



PERGAMON

AVAILABLE AT

www.ElsevierComputerScience.com

POWERED BY SCIENCE @ DIRECT®

Neural Networks 16 (2003) 1107–1140

Neural
Networks

www.elsevier.com/locate/neunet

A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes

Stephen Grossberg*, Dmitry V. Repin

Department of Cognitive and Neural System and Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA

Received 12 February 2003; revised 30 May 2003; accepted 30 May 2003

Abstract

Both animals and humans represent and compare numerical quantities, but only humans have evolved multi-digit place-value number systems. This article develops a Spatial Number Network, or SpaN, model to explain how these shared numerical capabilities are computed using a spatial representation of number quantities in the Where cortical processing stream, notably the inferior parietal cortex. Multi-digit numerical representations that obey a place-value principle are proposed to arise through learned interactions between categorical language representations in the What cortical processing stream and the Where spatial representation. Learned semantic categories that symbolize separate digits, as well as place markers like ‘ty,’ ‘hundred,’ and ‘thousand,’ are associated through learning with the corresponding spatial locations of the Where representation. Such What-to-Where auditory-to-visual learning generates place-value numbers as an emergent property, and may be compared with other examples of multi-modal cross-modality learning, including synesthesia. The model quantitatively simulates error rates in quantification and numerical comparison tasks, and reaction times for number priming and numerical assessment and comparison tasks. In the Where cortical process, transient responses to inputs are integrated before they activate an ordered spatial map that selectively responds to the number of events in a sequence and exhibits Weber law properties. Numerical comparison arises from activity pattern changes across the spatial map that define a ‘directional comparison wave.’ Variants of these model mechanisms have elsewhere been used to explain data about other Where stream phenomena, such as motion perception, spatial attention, and target tracking. The model is compared with other models of numerical representation.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Number; Spatial map; Speech category; Learning; Place-value system; Inferior parietal cortex; What and Where cortical streams

1. Introduction: human and animal numerical abilities

Both animals and humans can represent and compare numbers, but only humans can represent multi-digit place-value numbers. Choosing a larger prey to hunt, a tree with more fruit, or a flower with more honey illustrates how animal survival may be enhanced by being able to estimate and compare magnitudes and quantities. Given that many animals can estimate numerical quantities, it is natural to ask how this competence arose? This article describes a Spatial Number Network, or SpaN, model whose properties suggest that estimates and comparisons of numerical magnitude use specializations of more primitive neural mechanisms that have evolved in the Where cortical processing stream for purposes of motion perception, spatial attention, and target tracking. This explanation clarifies why

an analog spatial map for numerical representation has been discovered in the Where stream, notably in the parietal cortex (Naccache & Dehaene, 2001; Rickard et al., 2000).

The model also proposes how multi-digit place value numbers have arisen in humans through a learning process that links numerical categories in the What cortical processing stream with the spatial number map in the Where cortical processing stream. In particular, this learned map links learned auditory categories for individual number names and categories for place markers like ‘ty,’ ‘hundred,’ and ‘thousand’ with locations in the spatial number map. Place-value numbers are an emergent property of this inter-modality learning process.

The human capacity for mathematical thinking and intuition has been traced to a combination of linguistic competence and visuo-spatial representations (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). The SpaN model proposes a mechanistic substrate on which such higher mathematical processes can build. Other multi-modal

* Corresponding author. Tel.: +1-617-353-7858; fax: +1-617-353-7755.
E-mail address: steve@bu.edu (S. Grossberg).

learning processes are well known to occur in the brain (Stein & Meredith, 1993). Recently, such processes have been used to explain phenomena as far ranging as synaesthesia, metaphor, creativity, and language learning (Ramachandran & Hubbard, 2001a,b). Here we note how they may play a crucial role in initiating the peculiarly human use of multi-digit number systems via a fusion of linguistic and visuo-spatial representations. These results were reported in preliminary form in Repin and Grossberg (1999a,b).

Evidence for the shared numerical estimation capabilities of animals and humans takes several forms. In particular, for both humans and animals, the processing of larger quantities becomes increasingly difficult as reflected in larger reaction times and error rates, a finding known as the Number Size effect (Dehaene, 1997). The detection of a difference between two groups of objects that differ only in amount becomes easier, as reflected by faster reaction times and less errors, as the difference between the number of objects increases, a finding known as the Numerical Distance effect (Dehaene, 1997). These similarities in animal and human performance suggest that common mechanisms may control shared numerical abilities across species, and thus that the capacity for numerical competence may be derived from more basic neural mechanisms. The SpaN model lends support to both of these hypotheses by showing that neural mechanisms that are used for motion perception, spatial attention, and target tracking in the Where cortical processing stream can be specialized to form a spatial numerical map whose dynamics can simulate error and temporal characteristics of human and animal psychophysical data. No previous model known to us has quantitatively simulated such an extensive data set, nor shown how this competence could have arisen in the brain.

Given that animals as well as humans exhibit many similar numerical estimation properties, it is important to understand what processes underlie the human superiority in forming numerical estimates. Generally speaking, human superiority in forming numerical estimates is attributable to usage of symbolic notation. If unable to use such a symbolic

notational system, humans may perform no better than animals in certain estimation and comparison tasks, and task performance of both groups is often influenced by the same factors in a similar manner. Historically, however, it was not written symbolic notation, but spoken language that brought the concept of number into human life. Long before the appearance of the first concise numerical notation, in early Sumerian language dating from the third millennium BC, number-words reflected the structure of the numerical system, as shown in Fig. 1 (after Menninger, 1969). Numerical systems were developed independently by many civilizations in different parts of the world. The Sumerians, who inhabited the southern part of Mesopotamia, based on their system on gradations of the number 60, an influence that can be seen today in how time is measured in minutes and seconds (Fig. 1, left). The Celts in Europe as well as the Maya and the Aztecs in Mesoamerica used a vigesimal, or base-20, numerical system. Modern French still bears the legacy of the base-20 that interferes with its number-naming base-10 structure. Our modern numerical competence has a decimal system in its foundation that originated from Arab and Indian cultures.

Initially, most of the number systems were based on an *additive* principle. Egyptian and Roman systems (Fig. 1, middle, right) serve as good examples of how the symbols for units or hundreds are ordered and then grouped together such that their sum represents a new symbol for ten or thousand, respectively. The additive principle allowed use of a compressed representation of large numbers, such as 2374, but this representation was not as compressed and convenient for calculations as the modern number system based on a *multiplicative* principle. In a number system based on a multiplicative principle, maximum compression is achieved by means of place-values. Instead of having a new symbol for each of the powers of ten, as in the case of Egyptian hieroglyphs, the power of ten is encoded by its place information. Such a system was used by Babylonians as early as about 2000 BC, with only one principal

Sumerian			Egyptian		Roman	
Value	Cuneiform symbol	Number word	Value	Symbol	Value	Symbol
1		<i>aš</i>	1		1	I
10		<i>u</i>	10		10	X
60		<i>geš</i>	100		50	L
6010		<i>geš-u</i>	1000		100	C
60 ²		<i>šar</i>	10000		500	D
60 ² 10		<i>šar-u</i>	2374		1000	M
60 ³		<i>šar-gal</i>			2374	MMCCCLXXIII

Fig. 1. Sumerian, Egyptian, and Roman number systems. The Sumerian language illustrates how the structure of a number system was reflected in their number-words. Egyptian and Roman languages provide examples of number-systems formed according to an additive principle (see Cajory, 1928).

difference from the modern number system: they were lacking the concept of zero. In Babylonian notation, a number such as 3005 could not be expressed unambiguously, as the empty space was used instead of zeros. This limitation was the source of possible confusion and slowed down the development of mathematics.

Despite their many differences, the vast majority of number systems had key features in common: they utilized some form of compressed representation of an open-ended set of numbers by means of either an additive or multiplicative principle. Other similarities include the fact that the number-names employed as categories for the compressed representation often reflected the structure of the number system, such as in case of Sumerian and Roman systems. More importantly, even if the symbolic notation was based on an additive principle, the linguistic structure relied on a multiplicative relation. This is especially surprising to find with the Romans, who had a precisely ordered flexible verbal number sequence, but used a rather crude and cumbersome symbolic notation (Fig. 1, right).

An analysis of the historical development of numerical competence leads to the following conclusion: a natural task that produces common abstract concepts and common linguistic representations may suggest a common representation in the brain. It can be assumed that this representation arises from the more basic representation of numerical quantities that is shared by both animals and humans. The SpaN model proposes how the more primitive spatial numerical representation can be extended into a multi-digit numerical representation through learned interactions with number category names. These number category names are proposed to be learned in the What cortical processing stream as part of the normal course of language learning. The model shows how learned associations between number categories in the What processing stream and the spatial numerical representation in the Where processing stream naturally leads to place-value number representations as an emergent property of this interaction. Thus, the emergence of place-value number systems in many human cultures is proposed by the model to be an example of learned What–Where information fusion.

The article is organized into two parts to describe the data which these Where and What–Where processes can explain. Part I of the article describes animal and human psychophysical data and the Where model mechanisms that can be used to simulate them. Part II of the article describes additional psychophysical data about multi-digit place-value numbers and how What–Where learned associations can explain them.

