically much less than total time of simulation. $F(x)$ is the activation function given by the formula

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{x}{1+x} & \text{for } x > 0 \end{cases} \quad (2)$$

The total number of nodes receiving input is referred to as set size in rest of the article. In order to remain close to cortical neuronal dynamics we chose to keep the decay parameter $\lambda = 1$.

We modeled the dynamics according to the discrete form of Eq. (1). The activation of the nodes is updated at each step according to the following equation:

$$x_i(t) = \alpha F(x_i(t-1)) - \beta \sum_{j=1, j\neq i}^{N} F(x_j(t-1)) + I_i + noise \quad (3)$$

Following the example of Roggeman et al. (2010), we used a network of 70 nodes ($N = 70$) for the present work as higher values of N do not change the results qualitatively. All the simulations were run at fixed $\alpha = 2.2$.[2] The input was presented for 5 time steps and the simulations were run for a total of 50 time steps. A normal distribution of mean 0 and standard deviation 0.03 is used to sample the *noise* at each time step. As can be seen in Eq. (13) in Appendix A, that in the absence of external input the network settles to steady state dynamics characterized by its steady state activation $x(n)$, where n is the number of nodes in steady state activity. This quantity is uniquely determined by the network parameters α, β, and n (but not on the size of the network, i.e., N). So we use mean activation ($\bar{x}(n) = \sum x_i/N$)[3] as the main output of the network for the remainder of the paper. Table 1 lists the major parameters used for simulation throughout the study.

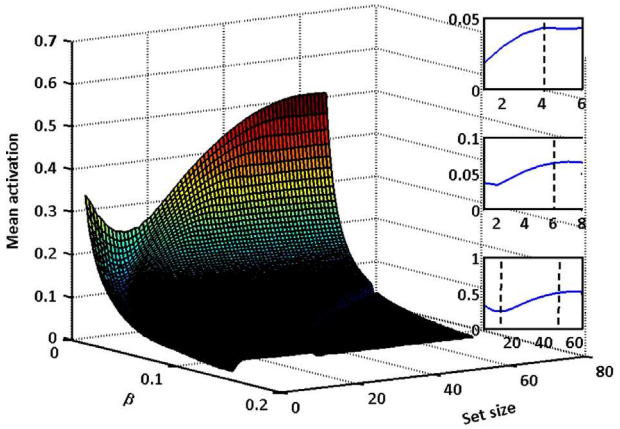


Fig. 2 – Mean activation vs. set size and β. The plot was derived at $\alpha = 2.2$ as an average of 100 simulations. The insets show the mean activation vs. set size plot for three particular β values – low: 0.01 (bottom inset), medium: 0.1 (middle inset), and high: 0.15 (top inset). The vertical line in the insets denotes the limits to the monotonic region. For instance we can see from the top inset that the mean activation ($\bar{x}(n)$) is monotonically increasing up to set size 4 for high β, but for low β (bottom inset) we can see that $\bar{x}(n)$ is monotonically increasing with numerosity for set sizes greater than 15. The dashed lines in the insets provide the limits of the monotonic region in the insets.

2.2. *Subitizing and estimation from network output at different inhibition levels*

Considering that the transient input is converted into steady state activity for the current network, we wanted to see the sensitivity of the network output, i.e., the mean activation ($\bar{x}(n)$) with the parametric variation of β and set size. We considered that the parameters suitable for numerosity estimation would be the ones that give a monotonically increasing variation of the mean activation with the increase in numerosity of the inputs. The plot of mean steady state activation of the network as a function of the set size and the

[2]The excitation parameter was chosen in line with Usher and Cohen (1999). The excitation parameter was varied systematically between 2.0 and 2.4 to check for stability. The convergence results are shown in Appendix C.

[3]Here x_i is the final activity of the node i at the end of simulation. It is either 0 or close to the value given in Eq. (13) in Appendix A. Or written in another way, $\bar{x}(n) = nx(n)/N$ in the absence of noise, where $x(n)$ is the steady state activation values of the nodes as given by Eq. (13).

inhibition parameter is given in Fig. 2. The input was clamped at level 0.33[4] for this and subsequent simulations. Other simulation parameters are given in Table 1.

It is evident that at medium inhibition the total activation of the network is sensitive to a smaller range values of numerosity (see the lower plot in inset of Fig. 2) while low inhibition seems to function better for larger numerosities (range of 15–40, see the upper plot in inset of Fig. 2). Moreover, the plot shows that at higher values of β the total activation of the network is fairly insensitive to larger numerosity and at low β the network activation does not monotonically increase as a function of numerosity for smaller numbers. The finding, that different inhibition parameters work best for representing different numerosities, confirms the previous report of Roggeman et al. (2010).

2.3. Truthfulness of representations

Although lower β-values are more sensitive to larger numerosities, the precision of the representation is compromised at lower inhibition values. In order to illustrate the point, we can construct a crude measure for the probability of faithfulness of representation ($P(\beta, setsize)$) as

$$P(\beta, setsize) = 1 - d_{i/o} \quad (4)$$

where $d_{i/o}$ is the mean Hamming distance[5] between input activation pattern and output activation pattern over N nodes parametrized over inhibition parameter β and *setsize*. The simulation parameters are the same as in Table 1. Fig. 3 illustrates that at lower inhibition the maximum value of $P(\beta, setsize)$ is around 0.6, although lower inhibition is better for estimation of larger numerosities (see Fig. 2). At lower β enumeration should be more prone to errors and inaccuracies due to more diffused representation. This finding is in line with numerous behavioral studies indicating that precision is reduced for estimation of large numerosities.

It is interesting to note that, from a purely representational point of view, Fig. 3 differentiates three types of phenomena – (1) at high inhibition (β=0.15), focused representation of a single object, (2) at medium inhibition ($\beta \sim 0.1$), specialized subitizing and (3) at low inhibition, estimation. All of these behaviors are emergent from the self-organized behavior of the network.