Part I: spatial mechanisms in the where processing stream

Our development of the spatial organization and neural mechanisms of the SpaN model was guided by

psychophysiological, neuroanatomical, and physiological data. When the parameters of these neural mechanisms are constrained by a small set of experimental data, the model can quantitatively simulate error and temporal characteristics of human and animal psychophysical data, such as Number Size and Numerical Distance effects, as well as number priming data. Section 2 gives an overview of relevant data, describes previous modeling, and puts the proposed model into the context of contemporary research in the field. Section 3 derives the Where mechanisms of the SpaN model in three steps and discusses the model's neuroanatomical and behavioral implications. Section 4 provides simulations of experimental data. Section 5 discusses the implications of the current work within a broader framework of modeling numerical abilities. Model equations for Where processing properties are provided in Appendix A.

2. Experimental data and modeling approaches

Ample evidence is available on the numerical competence of various animal species (for reviews, see Dehaene, 1997; Gallistel & Gelman, 1992). For example, animals as simple as honeybees are able to discriminate the amount and frequency of reward in appetitive conditioning experiments (Buchanan & Bitterman, 1998). Pigeons can abstract the information about the relative number of items in small visual arrays, independent of other parameters of the stimuli (Emmerton, Lohmann, & Niemann, 1997). Modality transfer experiments demonstrated that rats can learn the number of events in a visual or auditory sequence (flashes and beeps) and then respond to a mixed auditory–visual sequence with the same total number of events (Church & Meck, 1984). In these studies, numerosity discrimination ability usually dropped with the increasing number of items, thereby showing the Number Size effect.

Primates are able to perform simple arithmetic: In the experiments by Washburn and Rumbaugh (1991), chimpanzees chose the pair of two piles of chocolate bits that had a bigger total number of bits, even though each individual pile in the chosen pair had fewer pieces than the largest pile in the second pair. More mistakes were made with decreasing difference in the number of bits, thereby showing the Numerical Distance effect.

Animal studies have provided the most valuable data since they were unbiased and bore little influence of higher-order cognitive interference. However, animal data provide little or no chronometric information for numerical tasks. Reaction time data have been obtained mostly through psychophysical experiments with human subjects. Number reading studies (Bryzbaert, 1995) and subitizing (rapid numerosity estimation of the visual array with a small number of items, usually up to three or four) experiments (Mandler & Shebo, 1982) reported an increase in reaction times for bigger numbers and larger arrays, thereby

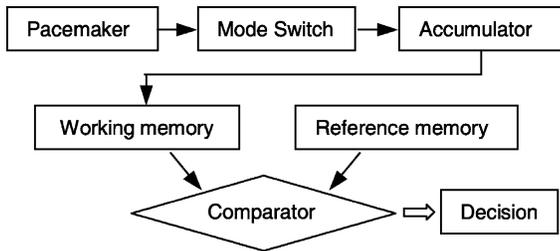


Fig. 2. Functional organization of the Meck and Church model. [Adapted with permission from Meck and Church (1983).]

demonstrating the temporal side of the Number size effect. Many studies addressed single-digit (Parkman, 1971), two-digit (Bryzbaert, 1995; Dehaene, Dupoux, & Mehler, 1990; Link, 1990), and multi-digit (Poltrock & Schwartz, 1984) number comparisons that showed longer reaction times for smaller distances between numbers for most cases, thereby demonstrating the temporal side of Numerical Distance effect. The amount of priming in number priming experiments was also shown to be a decreasing function of the numerical distance from the prime to the target number (den Heyer & Briand, 1986).

Data from brain-lesion patients have provided important insights about the structural composition of numerical abilities, such as the dissociation of verbal and quantitative knowledge, or of subitizing and counting (Dehaene & Cohen, 1994, 1997; Dehaene et al., 1999). Together with the psychophysical data from human infants and adults, and various animal species, they have provided explanatory targets for models of human and animal numerical abilities. Table 1 summarizes some illustrative and widely cited models. The classification is based on the domain of application (function, mechanism and representation addressed) and the experimental data explained. The table also includes the new SpaN model for comparison.

The most influential and widely cited models developed in the last 20 years include models by McCloskey and colleagues (McCloskey, 1992; McCloskey & Macaruso, 1995), Ashcraft (1987, 1992), the encoding complex hypothesis by Campbell and Clark (1992; Clark & Campbell, 1991), and the triple-code model of Dehaene (1992). These are functional models of numerical abilities that generally consider high-level verbal, phonological, graphemic representations and complicated cognitive tasks,

including numerical calculations. A wide variety of human data were explained qualitatively using these models.

The model of Gallistel and Gelman (1992), which was based mostly on animal data, attempts to link principles of animal cognition to human numerical competence. They considered the implications for development of numerical abilities and qualitatively accounted for such phenomena as subitizing, judging the order of two digits, and retrieving number facts. Their model used the mechanism of preverbal counting proposed by Meck and Church (1983) for duration and numerosity estimation. This mechanism was developed to explain how rats transfer numerical estimation knowledge between visual and auditory modalities. Wynn (1998) used a similar mechanism in her model of infant numerical abilities.

The Meck and Church (1983) information-processing model of counting and timing relied on a serial mechanism of numerical information accumulation (Fig. 2). In the model, an endogenously active pacemaker generates equally spaced pulses through time. Before reaching the accumulator, these pulses are gated by a mode switch, which operates either in a *run* or a *stop* mode for duration estimation, or in an *event* mode for counting. The basic idea is that the switch lets through a certain number of the pulses that are endogenously generated by the pacemaker through time. The accumulator adds up the pulses that are let through. During the training trials, the accumulator activation is read out to working memory, and is also stored in a reference memory. During the test trials, the animal is assumed to have a representation of the current accumulator value that can be compared to the information stored as a reference memory. The response (left or right lever press) occurred when the current accumulator value was closer to the value in a reference memory of a reinforced left or right response. To the present, no neural evidence has been found for a pacemaker with these properties.

Dehaene and Changeux (1993) proposed a neural network model that addresses the development of numerical competence in humans and animals (Fig. 3). In this model, visually presented objects in a scene are coded as Gaussian distributions, which then are projected onto a two-dimensional array of Difference-of-Gaussian (DOG) filters that represent the retina. Parallel processing by means of

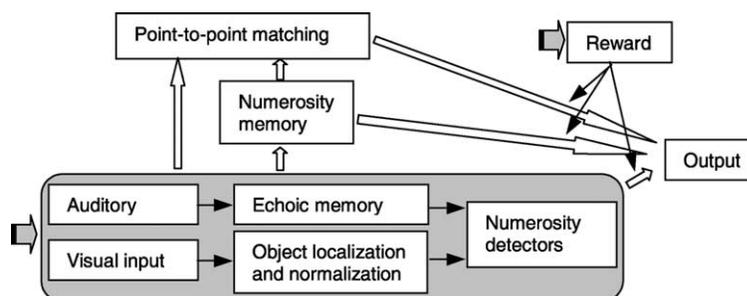


Fig. 3. Functional organization of the Dehaene and Changeux model. [Adapted with permission from Dehaene and Changeux (1993).]

a retina-like mechanism was mainly supported by evidence from subitizing (Mandler & Shebo, 1982). The cumulative activation of this array over all spatial locations activates numerosity detectors that are implemented as nodes with increasing thresholds. Different threshold levels correspond roughly to different numerosities. The model assumes that nodes, which are sensitive to these numerosities, are linearly ordered in a spatial array. The model does not provide a neural explanation of how the linear ordering arises. Auditory input activation is assumed to come from an echoic memory and to add up with the visual stream activation represented by the DOG array cumulative signal. Simulations provided quantitative results that could be compared to response distribution data from animal and human studies.

Additional output clusters were introduced to simulate a number of behavioral tasks related to number discrimination. The connection strengths between numerosity and output clusters were modified through training, where pairs of numerosities were presented as inputs. The connection weights were adjusted according to a Hebbian associative learning rule modulated by reward. After the learning process had been completed, the error rates for number discrimination were predicted as the values of the connection strengths.

3. SpaN model spatial map

Where processing in the SpaN model is defined by a neural network which models the spatial representation and comparison of numerical information in the brain. There are three key features that put the model into its own unique niche in the field. First, it describes the dynamics of the mechanisms of number acquisition and comparison on the scale of milliseconds; thus it explicates the internal structure of numerical processing, or the process microstructure, as opposed to a learned map which only reflects the outcome of numerical processing. Second, the SpaN model simultaneously provides a quantitative fit to error rate and reaction time data. This has not been done before within any single computational model. Finally, the neural mechanisms used in the model are specialized combinations of mechanisms that have previously been used to model processes of motion perception, spatial attention, and target tracking (Chey, Grossberg, & Mingolla, 1998; Gancarz & Grossberg, 1999; Grossberg, 1999a; Grossberg, Mingolla, & Viswanathan, 2001; Grossberg & Rudd, 1989, 1992). The model hereby clarifies how numerical abilities may have arisen as variations of more primitive processes in the *Where* or *How* processing stream of the brain (Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983).

The SpaN model *Where* processing utilizes three main processing stages: Preprocessor, Spatial Number Map, and Comparison Wave. A block diagram (Fig. 4) shows the main processing steps and corresponding model equations.

The following sections describe model implementation details as well as the specific data from brain anatomy, functional brain imaging, and psychophysical experiments that were used to guide model development.

Stage 1: Preprocessor. The preprocessor converts sensory input in different modalities into an analog signal, whose amplitude is roughly proportional to the number of items in a spatial pattern or the number of events in a temporal sequence. For example, given a set of objects in a visual field, the preprocessor assumes that attention shifts serially from one object to another. When each object is attended, a transient signal is generated. The transient detectors enable a continuously changing input to be transformed into a discrete series of output bursts that have a similar size and duration independent of input duration. These similarly calibrated bursts are then added by an accumulator neuron or population of such neurons. Thus, the amplitude of the accumulator activity represents the total number of items in the spatial pattern. Fig. 5 provides an illustrative simulation of how the preprocessor works. This serial mechanism responds to temporal sequences of events in the same way.

This accumulator mechanism plays a functional role similar to the Gaussian spatial filtering used in Dehaene and Changeux (1993), but it avoids a problem that may arise from using Gaussian receptive fields of multiple sizes to estimate numerosity; namely, a receptive field size may lump together several small objects as one, or parts of a large object may be counted as several small objects. The SpaN preprocessor also differs from the gated pacemaker with accumulator proposed in Meck and Church (1983). The SpaN model preprocessor does not require an endogenously active pacemaker, because its transient cells create reactive ‘pulses’ in response to the events themselves. Such transient cell responses are a common feature at the front end of models aimed at explaining data about visual motion perception, spatial attention, and target tracking (Baloch & Grossberg, 1997; Baloch, Grossberg, Mingolla, & Nogueira, 1999; Chey, Grossberg, & Mingolla, 1998; Grossberg, 1999a; Grossberg, Mingolla, & Viswanathan, 2001).

The preprocessor is not the key component of the model. It is realized in a simple way in the SpaN model to portray an early stage of sensory preprocessing and to clarify its possible relationship to other *Where* stream preprocessing mechanisms. Visual, auditory, or tactile sensory streams may produce different analog signals that are physically generated in different brain areas. The key issue is that the amplitudes of these signals are related to numerical properties of the stimuli in each modality. It is assumed in the model that the same numerosities in different modalities produce similar outputs after the integration of transients. Given that the preprocessor creates an analog amplitude to represent numerosity, the next problem is to figure out how this scalar quantity is transformed into a spatial representation of number.

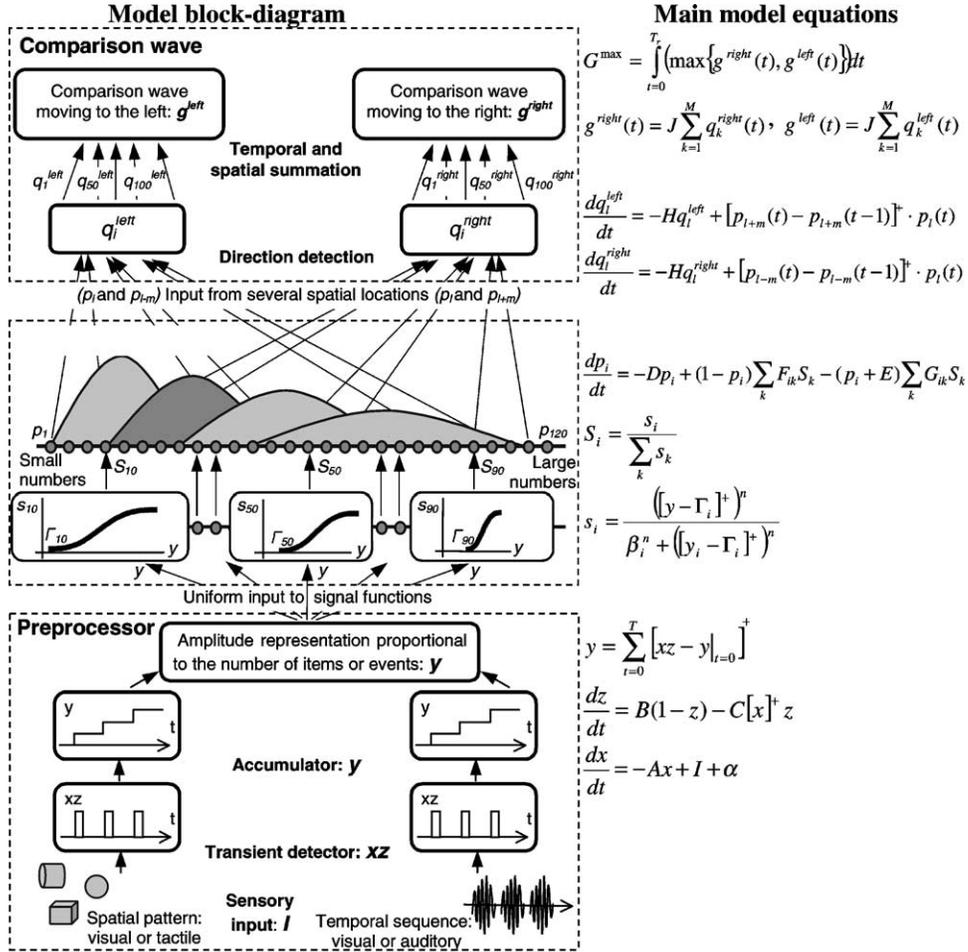


Fig. 4. Functional diagram of the SpaN model. *Preprocessor*: For each sensory input (pattern or sequence), a value of the integrator y is computed at each moment in time and then uniformly fed in to the spatial number map. *Spatial number map*: Each activity p_i of the map receives the normalized output S_i that is derived from the same integrator input y . The signal functions s_i that give the rise to S_i have increasing thresholds and slopes at each successive map cell i . Examples for cells 10, 50, and 100 are shown on the diagram. Each ‘bump’ on the spatial number map schematically represents the activation pattern p_i for sensory inputs with different number of items at the moment when the whole pattern is already processed. Activation of both left and right direction-sensitive cells (q_i^{left} or q_i^{right}) receives the input from two cells of the spatial number map (p_i and p_{i+m}). Both right and left magnitudes (g^{left} or g^{right}) at each moment are computed as summed activations of corresponding direction-sensitive cells (q_i^{left} or q_i^{right}).

Stage 2: Spatial number map. It has been hypothesized for over 30 years (Fairbank, 1969; Restle, 1970) that an analog spatial representation of number exists in the brain; see Dehaene (1997) for an overview. The hypothesis that this representation is modality-independent and exists at a low cognitive level is supported by modality transfer experiments with rats (Church & Meck, 1984) and pigeons (Emmerton, Lohmann, & Niemann, 1997). In these studies, animals were able to extract the numerical properties from the stimuli in visual and auditory modalities, and then to combine the numerical information on an abstract level to produce correct responses. Human adults can, of course, abstract numbers with little effort. Even 6–8-month-old human infants are able to select visual displays that match numerically identical auditory patterns (Starkey, Spelke, & Gelman, 1983).

The concept of an amodal representation being distributed over several spatial locations appeared in the model

by Dehaene and Changeux (1993). Recent brain imaging data (Dehaene et al., 1996; Pesenti, Thioux, Seron, & De Volder, 2000; Pinel et al., 1999; Rickard et al., 2000) have identified the inferior parietal cortex (IPC) as a convergence zone during various numerical tasks with inputs and outputs representing different modalities. IPC and adjacent regions are known to play an important role in various spatial tasks, including visuomotor integration (Nishitani, Uetela, Shibasaki, & Hari, 1999), navigation (Maguire et al., 1998), location working memory (Courtney, Ungerleider, Keil, & Haxby, 1996), and tactile object recognition (Deibert et al., 1999). Spatial attentional deficits are also believed to be associated with IPC damage (Buck, Black, Behrmann, Caldwell, & Bronskill, 1997). All this evidence strengthens the hypothesis that amodal numerosity properties are spatially represented. In parietal cortex, spatial relations may take the form of spatial maps (Andersen, Essick, & Siegel, 1985). The use of transient cells to preprocess inputs

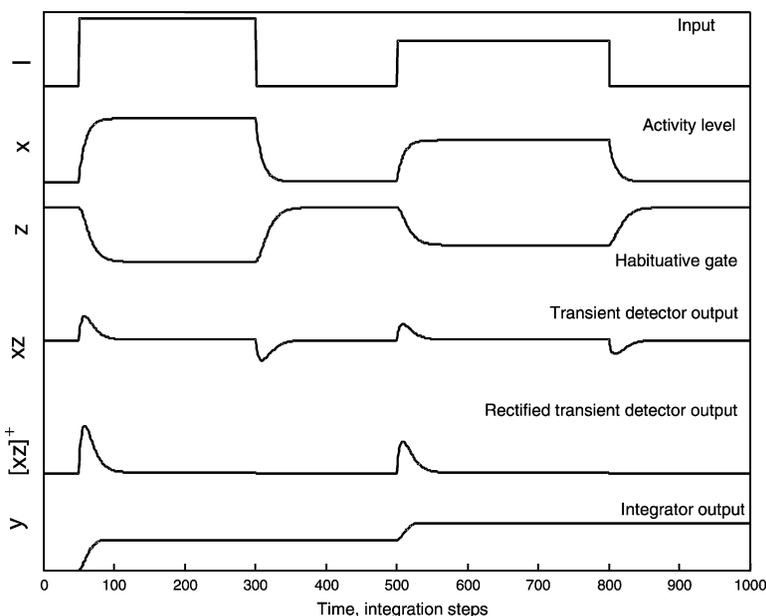


Fig. 5. Computer simulations of the preprocessor operation through time according to Eqs. (A1)–(A3).

to the spatial number map is consistent with the fact that such preprocessing of visual motion, attention, and tracking is part of the Where cortical stream to which IPC belongs.