2.4. Numerosity comparison

In order to simulate human numerosity comparison data, we calculated the probability that a given numerosity would be judged to be 'larger' than a given reference numerosity (16, in our case). The probability of responding 'larger' is calculated by running 100 simulations for numerosity values 10–24 for different β values (the parameters of the network are the same as in Table 1), as the proportion of times the mean

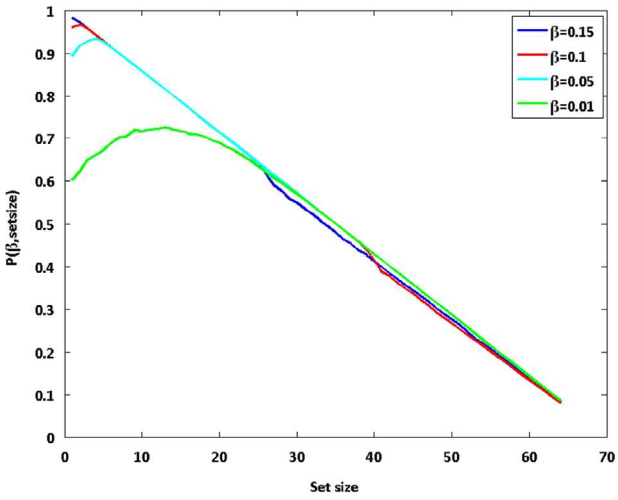


Fig. 3 – $P(\beta, setsize)$ plotted against *setsize* for different β. The modeling parameters are given in Table 1. The figure shows how probability of faithful representation decreases with decreasing β.

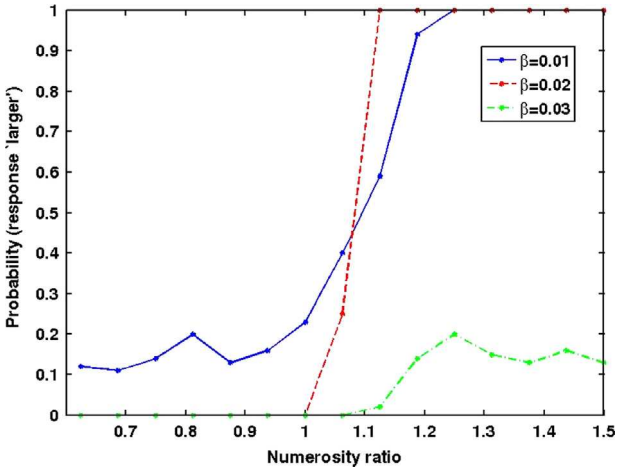


Fig. 4 – Probability of finding a given numerosity larger than the reference numerosity 16 plotted against numerosity ratios.

activation of the network for a given numerosity is larger than the mean activation of the network for the reference numerosity by threshold value (δ=0.01). In other words the probability that numerosity m will be perceived 'larger' than a predefined reference numerosity *ref* is given by

$$Prob_m(\text{response larger}) = \frac{Number\ of\ times\ \{\bar{x}_m(n) - \bar{x}_{ref}(n)\} > \delta}{Number\ \text{of simulations}} \quad (5)$$

where $\bar{x}_{ref}(n)$ and $\bar{x}_m(n)$ are mean activation values of the network when they are presented with numerosities *ref* and m respectively.

As shown in Fig. 4, performance differs depending on the β values. However, if we calculate the probability as an average over different β values (0.01, 0.011, 0.012 and 0.03), and perform a sigmoid fit against the numerosity ratio, we get a Weber fraction of 0.14 calculated at 75% performance value (r^2=0.93 at 95% confidence level), which matches closely with human data collected by Piazza et al. (2004) (see Fig. 5).

[4]Clamping here is analogous to electrophysiological practices where a constant d.c. like current with a certain amplitude is applied to the neuron/synapse for a finite time. Here the input amplitude is 0.33.

[5]The distance between two binary vectors that is equal to the number of bits that do not match between the vectors. For example, the Hamming distance between 010101 and 011111 is 2.

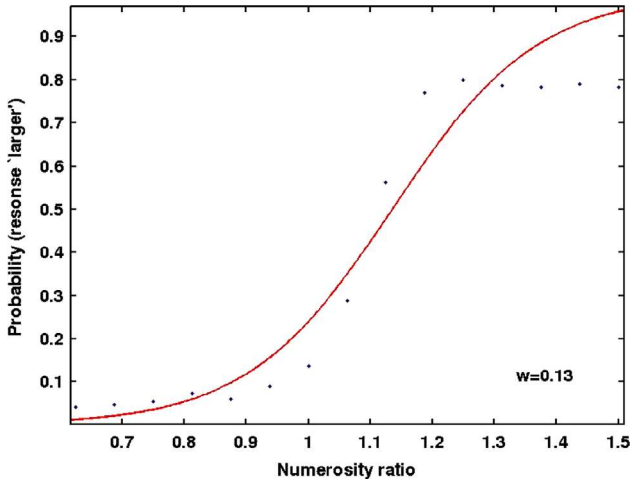


Fig. 5 – Simulation of human numerosity comparison data by averaging the probability values at different β (0.01, 0.011, 0.012, 0.03).

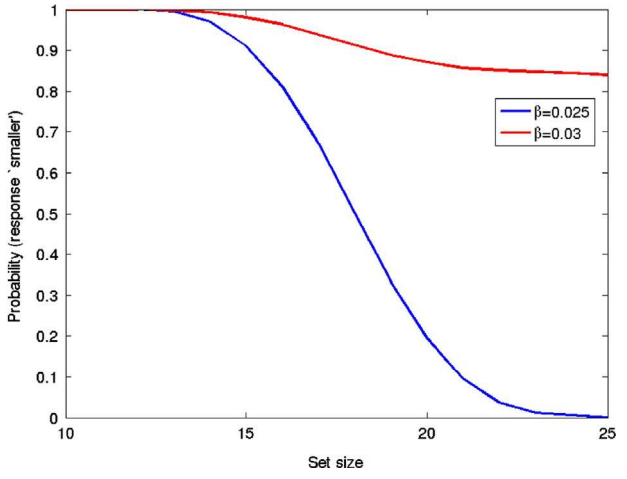


Fig. 6 – Probability of 'smaller' response for different numerosities for reference numerosity 16. The plot shows how at higher inhibition numerosity underestimation is possible.

Interestingly, at higher β values the probability of response 'larger' decreases and probability of response 'smaller' (i.e., 1 – probability of response 'larger') increases (see Fig. 6) showing that it might be possible that when comparing numerosities at higher β, one might see underestimation.

For smaller numerosities, we wanted to see if our model shows the classic *distance effect* and follows *Fechner's law* (Dehaene and Changeux, 1993). Distance effect refers to the finding that performance is better when the distance between the numbers to compare increases. Fechner's law is an extension of Weber's law, and states that perceptual judgments of magnitude are proportional to the logarithm of the stimulus intensity. Similar to Eq. (5) we calculate the discrimination probability of each numerosity $i=1$–5 against the reference numerosity $ref=1$–5 as

$$Prob_i(\text{response different}) = \frac{Number\ of\ times\ |\bar{x}_{ref}(n)-\bar{x}_i(n)|>\delta}{Number\ \text{of simulations}} \quad (6)$$

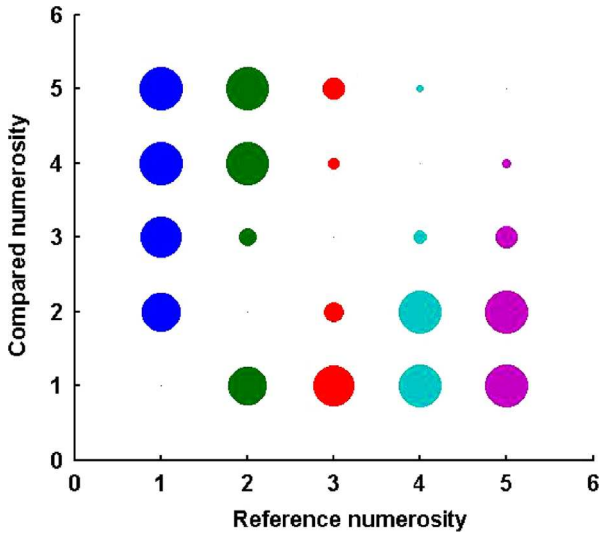


Fig. 7 – We show pairwise discrimination probability between different numerosities 1–5. The probability is proportional to the areas of the circles at each point. The largest area is 1 and the smallest ones adjacent to top right corner being 0.03. As we move towards the top right corner the circle sizes decrease, i.e., the overall precision of discrimination suffers with increasing numerosity. As one moves away from the diagonal in figure, discrimination performance increases, demonstrating *distance effect* where numerosity discrimination performance is better if the compared numerosities are more distant. And the discrimination probability between 1–2, 2–3, 3–4 and 4–5 pairs show a decrease in precision for higher numerosities consistent with *Fechner's law* (see text for more details).

where the symbols have the same meaning as above. As we are operating in the subitizing range we chose to calculate the probability as an average of $\beta=0.06$, 0.08, 0.1, 0.14–0.15 ($\delta=0.01$) for this simulation. Fig. 7 represents the $Prob_i$ (response different) matrix for each reference numerosity (1–5) and compared numerosity (1–5), the area of the circles representing the magnitude of the probability (largest circles at top left and bottom right have value 1). Observing the probability values adjacent to the diagonal line, we can see that the overall precision decreases with increasing numerosity (i.e., from 1 to 5). Whereas for same numerical distance (like 2–3, 3–4, 4–5 pairs), we see that the probability for discrimination is higher at lower numerosity. The results indicate that performance is better when the difference between the numbers to compare is larger (*Distance Effect*) and the decrease in precision observed for higher numerosities is consistent with *Fechner's Law* (although results in Fig. 7 conform to Fechner's logarithmic scaling, they are not shown here). Thus Fig. 7 illustrates that the *distance effect* and *Fechner's law* emerge from the model.

2.5. *Empirical evidence for the inhibition parameter*

In line with the results of Melcher and Piazza (2011), the behavior of this on-center off-surround model depends on the inhibition between nodes (Fig. 1). In the case of high

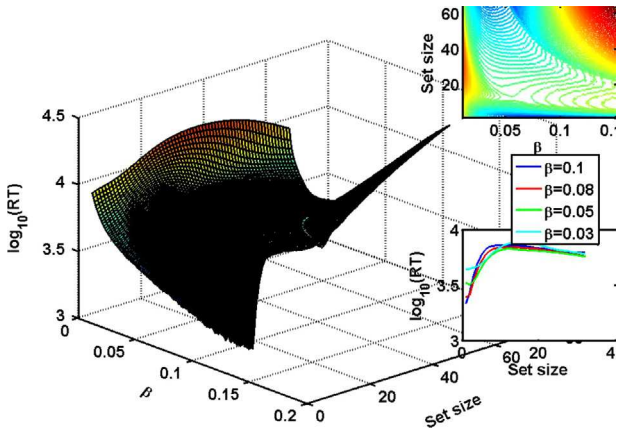


Fig. 8 – A parametric display of the reaction time distribution as calculated from the model. The values are meant to be proportional to log(RT) rather than directly representations. The top inset shows the contour plot for the energy surface and bottom inset shows the prediction for the reaction times for set sizes 1–32 at $\beta=0.03$, $\beta=0.05$, $\beta=0.08$ and $\beta=0.1$.

inhibition (top inset of Fig. 2), the pattern resembles that reported previously in fMRI studies of visual working memory, with activation increasing up to around 4 items (Todd and Marois, 2004, 2005; Kawasaki et al., 2008). Varying the inhibition parameter (β) results in different patterns of activation. The graph for medium inhibition (middle inset of Fig. 2) successfully models the activity of fMRI data for enumeration tasks Roggeman et al. (2010). Here the activation flattens out at about 6 items. For low inhibition, the prediction is shown as a concave curve shown in the bottom inset of Fig. 2. The model's prediction of a concave curve is well supported by data published in Roggeman et al. (2010). Their data was collected for the lateral intraparietal (LIP) area, and thus provide a hypothesis for localization of such a numerosity network in the brain that is also consistent with other numerosity related findings (e.g., Roitman et al., 2007).