Numerical organization in humans is known to have both ordering and directional properties. The experiments by Dehaene and colleagues (Dehaene, Bossini, & Giraux, 1990; Dehaene, Dupoux, & Mehler, 1993) demonstrated that left-hand responses were faster than right-hand for the smaller numbers within a given set of numbers, and conversely for the larger numbers. This effect, called SNARC (Spatial-Numeric Association of Response Codes), implies that left-to-right representation exists for numbers going from small to large, at least for subjects raised in Western cultures, while for Middle Eastern subjects (e.g. Iranians) the effect was less pronounced or reversed. A similar directional dependence was reported by Bryzbaert (1995), who found that subjects were faster to respond to the left-to-right ordered pairs (24, 26) than to the same pairs ordered inversely (26, 24).

Several studies have reported subjects actually seeing a vivid image of an ordered structure or a regular geometric shape when performing operations with numbers. In the work of Seron, Pesenti, Noel, and Deloche (1992), approximately 10% of all subjects used some sort of spatial representation, ranging from a one-dimensional line-like to a three-dimensional spiral-shaped structure, when mentally manipulating with numbers up to number values of about 100. The vast majority of these subjects reported ‘seeing’ the numbers used in numerical tasks. These results are consistent with the SpaN model hypothesis, mentioned in Section 1 and developed in Part II, that inter-modal learning, notably learning between language representations of numerical categories and the spatial number map, are used by humans during numerical tasks. Before turning to this

learning process, it is first necessary to model how the more primitive spatial map arises on which it builds.

Taking into account the above evidence about the analog nature of the numerical representation, its directional properties, and data from the IPC studies, it is reasonable to assume that a topographically organized spatial map is used to represent numerical information. In the SpaN model, such a spatial number map is implemented using well-known neural mechanisms; namely, every cell receives its input through a nonlinear signal function before activating excitatory and inhibitory on-center off-surround kernels. How do such commonplace neural mechanisms generate a spatial map? In particular, how can they translate a scalar output signal from the Preprocessor, whose amplitude increases with numerical size, into a spatial map whose position of maximal activation changes as the number increases? How can the Number Size and Numerical Distance effects arise from such basic mechanisms?

The SpaN model predicts that this is accomplished for the spatial number map using the same strategy that has been used to model other spatial brain maps that convert a scalar magnitude into a positional shift across the map. In particular, the model predicts that both the thresholds and the slopes (or sensitivities) of the signal functions increase from the left to the right side of the map (Fig. 4). This Position-Threshold-Slope hypothesis was also used to derive a spatial map from an analog signal by Grossberg and Kuperstein (1986/1989) in their model of saccadic eye movement control, which is another example of a Where stream process. Neurophysiological data wherein thresholds and slopes covary across cells involved in eye movement control have been reported by several investigators (Luschei & Fuchs, 1972; Robinson, 1970; Schiller, 1970). Such a correction between Position, Threshold, and Slope converts

an increasing analog input into a topographical shift across the spatial map in the following way. The scalar input excites all signal functions equally. A small input can strongly activate only cells which receive inputs from signal functions having small thresholds. Such cells are clustered towards the left end of the map. As the input increases, the cells whose signal functions have small thresholds fire more vigorously. However, cells whose signal functions have somewhat larger thresholds also start to fire. In addition, the *rate* of firing by cells with larger thresholds overtakes and exceeds that of the cells with smaller thresholds, because the *slopes* of their signal functions are larger, even though their thresholds are also larger. As the input increases even further, the cells with even larger thresholds start to fire more than all the others, and so on. The location of the maximally activated cells hereby shifts across the map as the input increases.

Because larger inputs can activate cells with both smaller and larger thresholds, the *number*, and the spatial span, of active cells can increase with input amplitude even as the peak response shifts to the right. This tendency is called the Weber law property, which also prevails in numerical data, notably the Number Size effect (Dehaene, 1997). In order to control this distributed activation pattern, competitive interactions between the cells, notably on-center off-surround interactions among cells that obey membrane equations, respond to the inputs received through the array of signal functions by normalizing and spatially sharpening them (Grossberg, 1973, 1980).

These qualitative properties were turned into quantitative data fits by using the response probability curves observed in the classical Mechner (1958) experiments; see Fig. 6.

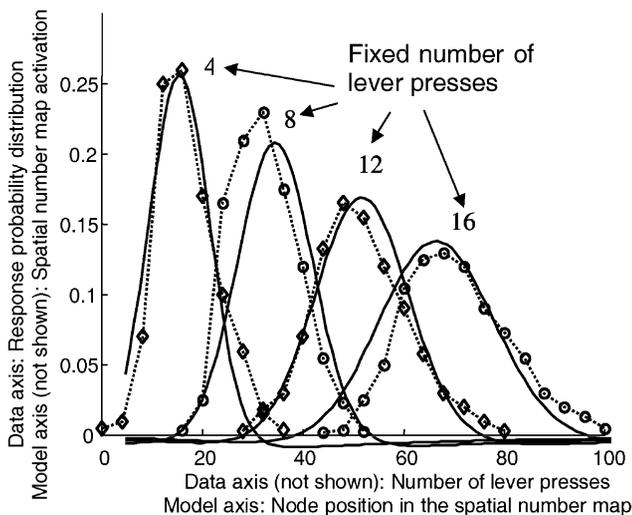


Fig. 6. Data: Rats were trained to press a lever a fixed number of times before switching to another lever in order to get a reward. Data points (diamonds, circles) represent the proportion of responses made to match one of the four required fixed numbers of lever presses (4, 8, 12, or 16), see text for description of the experimental paradigm. [Adapted with permission from Mechner (1958)]. Model: Solid lines show equilibrium activities $p_i(\infty)$ of the spatial number map for four inputs corresponding to 4, 8, 12, and 16 events.

The thresholds and slopes of the spatial map were selected such that the difference (in a least mean square sense) between the experimental and simulated response probabilities was minimized. The probability of response as simulated in the SpaN model is related to the activity distribution in the spatial number map. For a full description of how the model simulates experimental data, see Section 4.

Stage 3: Comparison wave. How are the relative sizes of two numbers compared using the Spatial Number Map? Clearly such a comparison process will generate a wave-like redistribution of activation across the map. The SpaN model shows that such a wave can be directly used to simulate many data about numerical comparison. An exciting aspect of this hypothesis is that a homologous wave has been used to explain many data about motion perception, spatial attention shifts, and target tracking (Baloch & Grossberg, 1997; Francis & Grossberg, 1996; Grossberg, 1999a; Grossberg & Rudd, 1989, 1992). Thus the numerical comparison process may be homologous to other, more primitive, Where stream processes. Link (1992) has used a random walk process to generate comparisons between stimuli. Such a random walk has some properties of a comparison wave, and his hypothesis takes on new meaning using the neural mechanisms that will now be described.

Presentation of two or more numerical inputs causes a redistribution of activation across the spatial number map. For example, suppose that a small number is presented first and produces activation on the left side of the map (Fig. 7). A larger number presented next would build its activation at some location to the right of the smaller one, while some signal remaining from the small number is still decaying. The map activations due to both inputs are added. The sum of this correlated, but spatially displaced, growth and decay of activation is a bell-shaped activation whose maximum moves continuously from the location of the first input to that of the second input. This traveling wave of activation is called a comparison wave because the properties of this dynamical redistribution of activation naturally explain data about numerical comparison.

To selectively detect the direction and amplitude of the comparison wave, another population of cells receives input from the spatial number map. Each cell in this population is sensitive to either the right or the left direction of motion across the spatial number map. Such ‘directional transient

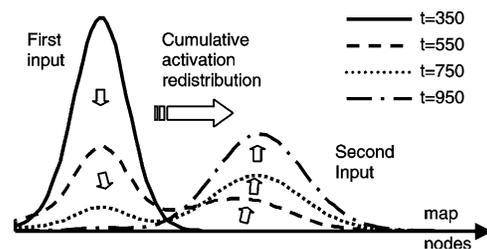


Fig. 7. Computer simulation of the dynamic redistribution of activation across the spatial number map through time (p_i). Different curves correspond to different times, calibrated in integration step units.

cells' are also familiar in neural models of motion perception; e.g. Baloch et al. (1999), Chey, Grossberg, & Mingolla (1998) and Grossberg, Mingolla, & Viswanathan (2001). The activations of all right direction-sensitive cells are added up to yield the right comparison wave output, while outputs from all left direction-sensitive cells add to yield the left comparison wave output. If the comparison wave to the left is larger, then 'smaller' is the judgment; if the right wave wins, then 'larger' is the judgment. Fig. 8A depicts a simulation in which a larger number followed by a smaller number gives rise to a larger wave to the left. In addition, greater distances between the inputs to the spatial number map give rise to bigger comparison waves, thereby producing a more easily discriminated difference. This property of comparison waves helps to explain the Numerical Distance effect. See Fig. 8B for an illustrative simulation. Most brain-imaging techniques, such as PET and fMRI, do not yet possess a sufficient temporal resolution (on the order of 10–100 ms) to support or disprove this hypothesis. Several ERP studies (Ullsperger & Grune, 1995) have, however, detected smaller P300 amplitudes for smaller numerical distances.