2.6. *Estimation of reaction times from the model*

In order to explore the parametric space of the network we modeled an energy function on the following lines (see Appendix B)

$$H=\sum_i H_i \propto -\sum_i \int \left(1-\alpha\left(\frac{F(x_i)}{x_i}\right)^2\right)\dot{x}_i^2 dt \qquad (7)$$

$\dot{x}_i = dx_i/dt$ and H_i is the energy for a particular node i.

The reaction time distribution is assumed to correlate with the total amount of allowed fluctuation energy at a given state (as shown in Appendix B), i.e.,

$$RT \sim \sum_i \int \left(1-\alpha\left(\frac{F(x_i)}{x_i}\right)^2\right)\dot{x}_i^2 dt \qquad (8)$$

The simulated values for *RT* were obtained as an average of 100 simulations. The modeling parameters are given in Table 1. Fig. 8 shows the *RT* values for a given inhibition parameter and set size for a network with 70 nodes. It seems to be consistent with the general idea that subitizing (enumeration of small numbers) is faster than estimation (enumeration of large numbers).

However, this simulation also offers some interesting features when considering the mean activation value of the network. From Fig. 2 we can see that higher inhibition ($\beta\sim0.1$) is better for representing small numerosity ranges including the subitizing range, while lower inhibition ($\beta\sim0.03$) is more suited for larger numerosity estimation. Fig. 8 (bottom inset) shows the estimated *RT* values for the different levels of inhibition (low or high). It shows a nearly linear increase in reaction times for numerosities in the subitizing range. The reaction times for the large numerosity ranges remain very similar for a wide range of β value. This effect has been well known and demonstrated by Kaufman et al. (1949). The *RT* values for enumeration in small numerosity range at low β tend to be higher. So if a network tries to enumerate smaller numerosity at low inhibition, then the reaction time should be higher than reaction time for enumerating the same at higher inhibition.

3. Discussion

3.1. *The model*

In the current work we presented a single layer recurrent neural network with complete interconnectivity between nodes that cooperate and compete with each other through mechanisms of self-excitation and lateral inhibition after being presented with finite transient normalized input. Once the transient input is taken away, the network settles down to a steady state activity ($x(n)$) governed by the strength of lateral inhibition (β) between nodes. The network's behavior can be explored by varying β and observing values of $x(n)$ for different sizes of input (set sizes).

We have used mean activation ($\bar{x}(n)$) as the main output of the network in order to parameterize it against different β and set sizes. The condition that mean activation of the network should increase monotonically with set size gave us two regimes of set sizes governed by two different β values – higher β for smaller numbers and lower β for larger numbers (see Fig. 2).[6] The measure for the probability of faithfulness of representation ($P(\beta, setsize)$) shows, that although lower β is needed for enumeration of larger numbers, the representation becomes coarser and thus more prone to errors. Moreover, we saw that although $P(\beta, setsize)$ is almost linear for larger numerosities, it is near constant for smaller numerosities, thus indicating a possible reason that precision levels for enumeration differ significantly between the subitizing and the estimation range (see Fig. 3).

We estimated the probability that a particular numerosity would be judged as 'larger' than a given reference numerosity (16, in our case) on the basis of the fraction of times the mean

[6]Here we have used the additive variant of on-center off-surround activity rather than the shunting one (where excitation terms are multiplied by $B-x$ and inhibitory terms are multiplied by $x-C$, where B and C are excitation and inhibition parameters, see Grossberg (1973), as the additive function gives us the monotonic property of the mean activation of the network with numerosity unlike the shunting variant.

activation $\bar{x}(n)$ for that particular network is larger than that of the reference numerosity. Empirical data observed by Piazza et al. (2004) shows that for reference numerosity 16, numerosity comparison data could be fitted to a sigmoid curve corresponding to Weber fraction (w) of 0.15. Although one particular β could not explain the spread of the data, using an average of different β values allowed for a good fit ($w=0.14$, $r^2=0.93$ at 95% confidence level) to human numerosity comparison data collected by Piazza et al. (2004) (see Fig. 5). At higher β, the model could successfully capture (see Fig. 7) the *distance effect* (the same-different judgment performance is better when the distance between the numbers to compare increases) and *Fechner's law* (perceptual judgments of magnitude being proportional to the logarithm of the stimulus intensity).

Finally, the mean activation values for different set sizes were consistent with fMRI response data. Roggeman et al. (2010) also presented results showing a different level of inhibition β involved in numerosity comparison as opposed to explicit enumeration. The β level found in their comparison experiment is line with the β values taken for modeling numerosity comparison data. From these experiments LIP area emerges as a possible neural substrate for the proposed network.

3.2. Comparison with other models

There have been limited attempts at a unified computational approach towards numerosity. This is due to the fact that human performances in subitizing and estimation ranges are clearly distinct – both in terms of performance measures like Weber fractions (Burr et al., 2010), precision (Revkin et al., 2008) as well as reaction time (Kaufman et al., 1949). The differences in empirical data for the two numerosity ranges suggested the possibility of different enumeration mechanisms involved in subitizing and estimation. Dehaene and Changeux (1993) suggested a model based on supervised learning through ordered numerosity detectors and were able to demonstrate the *distance effect* and *Fechner's law* for numerosities 1–5. Stoianov and Zorzi (2012) trained hierarchical generative networks to use the statistical properties of images in order to develop numerosity detectors in an unsupervised manner at the highest level of the network. Their work was able to simulate adult human numerosity comparison data in the larger ranges of numerosity. Our model falls somewhere in-between these two extremes. Our network is unsupervised to the extent that there is no reinforcement involved. However, the network has no learning component like the previous two networks. We can say that our network demonstrates the self-organization process required for the formation of numerosity detectors that arrive through learning in a hierarchical network proposed by Stoianov and Zorzi (2012).

We proposed a network that responds to numerosity as a normalized input and we used the output steady state activity to characterize the network response. Interestingly, it removes a limitation of the networks proposed by Stoianov and Zorzi (2012) and Dehaene and Changeux (1993). The former does not address the subitizing range whereas the latter does not work well in the estimation range. Thus the earlier models seem to assume a break between estimation and subitizing ranges.

The model developed by Grossberg and Repin (2003), based on a similar on-center off-surround architecture, was used to account for data in numerosity comparison tasks and provided an explanation of how the mapping from spatial to actual linguistic number categories might take place. In the current work, the question was what mechanisms might underlie the spatial map itself, whose activation may then be read off to produce actual responses through other higher order networks (similar to the one used in Grossberg and Repin (2003)). In agreement with Grossberg and Repin (2003) and also Roggeman et al. (2010), our results show how capacity limits can emerge out of inhibitory interactions between nodes in a spatial saliency map.

In our case, the constraint that network mean activation should monotonically increase with numerosity allows the two regimes to emerge naturally on the basis of different recurrent inhibition within the same network. Task dependent fMRI response correlates well with the predicted pattern for mean activation (e.g. in Roggeman et al. (2010)), and thus gives a possible physical basis for inhibition parameter as well as indicates a possible neural substrate in the networks of lateral intraparietal cortex (LIP) common to visual–spatial attention as well as visual short-term memory networks (Gottlieb and Goldberg, 1999; Vogel and Machizawa, 2004). The network also seems to provide an interesting idea that numerosity comparison data emerges as an average response across a range of recurrent inhibition thus matching the success of Stoianov and Zorzi's network. The model also allows for variable precision between different ranges of numerosity as depicted in Fig. 3 and also able to account for *distance effect* and *Fechner's law* similar to Dehaene and Changeux (1993) (see Fig. 7).

The key feature of the present network is that it is able to give a comprehensive account of a host of observed phenomena within both subitizing and estimation range through a single layer network with one tuning parameter (β). Because of the simplicity, such network could be theoretically analyzed through energy function formulation (see Appendices A and B), which is a novel contribution of our approach. The advantage of our network is that such a generic network can not only explain previous experimental data but also give novel predictions. It can also be easily coupled with other components in order to explore learning and multi-modal interactions.

3.3. Possible mechanisms for modulating the inhibition parameter

An important feature of the model proposed here is that the network gives task-dependent response profiles that are contingent upon varying the strength of lateral inhibition β. So it is natural to ask the question regarding possible mechanism for changing inhibition. Task dependent changes might be motivated by top-down influence. However, modulation of inhibition within the same task leads to the possibility that there might be an internal mechanism for β modulation. A simple additional summator node that

receives input from all the nodes and gives feedback to the entire network based on a threshold might provide a mechanism for modulating the recurrent inhibition from within the network.

3.4. *Possible implications of modeling reaction time*

Following the success of the network in predicting behavioral performances in numerosity comparison, we tried a different approach. In order to extract new testable predictions from the model we formulated a possible expression for reaction time or its close correlate. We took averages of the *RT* values at two different inhibition ranges (high: 0.08–0.11, low: 0.03–0.06) in order to elucidate what happens to *RT* when one tries to enumerate in the small numerosity range with low β or enumerate in the large numerosity range with high β. The model predicts an asymmetry or hysteresis in the reaction times for enumeration of small and large numerosities depending on whether they were preceded by a set of small or large numerosities. We reason that the network will generally be in a low β state if it is engaged in larger numerosity enumeration tasks and in a state of higher β if it is engaged in repeated enumeration tasks with smaller numerosity. Thus if the network is presented with a small numerosity enumeration task followed by a chain of large numerosity estimation tasks, the reaction times should be higher, whereas no such significant increase in reaction time should be observed in the other direction (i.e., large numerosity estimation task followed by a set of small numerosity enumeration tasks).

To get a clear prediction, we computed the average *RT* values for low β range (0.02–0.05) and high β range (0.08–0.11). Fig. 9 clearly shows that task switch between small numerosity enumeration and large numerosity enumeration should yield asymmetric switch cost depending on whether the switch was small to large or large to small enumeration. Moreover, we saw from Fig. 6 that enumeration of large numerosity from a higher β state increases probability of underestimation. Thus we arrive at two interesting predictions for an experiment, where a series of enumeration trials where varying lengths of small numerosity enumeration blocks are interspersed with varying lengths of large numerosity enumeration trials. First, for small numerosity enumeration followed directly by large numerosity estimation trials, the model would predict higher *RT* compared to standard small numerosity enumeration trials. On the other hand, for large numerosity trials directly followed by small numerosity enumeration, we would predict more underestimation effect than standard large numerosity enumeration trials.

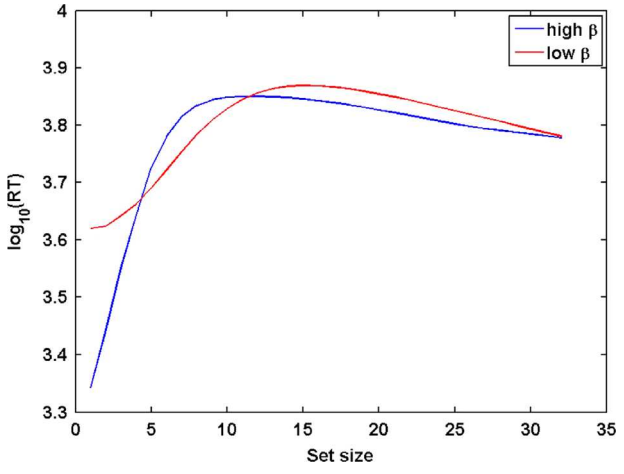


Fig. 9 – Average *RT* values for two ranges of β: high – 0.08–0.11 and low – 0.02–0.05. The larger numerosity ranges hardly differ in *RT* between the two ranges of inhibition, however the small numerosity range differs quite significantly in *RT* between the two beta ranges.