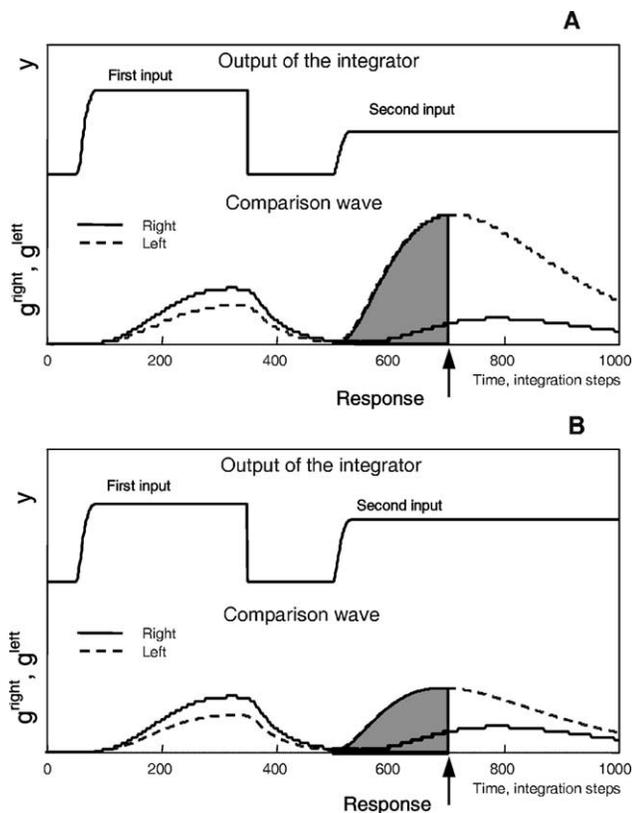


Fig. 8. Computer simulation of the comparison wave. The second input is smaller than the first one, so the wave moving to the left (g^{left}) wins (dashed line). (A) The second input is significantly smaller than the first one; thus, a larger left wave is produced. (B) The second input is a little smaller than the first one; thus, a smaller left wave is produced, which is still larger than the right wave. The model response is computed from the shaded area under the curve (see Appendix A for details).

4. Simulation results

Two types of psychophysical experimental paradigms were simulated with the SpaN model. In the first, the task was to recognize the numerical stimuli. Reaction time was recorded as the time interval from the stimulus onset until the response was initiated. In the SpaN model, the reaction time was simulated as the moment when the maximum activation of the spatial number map ($\max_{i=1\dots 120} (p_i)$ in Eq. (A6)) reached a threshold value. In the second experimental paradigm, two numerical stimuli were compared and reaction time was measured from the onset of the second stimulus or of both stimuli if they were presented simultaneously. The SpaN model assumed serial internal preprocessing even in case of simultaneous stimulus presentation. The probability of response in SpaN was presumed to be proportional to the comparison wave magnitude (g^{left} or g^{right} in Eq. (A10) or (A11)) at the time of the response. The reaction time was determined as the moment when the comparison wave magnitude $\max\{g^{\text{left}}, g^{\text{right}}\}$ reached a threshold value. For both paradigms, this value was fixed for the set of model runs simulating a single experimental setup but could differ for different experiments.

In the SpaN model simulations, it was assumed that preprocessing takes the same amount of time for each numerical stimulus. This assumption is consistent with the use of numerical categories as the inputs in many numerical experiments with adult humans, since the vast majority of number reading and comparison reaction time data come from human subjects presented with Arabic numbers in visual format. This format does not require numerical accumulation mechanism for most experiments when adult subjects are involved. Rather, when the number is recognized, it activates the corresponding number category which, in turn, activates the spatial representation. Modeling of the interaction between preprocessing, learned categorical representations, and the spatial representation is given in Part II of this article.

The simulations were implemented in MATLAB environment and run on a 300 MHz PentiumII PC. An array of 120 cells was used for both the spatial number map and the comparison wave direction-sensitive cells. Throughout the simulations, all free parameters in Eqs. (A1)–(A12) were fixed; see Appendix A for these parameter values. All target data were plotted as dashed lines, and all model results were plotted as solid lines. In evaluating model fits to data, as with all neural models, it is appropriate to evaluate how many *processes*, not how many *parameters*, are used. In the present simulations, an accumulator based on transient cell responses, a spatial map based on correlations between cell threshold and sensitivity, and a comparison wave induced by the times of numerical occurrence provide a conceptually simple description of the underlying brain processes, by using variations of processes that play multiple roles in the Where

cortical stream. In addition, all data fits in the present model are based upon emergent, or interactive, properties of these processes, which makes it difficult to optimize data fits. This is particularly true with the data that are fit using the comparison wave, as will be noted in greater detail below, so getting the degree of quantitative fit that is herein exhibited illustrates how robust the model mechanisms really are.

4.1. Response distribution

The most informative experimental studies provide not only error rate data, but also the complete response distribution. In the work of [Mechner \(1958\)](#), rats were trained to press a lever for a fixed number of times before switching to another lever in order to get a reward. Data points ([Fig. 6](#)) represent the proportion of responses made to match one of the four required fixed numbers of lever presses of 4, 8, 12, or 16 (dotted lines). The animal's task in each block of trials was to perform a fixed number of bar presses (either 4, 8, 12, or 16) on one lever, and then switch to press a second lever. If the number of presses before the switch was correct, the animal was rewarded. If incorrect, then no reward was given. Because animals made counting errors, their performances were distributed among the correct number and either fewer or more responses. Thus, the experiment produced four distributions, which are reprinted in [Fig. 6](#). The abscissa represents the number of pre-switch bar presses and the ordinate is the proportions of trials (in a given block) on which that number was observed. In each block, the animal's modal performance was the target number of responses, and the proportions representing errors fell off smoothly on both sides.

The basis of numerical representation in the model is the spatial number map. Its patterns of activation reflect the differences of processing of different numerical judgments. When the stimulus is present for a sufficient amount of time, and the required response has to be produced without any time constraints, the model assumes the response distribution to be directly related to the equilibrium activation pattern of the map. Said in other way, it is assumed that the subject randomly samples from the spatial number map distribution with a probability that is proportional to the map activity. The solid lines in [Fig. 6](#) show the SpaN map activation for the four stimulus magnitudes related as 4:8:12:16. The patterns for bigger numerosities are more distributed along the map (Weber law), as in the data. There is also a small bias towards larger stimuli, which is reflected in the asymmetry of the distributions: they have more activation to the right of the mean than to the left. The model's ability to *qualitatively* fit these data, notably the Weber law properties, arises directly from the model hypothesis about covariation of signal function thresholds and slopes, followed by intercellular normalizing competition. The *quantitative* fit of model to data was achieved by

selecting particular spatial number map parameters, which were then held fixed to simulate all the other experiments.

4.2. Error rate data

The Numerical Distance effect exhibits itself through experiments wherein two stimuli with different numerical properties are used in each trial. The response error rate is evaluated as a function of the numerical distance between numbers or quantities of objects, which is varied throughout the experiment. These studies demonstrate that the error rate is higher for adjacent numbers and then goes down as numbers are chosen further apart. [Fig. 9A](#) shows the data for humans comparing two-digit numbers and chimpanzees selecting the larger of two piles of chocolate bits.

In the SpaN model, presentation of two inputs in a single trial generates a comparison wave as a result of the transition between the successively activated input representations within the spatial number map. Both left and right waves may appear during the process of redistribution of activation across the map, but the wave with larger magnitude wins and thereby provides the information about the direction of the comparison process. A larger distance between the two numbers results in a greater spatial separation of the corresponding activations of the spatial number map ([Fig. 6](#)), and the redistribution of the activation

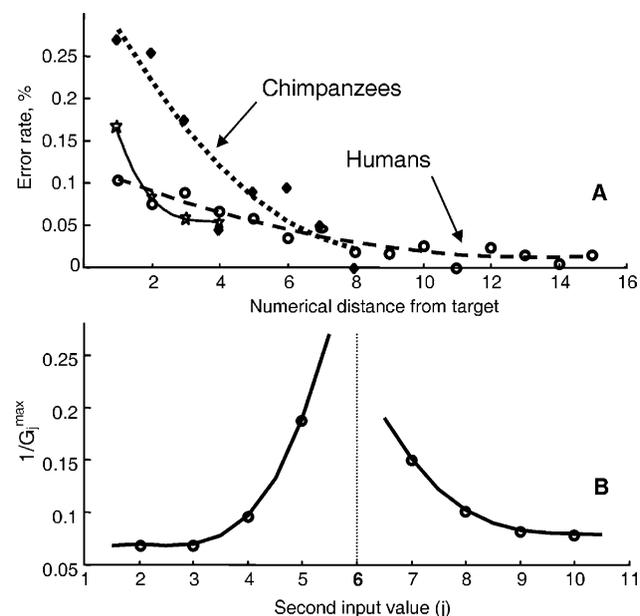


Fig. 9. (A) Data showing error rate for chimpanzees selected the larger pile of chocolate bits (solid diamonds, dotted line is the best cubic fit) [Adapted with permission from [Washburn and Rumbaugh \(1991\)](#).] and people comparing two-digit numbers to a fixed standard of 65 as a function of the numerical distance from the target (open circles, dashed line is the best cubic fit). [Adapted with permission from [Dehaene et al. \(1990\)](#).] Model simulations: Average predicted error rate for humans comparing number-inputs (open stars, solid line is the best cubic fit). (B) Model simulated error rate for a set of number pairs (2, 6) to (10, 6), calculated as the inverse of G_1^{\max} (Eq. (A12)) as a function of distance between the first and the second inputs.

occurs *between* the spatially separated positions along the map as opposed to activation decay and rise at almost the same position in the map. Presence of a substantial along-the-map component of activation redistribution produces a larger amplitude of the comparison wave. The SpaN model hypothesizes that, the larger the amplitude of the comparison wave in one direction with respect to the other direction, the more reliable and accurate is the response. In particular, the model hypothesizes that the error rate covaries with the inverse of the maximum amplitudes of the two waves. Due to the dynamic nature of the comparison waves, it is very difficult to optimize the inverse of the maximum amplitude of the two waves to fit these data. Model parameters were therefore selected by trial-and-error.