4. Conclusion

The current results have implications towards the debate regarding whether subitizing and estimation reflect unique mechanisms that utilize different neural correlates (Dehaene and Changeux, 1993; Whalen et al., 1999; Revkin et al., 2008; Melcher and Piazza, 2011). The current model shows how the same network could account for both subitizing and estimation performance. In terms of neural instantiation, a single flexible network could account for both types of performance. At a later stage, population decoding could map the pattern of activity onto approximate number. Given the close link between attention and subitizing, Burr et al. (2010) have suggested that while a shared pre-attentive mechanism might operate for both subitizing and estimation, additional attentional resources are pooled for subitizing. We suggest an alternative, in which it is the task dependent top-down modulation of the recurrent inhibition that separates the subitizing and estimation ranges. It is interesting to note that the model predicts higher energy values for higher inhibition (see Appendix B). So it might be that energy expenditure has to be made in order for the network to operate at a higher inhibition regime. And perhaps attention is needed to change to higher inhibition. Overall, the current work demonstrates how a saliency map model related to capacity limits and individuation found in basic sensori-motor processes can be used to individuate small numerosities or represent the numerosity of an ensemble of items.

Acknowledgments

This work was supported by a European Research Council Starting Grant (agreement n. 313658) to D.M. and by the India–Trento Programme for Advanced Research (ITPAR) support to RSG.

Appendix A

Stability analysis

The differential equation governing the time-evolution of the network of *N* nodes is given by

$$\frac{dx_i}{dt} = -\lambda x_i + \alpha F(x_i) - \beta \sum_{j=1, j\neq i}^{N} F(x_j) + I_i + noise \qquad (9)$$

$x_i(t)$ is the activation of node i at time t. I_i represents the intensity of external input ($\forall i,\ 0 \le I_i \le 1$), it is zero if the stimulus is absent for a particular node at that time point. Input is only presented for a finite amount of time, typically much less than total time of simulation. $F(x)$ is the activation function given by the formula

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{x}{1+x} & \text{for } x > 0 \end{cases} \tag{10}$$

The network should reach steady state activity when the external input is taken away. If we disregard noise, at steady state, i.e. when, $dx_i/dt = 0$,

$$\lambda x_i = \alpha F(x_i) - \beta \sum_{j=1,\, j \neq i}^{N} F(x_j) \tag{11}$$

As the equation is symmetric under permutation of units, the system should have symmetric solutions characterized by number of active units n, and their activation $x(n)$, all other units having 0 activation.

$$x(n) = \left(\frac{\alpha - (n-1)\beta}{\lambda}\right) F(x(n)) \tag{12}$$

Using Eq. (10), we get

$$x(n) = \left(\frac{\alpha - (n-1)\beta}{\lambda}\right) - 1 \tag{13}$$

Noise can bring in additional fluctuation that can destabilize the solution for a pair of active modes (with equal activation according to Eq. (13)), unless the difference of activations between the said nodes $\Delta x = x_i - x_j$ decays. Using Eqs. (9) and (13) we get

$$\frac{d\Delta x}{dt} = \Delta x\left[-\lambda + \lambda^2\left(\frac{\alpha+\beta}{(\alpha-(n-1)\beta)^2}\right)\right] \tag{14}$$

Thus the fluctuation decays only if $d\Delta x/dt < 0$, i.e.,

$$\frac{\alpha+\beta}{(\alpha-(n-1)\beta)^2} < \frac{1}{\lambda} \tag{15}$$

As we can see that the decay parameter, excitation parameter and inhibition parameter are not completely independent for stable solutions. For the present purposes we use $\lambda = 1$.

Appendix B

Formulation for the expression of reaction time

In order to explore the stability of the network in a more rigorous manner, we worked to derive an expression for a Hamiltonian (or energy function) for such a network from first principles. A desirable property for such a Hamiltonian should be that it exhibits properties of a Lyapunov function under suitable range of parameter choices.

According to classical mechanics for a set of generalized co-ordinates x, the time evolution according to the Hamiltonian principle is given by

$$\frac{\partial x}{\partial t} \propto \frac{\partial H}{\partial \dot{x}} \tag{16}$$

$$\frac{\partial \dot{x}}{\partial t} \propto -\frac{\partial H}{\partial x} \tag{17}$$

where H is the Hamiltonian (or the energy) of the system. In the present system the activations of the nodes can be taken as generalized co-ordinates for the system. Given Eqs. (16) and (17) we can roughly say that at any point in time evolution (considering $\lambda = 1$ and proportionality constant as 1) disregarding noise, the Hamiltonian equation for a particular node i is given by

$$dH_i \propto \frac{\partial H}{\partial \dot{x}_i} d\dot{x}_i + \frac{\partial H}{\partial x_i} dx_i = k_1 \dot{x}_i d\dot{x}_i - k_2 \ddot{x}_i dx_i \tag{18}$$

where k_1 and k_2 are proportionality constants (it is reasonable to assume that $k_1 \neq k_2$). Now from Eq. (9) we have

$$d\dot{x}_i = -dx_i + \alpha\left(\frac{F(x_i)}{x_i}\right)^2 dx_i \tag{19}$$

$$\ddot{x}_i = -\dot{x}_i + \alpha\left(\frac{F(x_i)}{x_i}\right)^2 \dot{x}_i \tag{20}$$

Inserting Eqs. (19) and (20) in Eq. (18) we get

$$dH_i = -(k_1 - k_2)\left(1 - \alpha\left(\frac{F(x_i)}{x_i}\right)^2\right)\dot{x}_i^2 dt \tag{21}$$

and the total energy

$$H = \sum_i H_i \propto -\sum_i \int \left(1 - \alpha\left(\frac{F(x_i)}{x_i}\right)^2\right)\dot{x}_i^2 dt \tag{22}$$

From Eq. (13) we can substitute terms in steady state to get

$$H \propto -\sum_i \int \left(1 - \frac{\alpha}{(\alpha-(n-1)\beta)^2}\right)\dot{x}_i^2 dt \tag{23}$$

Now if $dH < 0$ and thus a monotonically decreasing Lyapunov type function in absence of external input, we have the stability condition as

$$\frac{\alpha}{(\alpha-(n-1)\beta)^2} < 1 \tag{24}$$

Comparing this to Eq. (15), we can see that the conditions derived from the energy value is slightly different and diverges greatly for higher β. This is due to the fact that Eq. (15) excludes winner-take-all mechanisms operating at higher inhibition, whereas the energy function does not. And it is evident that for all β,

$$\frac{\alpha}{(\alpha-(n-1)\beta)^2} < \frac{\alpha+\beta}{(\alpha-(n-1)\beta)^2} < 1 \tag{25}$$

and thus the energy function is a very suitable candidate for the network as it is in line with the stability analysis derived from the dynamics of the network. However, there was an implicit assumption in the analysis presented here, mainly $k_1 > k_2$, which is justified as otherwise the system will not reach convergent solution. Interestingly as mean activation values are higher at lower β, it stands to reason that higher inhibition will also have higher energy than the lower ones.

In order to apply the equation to get to reaction time (RT) distribution, we assume that reaction times correlate with the allowed fluctuation in energy (i.e., $-\int dH$). More energy to dissipate, more the reaction time. And thus

$$RT \sim -\int dH \propto \sum \int \left(1 - \alpha\left(\frac{F(x_i)}{x_i}\right)^2\right)\dot{x}_i^2 dt \tag{26}$$

We know from Mohamed et al. (2004) and Yarkoni et al. (2009) that reaction times do correlate with highest values of

cortical activation and also be used to predict trial-by-trial variability in reaction times. We have seen from Fig. 2 and the corresponding discussion in Section 2 that experimental fMRI activation patterns correspond to the integral activity of the network, which in turn correlates with the *RT* given in Eq. (26), thus the choice of the formula for *RT* seems reasonable.

Appendix C

Effect of αand noise on convergence

Here we want to show how changes in α and *noise* affect the overall behavior of the network. In order to do so we ran simulations for mean activation $x(n)$ and log 10(RT) values

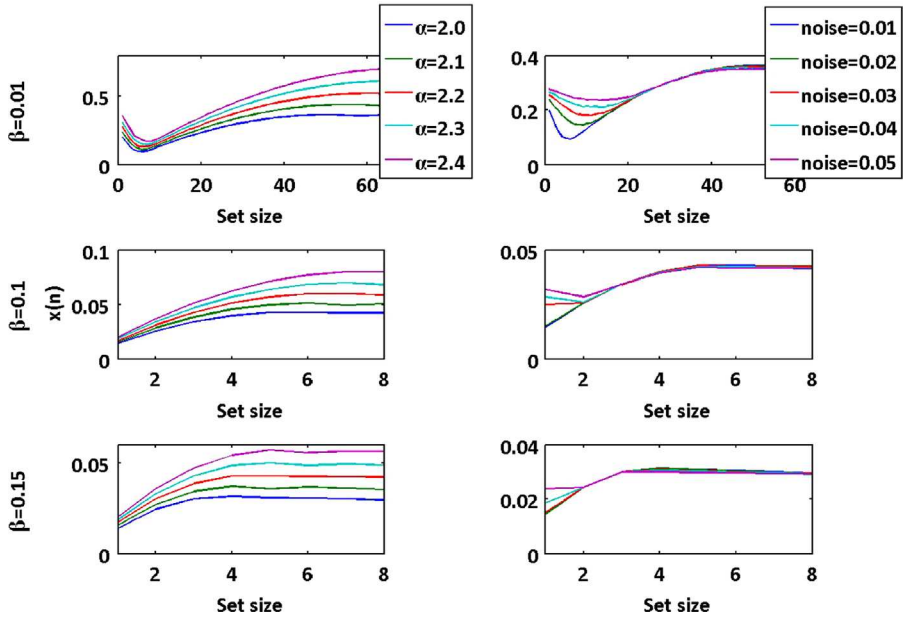


Fig. 10 – Variation in mean activation ($x(n)$) pattern plotted for three β values (0.01, 0.1, and 0.15) and their variation with changing α and *noise*. The left column shows variation of $x(n)$ with α at constant level of *noise* (standard deviation 0.01) and the right column shows the variation of $x(n)$ at constant $\alpha=2.0$ at different noise levels.