The first input (equal to 6 units) was presented to the network for a fixed time (450 time steps). After a brief delay (100 time steps), the second input was presented. Accumulation of signals from directional-sensitive cells started at the same time lasts for a fixed number of time steps T_r for each input (Eq. (A12)). Fig. 9B shows the plot of the inverse of G_j^{\max} for different pairs of inputs, where a numerical input 6 was always presented first followed by one of the inputs 2, 3, 4, 5, 7, 8, 9, or 10. The smallest comparison waves (largest inverse values of G_j^{\max}) occurred for the inputs closest to input 6, with the fall-off slowing down with increasing distance from input 6. The slowing-down effect results from a decreasing correlation of the fall and rise of the activations of the two inputs as the distance between them on the spatial number map increases. That is why the difference between the waves that occur in (6, 9) and (6, 10) comparisons is smaller than that of for (6, 7) and (6, 8) comparisons. The results of these simulations exhibit similar properties to those found in the experimental data (Fig. 9A).

4.3. Chronometric data

Because the SpaN model is a real-time model that incorporates time as an explicit independent variable, it can simulate the microstructure of the processes underlying numerical abilities, including data about reaction times. For each of the simulated experimental paradigms, the reaction time was determined as $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$. The time interval t_{fixed} was assumed equal for all the input stimuli. It includes the times needed for a motor response (e.g. a key press) or verbal response required by the experimenter, or for visual processing of a one-digit number. The value of t_{fixed} was individually selected for each of the various experimental conditions. The output of the model t_{model} was determined as the time when the comparison wave activation G^{\max} (Eq. (A12)) reached a fixed threshold value. In number reading experiments, this threshold was applied to the activation of the spatial number map; for numerical comparison, it was applied to the activation of the comparison wave. Different threshold values were used for each of the two paradigms above (threshold values and

values of t_{fixed} are reported in the figure captions). The scaling factor of 1/2 was used to convert model output from integration time step units into milliseconds.

In the current simulation setup, the reaction time for comparison tasks depends the rate of the comparison wave built-up. This rate can vary as a function of stimulus magnitude, the difference between consecutive stimulus magnitudes, and the interstimulus interval (ISI). The first two factors simulate the target distance and size effects, while the last one brings up a more challenging issue that is considered in Section 5.

In many experimental paradigms, number-reading times can be assessed with the help of an eye-tracking technique. The duration of fixation on the numerical stimuli can be used to estimate the time used for cognitive processing after the visual system processing time is subtracted. The latter is assumed to be equal for all single-digit numbers in Arabic form. Fig. 10(diamonds, dashed line) displays the first-gaze duration (FGD) times in Arabic number reading experiments (Gielen, Bryzbaert, & Dhont, 1991). Here, the Number Size effect yields an approximately linear increase in reading time with number magnitude. Fig. 10 shows simulation results for inputs with magnitudes ranging from 1 to 10 (circles, solid line). The results are comparable to experimental data and exhibit a similar linear increase in reaction time with increasing stimulus magnitude.

This effect emerges from the fact that the spatial number map signal functions have thresholds and slopes that increase from the left (smaller numbers) to the right (larger numbers) side of the map (Fig. 4). In particular, lower threshold and slope values yield faster activation build-up for smaller numbers than larger threshold and slope values for the larger numbers. The linear fit is, however, an emergent property of interactions across the entire spatial

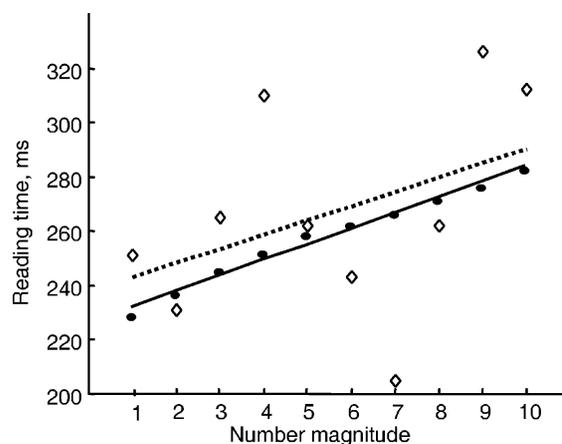


Fig. 10. Number reading times. The data (diamonds, dashed line, a best linear fit) measure first-gaze durations (FGD) during number reading experiments. [Adapted with permission from Gielen, Bryzbaert, and Dhont (1991).] Model simulations: circles, solid line, a best linear fit. Total $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$, where t_{model} is the time when $\max(p_i)$ reached a fixed threshold Th . Simulation parameters: $t_{\text{fixed}} = 205$ ms, $Th = 0.13$.

map, and is not obvious from an inspection of any subset of the chosen parameters.

Experiments wherein the two stimuli are presented with a short delay between them allow estimation of priming effects. For numerical stimuli, two types of priming are known. The first one is semantic priming, where the response time for number–number pairs is contrasted with that of number–letter pairs. The second type of priming is observed for number–number pairs, when the numerical distance between the numbers in each pair is varied. The second type of priming is related to the temporal side of the Numerical Distance effect and has been demonstrated experimentally (Bryzbaert, 1995) to be a linear function of a prime-target (Fig. 11A, diamonds, dotted lines).

Number priming results are obtained from the dynamics of the spatial number map. The first input (prime) is selected from the set of inputs of 1 through 15, and is presented for a fixed time (450 time steps). The second input (target) is presented after a brief delay (100 time steps). Reaction time is measured from the onset of the target input to the moment when the spatial number map activation reaches a fixed threshold. Simulation results for two target inputs 5 and 8

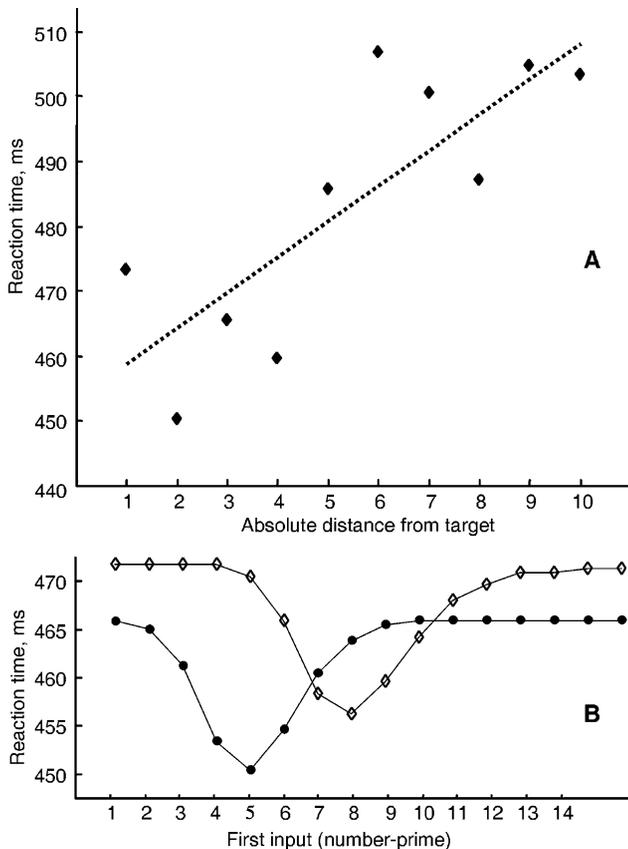


Fig. 11. Number priming. (A) Best linear fit (dashed line) to the data of Bryzbaert (1995). [Adapted with permission from Bryzbaert (1995).] (B) Model simulations for number targets 5 (circles) and 8 (diamonds). Total $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$, where t_{model} is the time when $\max\{p_i\}$ reached a fixed threshold Th. Simulation parameters: $t_{\text{fixed}} = 360$ ms, Th = 0.13.

are shown in Fig. 11B. Priming is a linear function of the distance from target in a limited region around the prime. Then it flattens out for bigger distances.

These trends are due to the following model properties. Target input starts to build up when the activation of the prime is still present. The activation level of the prime and its position in the spatial number map (Fig. 6) are the two factors that control the magnitude of the priming effect. Prime inputs that are closest to the target produce the largest residual activation at the position of the target. A larger distance between the positions of the prime and the target activations along the spatial number map result in a decreasing priming effect. Even if the prime input has a higher absolute level of activation at its position on the map, its residual activation at the position of the target, which gives rise to priming, is smaller than for the number that has a lower absolute level of activation but is closer to the target position. A rationale for why the effect may not flatten out in vivo is proposed in Section 5.

The Number Size and Numerical Distance effects may also be derived from temporal characteristics of the numerical comparison process. Experimental data of Parkman (1971) show that, for a given distance between the pair of numbers, the comparison time increases as the magnitude of the numbers increase. Thus, to compare ‘3’ and ‘5’ takes less time than to compare ‘6’ and ‘8.’ Fig. 12 plots data (filled diamonds, dashed line) which demonstrate the effect for single-digit number comparisons.

To simulate this effect, two inputs were presented to the SpaN model. The first input was presented for a fixed time (450 time steps), the second one followed with a 100 time step delay. Number pairs in the range from (3, 5) to (10, 12) with a constant distance of two units between them served as the inputs. Eight input pairs

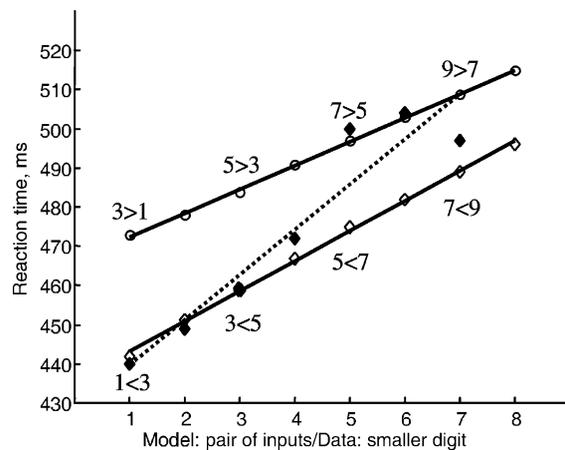


Fig. 12. Number comparison times as a function of number magnitude. Data (solid diamonds, dashed line, best linear fit) from humans comparing one-digit numbers. [Adapted with permission from Parkman (1971).] Model simulations: solid lines (best linear fit); open circles show the right comparison wave results, open diamonds those from the left comparison wave. Total $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$, where t_{model} is the time when G^{max} (Eq. (A12)) reached a fixed threshold Th. Simulation parameters: $t_{\text{fixed}} = 320$ ms, Th = 65.0.

were presented to the network in both ascending (3,5) and descending (5,3) order. Fig. 12 gives the times determined from the comparison wave magnitude build-up rate for both right and left directions (open circles: right, open diamonds: left; solid lines). The lower maximum levels of the spatial number map activation for larger numbers (Fig. 6) yields a weaker effect of activation redistribution. This results in a smaller comparison wave between a pair of larger numbers than between a pair of smaller numbers that are separated by the same distance. The reaction time increase is linear with number magnitude, as in the empirical data.

The influence of the numerical distance on the comparison time was studied in experiments of Link (1990), wherein two-digit numbers in Arabic notation were visually presented to human subjects. The data plotted in Fig. 13(diamonds, dashed lines) show the longest comparison times for nearby numbers, and a rapid decrease for bigger interstimulus distances. In the SpaN model, presentation of the pairs of inputs (2, 6), (3, 6), (4, 6), (5, 6) and (6, 7), (6, 8), (6, 9), (6, 10) to the network were used to simulate comparison time dependence on the interstimulus distances. The same temporal characteristics of the stimuli as in the previous simulation were used. Response times were determined from the larger (left or right) comparison wave magnitude. A larger numerical distance produces a larger separation between the input activations on the spatial number map, resulting in a larger comparison wave magnitude. The larger wave reaches the fixed threshold chosen for this simulation faster than does the smaller wave. Fig. 13(circles, solid lines) demonstrates that the reaction time drop was faster closer to the common stimulus (input equal to six units) and slowed down further away from it in the same fashion as in the experimental data.

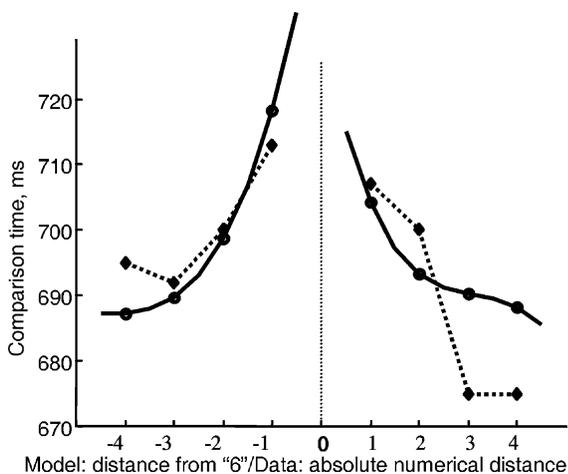


Fig. 13. Number comparison times as a function of numerical distance. Data (diamonds, dashed lines) from people comparing two-digit numbers presented in visual Arabic notation. [Adapted with permission from Link (1990).] Model: circles, solid lines. Total $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$, where t_{model} is the time when G^{max} (Eq. (A12)) reached a fixed threshold of Th . Simulation parameters: $t_{\text{fixed}} = 584$ ms, $Th = 65.0$.

5. Discussion of Where processing properties

The SpaN model proposes a real-time neural architecture for modeling basic facts of numerical competence that are shared by human and animals. The model uses specialized versions of mechanisms—such as transient cells, spatial maps, and comparison waves—that have also been used to model other types of data concerning the Where cortical processing stream. Using these mechanisms, key experimental paradigms about numerical estimation and comparison can be simulated, and quantitative estimates for error rates and reaction times can naturally be obtained as emergent properties of these mechanisms.

Overall, the SpaN model explains a similar set of error rate data as the model of Dehaene and Changeux (1993). On the other hand, the comparison wave mechanism that we propose is a conceptually different approach from the one proposed by Dehaene and Changeux. In their model, the discrimination is based on the value of connection weights that are proposed to be learned during multiple presentations of the same pairs of numerosities over a longer time scale. In particular, Dehaene and Changeux propose that learned weights connect each of the three layers of their network (numerosity detectors, memory clusters, and matching clusters) to the output clusters. These connection weights were modified during training in such a way that “the network had to activate one output cluster if N2 was larger than N1, and another output cluster if N2 was smaller than N1” using a reward to strengthen a connection if the predicted output was correct. This approach raises a number of questions about how the output clusters could themselves be self-organized to represent the comparison being made, and also about whether reinforcement learning is needed to make all comparisons of larger and smaller pairs of numbers. Independent of such concerns, one can imagine that a comparison based on dynamic properties of the Where stream and higher-level associative learning mechanisms are not necessarily mutually exclusive. The comparison wave shows how the microstructure of the spatial number map itself, on a millisecond time scale, may yield these effects when these dynamics are sensed by directionally selective populations of output cells.

In addition to error rates, the SpaN model goes beyond the scope of previous models by accounting for four different types of chronometric data. A common temporal scale that integrates inputs in real time was used in all simulations. This is an important feature, since the same transformation of integration time steps into milliseconds is used in each case, thus providing a common linking hypothesis between the two types of data.

One of the subtleties in the data explained concerns the approximately linear increase in number reading times (Fig. 10) that was similar to gaze duration times for single digit reading (Gielen, Bryzbaert, & Dhont, 1991). Some psychophysical studies have reported logarithmic behavior with respect to number magnitude, such as subitizing

experiments (Mandler & Shebo, 1982) and two-digit number reading experiments (Bryzbaert, 1995). We suggest that other properties of the stimuli than their magnitude could have contributed to the results in some cases. For example, there could be parallel rather than serial visual processing mechanisms operating in the subitizing and the number of digits studies, including effects of verbally mediated categorization, using mechanisms such as those described in Part II of this article.

In the subitizing case, processing of a visual array of items is unlikely to be a serial process, as reflected by a disproportionately small reaction time increase when the number of items in the array changes from one to three. The serial-parallel processing dilemma related to the mechanism of subitizing has been presented in the literature for many years, attributing subitizing to a process of pattern recognition (Mandler & Shebo, 1982) or preverbal counting (Gallistel & Gelman, 1992). The present work does not concentrate on the investigation of subitizing. Rather, it assumes that whatever the mechanism is, it activates a certain representation (the SpaN model preprocessor) that reflects the number of items in a subitizable array. In the two-digit number reading experiments, processing of a two-digit number involves verbally mediated categorization that is possibly superimposed onto the number-magnitude related effect.

Most of the simulations captured the behavior of a psychophysical variable in the whole range of the stimuli employed (Section 4). The exception was the priming data (Fig. 11). Here, only a range of numbers, centered at the stimulus-prime, exhibited the experimentally demonstrated linear increase in reading time, while larger prime-target distances resulted in no or little priming in the model, as opposed to the linear priming that is experimentally observed (Bryzbaert, 1995). The following hypothesis provides a possible explanation of this discrepancy: cognitive processing of two-digit numbers may depend on the other properties of the stimuli, like the number of digits (indicated in Bryzbaert, 1995). This may imply that the two-digit number is not processed all at once. Rather, a dissociation may occur between the processing of each of the two digits. If this hypothesis is correct, then the priming effect cannot be fully attributed to number magnitude, because part of the effect may come from a categorical perception of numbers that is due to interactions between the separate processing of each digit. Given the above assumptions, which are supported by our simulations of multi-digit number data, in Part II of this article, the model results can be interpreted as a genuine magnitude-related priming, which is valid for small numbers, or for a certain range of numbers, and that is not influenced by composite categorical-magnitude processing.