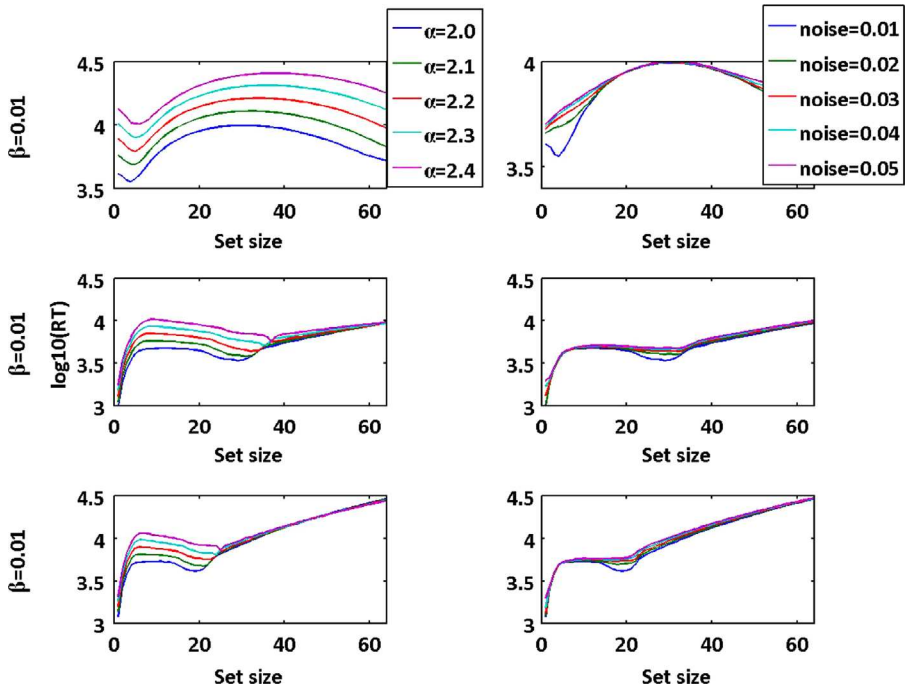


Fig. 11 – Variation in Reaction Times log 10(*RT*) pattern plotted for three β values (0.01, 0.1, and 0.15) and their variation with changing α and *noise*. The left column shows variation of log 10(*RT*) with α at constant level of *noise* (standard deviation 0.01) and the right column shows the variation of $x(n)$ at constant $\alpha=2.0$ at different noise levels.

calculated for (1) constant *noise* (sampled from normal distribution of mean 0 and standard deviation 0.01) and variable α (2.0–2.4) and (2) constant α (2.0) and variable *noise* (sampled from normal distribution with mean 0 and standard deviation 0.01–0.05). The rest of the simulation parameters remain the same as in Table 1. The results are given in Figs. 10 and 11. Figures depict how different interesting behavior emerges out of manipulation of *noise* and α. However, there is a lot of similarity in the qualitative behavior as well. So we have chosen $\alpha = 2.2$ and *noise* of standard deviation 0.03 in the main simulations reported in the paper in order to probe a wide range of behaviors while varying the least number of parameters.

REFERENCES

Bogacz, R., Usher, M., Zhang, J., McClelland, J.L., 2007. Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 362, 1655–1670.

Burr, D., Ross, J., 2008a. Visual sense of numbers. Curr. Biol. 18, 425–428.

Burr, D., Ross, J., 2008b. Response: visual number. Curr. Biol. 18, R855–R856.

Burr, D., Turi, M., Anobile, G., 2010. Subitizing but not estimation of numerosity requires attentional resources. J. Vis. 10, 1–10.

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

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

Gordon, P., 2004. Numerical cognition without words: evidence from Amazonia. Science 306, 496–499.

Gottlieb, J., Goldberg, M.E., 1999. Activity of neurons in the lateral intraparietal area of the monkey during an antisaccade task. Nat. Neurosci. 2, 906–912.

Grossberg, S., 1973. Contour enhancement, short term memory, and constancies in reverberating neural networks. Stud. Appl. Math. LII, 213–257.

Grossberg, S., Repin, D.V., 2003. A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. Neural Netw. 16 (8), 1107–1140.

Hauser, M.D., Tsao, F., Garcia, P., Spelke, E.S., 2003. Evolutionary foundations of number: spontaneous representation of numerical magnitudes by cotton-top tamarins. Proc. R. Soc. B: Biol. Sci. 270, 1441–1446.

Kaufman, E.L., Lord, M.W., Reese, T., Volkmann, J., 1949. The discrimination of visual number. Am. J. Psychol. 62, 498–525.

Kawasaki, M., Watanabe, M., Okuda, J., Sakagami, M., Aihara, K., 2008. Human posterior parietal cortex maintains color, shape and motion in visual short-term memory. Brain Res. 1213, 91–97.

Melcher, D., Piazza, M., 2011. The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. PLoS One 6, 1–11.

Mohamed, M.A., Yousem, D.M., Tekes, A., Browner, N., Calhoun, V.D., 2004. Correlation between the amplitude of cortical activation and reaction time: a functional MRI study. Am. J. Roentgenol. 183 (3), 759–765.

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.

Piazza, M., Fumarola, A., Chinello, A., Melcher, D., 2011. Subitizing reflects visuo-spatial object individuation capacity. Cognition 121, 147–153.

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

Revkin, S.K., Piazza, M., Izard, V., Cohen, L., Dehaene, S., 2008. Does subitizing reflect numerical estimation?. Psychol. Sci. 19, 607–614.

Roggeman, C., Fias, W., Verguts, T., 2010. Salience maps in parietal cortex: imaging and computational modeling. NeuroImage 52, 1005–1014.

Roitman, J.D., Brannon, E.M., Platt, M.L., 2007. Monotonic coding of numerosity in macaque lateral intraparietal area. PLoS Biol. 5, 1673–1682.

Stoianov, I., Zorzi, M., 2012. Emergence of a 'visual number sense' in hierarchical generative models. Nat. Neurosci. 15, 194–196.

Todd, J.J., Marois, R., 2005. Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. Cogn. Affect. Behav. Neurosci. 5, 144–155.

Todd, J.J., Marois, R., 2004. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428, 751–754.

Trick, L.M., Pylyshyn, Z.W., 1994. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychol. Rev. 101 (1), 80–102.

Usher, M., Cohen, J.D., 1999. Short term memory and selection processes in a frontal-lobe model, model. In: Heinke, D., Humphreys, G.W., Olson, A. (Eds.), Connectionist Models in Cognitive Neuroscience. Springer-Verlag, London, pp. 78–91.

von der Malsburg, C., Buhmann, J., 1992. Sensory segmentation with coupled neural oscillators. Biol. Cybern. 67, 233–242.

Vogel, E.K., Machizawa, M.G., 2004. Neural activity predicts individual differences in visual working memory capacity. Nature 428, 748–751.

Whalen, J., Gallistel, C.R., Gelman, R., 1999. Nonverbal counting in humans: the psychophysics of number representation. Psychol. Sci. 10, 130–137.

Xu, F., Spelke, E.S., 2000. Large number discrimination in 6-month-old infants. Cognition 74, B1–B11.

Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., Braver, T.S., 2009. BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. PLoS One 4 (1), e4257.