In both number priming and number comparison experiments, the response time may depend on the interval between the presentation of the two consecutive stimuli, namely, the ISI. In the experiments by Gielen, Bryzbaert,

and Dhont (1991), the gaze duration is used to estimate the time interval needed by the subjects to process a numerical stimulus under no time stress. This interval ranged from 200 to 600 ms for one-digit and two-digit numbers. An ISI of less than 200 ms for single-digit numbers is thus likely to introduce some kind of interference in the processing of the first stimulus and lead to disappearance or even reversal of the Numerical Distance effect in the comparison task (Bryzbaert, 1995). The SpaN model uses ISIs of about 250 ms in the simulations. In this parameter range, the redistribution through time of spatial number map activations produces the comparison wave, leading to a genuine distance effect. This effect was also demonstrated by Link (1990) in his numerical comparison experiments with even larger ISIs (1500 ms). Very long ISI (seconds) may produce a complete decay of the activation produced by first stimulus, and when the second stimulus comes in, it may become necessary to somehow recall the first stimulus from memory. On the other hand, a very short ISI, or simultaneous stimuli presentation, may not significantly influence the pattern of results, as demonstrated in Bryzbaert (1995) for ISIs of 0, 200, and 400 ms, which may be interpreted as support for the hypothesis of serial processing of each stimulus.

The brain area within the Where processing stream that seems to be used for numerical representation is the IPC (namely, the areas around the inferior-parietal–occipital junction (Dehaene et al., 1996; Pesenti et al., 2000; Pineda et al., 1999; Rickard et al., 2000)). The SpaN model predicts that this area has a topographic organization that has an ordering related to numerical properties of the stimuli that it processes. This map is predicted to arise through a combination of correlated cell activation thresholds and sensitivities, followed by normalizing on-center off-surround competition. Such correlated signal function thresholds and slopes have also been used to explain other types of data, and have experimental support in several parts of the brain. For example, models of saccadic eye movement (Grossberg & Kuperstein, 1986, 1989; Grossberg et al., 1997), consonant processing during speech perception (Cohen & Grossberg, 1997; Grossberg, 1970), and binocular surface brightness perception (Grossberg & Kelly, 1999) all exploit correlated properties of signal function thresholds and slopes. So do models of adaptively timed learning. Fiala, Grossberg, and Bullock (1996) model adaptively timed classical conditioning of the eyeblink reflex using a calcium gradient across a spatially distributed population of cells in the cerebellar cortex, and summarize biochemical data to support this hypothesis. Here, the calcium gradient causes different cells to respond at different rates to a shared analog input signal. A Weber law also occurs in these adaptive timing data, and can be explained as an emergent property of intercellular interactions across these spatially distributed cells. The adaptive timing example may be viewed as a temporal analog of the spatial mapping principles that are shared in the previous examples. In all,

the basic mechanism that gives rise to the spatial number map may reflect a more general brain design.

The use of a comparison wave is natural wherever a shift of spatial attention is commanded by a sequence of events. Grossberg and Rudd (1989, 1992) described the simplest example of such a wave; namely, when the Gaussian activation across space due to one input is decreasing as the activation across a spatially overlapping Gaussian due to a second input is increasing, then such a wave can occur. The properties of such a wave qualitatively simulate many data about long-range apparent motion perception and clarify how motion perception may be related to shifts of spatial attention and target tracking (Grossberg, 1999a). In these examples, some of the cell properties that create the wave can habituate when the cells are activated. If these processes are entirely homologous, then it is sensible to ask if similar habituation may occur in the numerical comparison wave? If it does, then this property would predict that multiple repetitions of a wave, say, from 2 to 6, might habituate properties of a wave, say, from 3 to 5.

A good quantitative fit for reaction time data was obtained from information that is read out from different parts of the model: from the spatial number map for reading times and priming, and from the comparison wave for numerical comparison tasks. We hypothesize that these two output pathways are used to process different sorts of information about numerical estimation and comparison, much as the form and motion systems of the visual cortex are used to estimate different types of information about individual and sequential visual scenes. This proposal predicts that different output pathways may be used for numerical representation in numerical acquisition and retention tasks versus numerical comparison tasks. Recent functional imaging studies are consistent with this proposal (Pesenti et al., 2000; Rickard et al., 2000) but do not probe the predicted neural mechanisms.

Part II: What number categories and What-to-Where associative mapping

With this background of experimental and modeling results about Where processing contributions to numerical representation and estimation, we can now discuss how What numerical categories, when combined with What-to-Where learned associations to the Spatial Number Map, can lead to multi-digit representations that give rise to a place-value number system as an emergent property.

6. Three types of models for multi-digit number comparison

Psychophysical data related to multi-digit number processing include studies on number reading, comparison,

and simple arithmetic. A major controversy arising from the data concerns the response times in numerical comparison experiments (Brybaert, 1995; Dehaene et al., 1990; Hinrichs, Yurko, & Hu, 1981; Poltrock & Schwartz, 1984). All results agree on the general trend when the numbers are compared to a fixed standard: response time becomes longer as the difference between the presented number and the standard becomes smaller. This trend reflects the temporal side of the Numerical Distance effect (Dehaene, 1997). In addition to these response time differences, the Numerical Distance effect is exhibited in an increasing error rate as the difference between the numbers being compared decreases. The controversial portion of the data is related to how the reaction times behave at a decade boundary (for two-digit numbers). Experiments on two-digit (Dehaene et al., 1990) and multi-digit (Poltrock & Schwartz, 1984) number comparisons reported no fine-grain patterns in the reaction time data beyond the conventional Numerical Distance effect. In contrast, the study by Hinrichs et al. (1981) mentioned a statistically significant increase in the reaction time change for the two boundaries between the decades (49–50 and 59–60) versus the adjacent intervals in the number comparison experiment with stimuli ranging from 11 to 99 and a fixed standard of 55. The experiments by Brybaert (1995) demonstrated a reverse distance effect (reaction time increase for larger numerical distance) for the two-digit numbers across the decades boundaries (Fig. 14). Neither Hinrichs et al. (1981) nor Brybaert (1995) proposed a mechanism to explain these observed paradoxical results for numerical comparison between decades or on decades boundaries.

Traditionally, explanations of psychophysical data about multi-digit number comparison were based on information about the symbolic and linguistic structure of numbers—called the *lexicographic* approach—or the magnitude

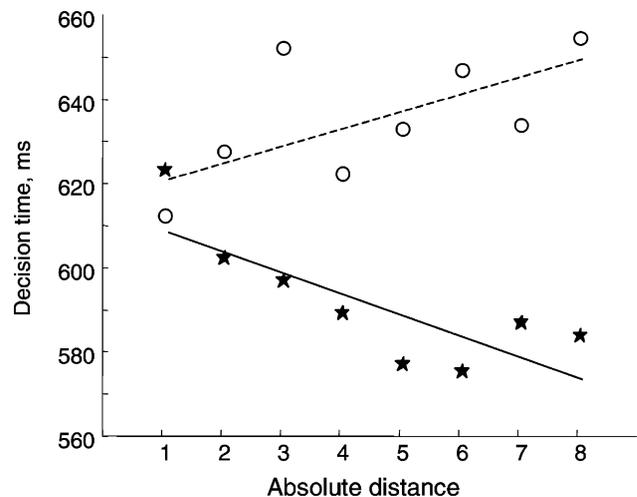


Fig. 14. Decision times for two-digit number comparison as a function of the absolute distance between the target and comparison numbers, whether or not the target and the standard shared the same decade. Open circle: different decade; solid star: same decade. [Adapted with permission from Brybaert (1995).]

representation of numbers—called the *holistic* approach—or a combination of both. The lexicographic approach predicts numerical comparison times based solely on the leftmost digit information (the decades digit in the case of two-digit numbers), completely ignoring the other information (Poltrock & Schwartz, 1984). The holistic approach—in which the symbolic numerical notation first would be converted to a magnitude representation and only then would the two numbers be compared—is supported by the experiments by Dehaene et al. (1990). A combination of the two approaches above was proposed in Hinrichs et al. (1981). According to their hypothesis, for two-digit numbers, the result of the units comparison could influence the result of the decades comparison that, by itself, was providing the correct result. The Hinrichs et al. (1981) hypothesis, called the *interference* model, was questioned by Dehaene et al. (1990), based on the results of their experiments with asynchronous presentation of decades and units digits during the two-digit number comparison task. These experiments yielded no difference in the error rates and reaction times for the conditions when either decades or units digits were presented 50 ms earlier than the other digit. In ruling out the interference model in favor of the holistic model, a strong emphasis was placed on the relative processing speed of the units and decades digits. According to the Dehaene et al. (1990) argument, the earlier presentation of the units digit should have increased the reaction time, while the earlier presentation of the decades digit should have reduced the reaction time, results which have not been observed in the experiments.

As noted above, the present work develops a model of cognitive numerical representation in the human brain that incorporates both lexicographic and holistic components. The lexicographic mechanisms begin to explain the complex structure of the modern numerical system, including the place-value principle that allows a compressed representation of the open-ended set of numbers. This aspect of the model shows how learned number-name categories are involved in numerical representation, and thus clarifies how cognitive processes begin to enter the symbolic number system. The holistic approach provides a basis for the spatial representation of numerical information in the brain and fundamental mechanisms underlying the number comparison processes. Why does not a spatial representation alone have the capacity to represent arbitrarily large finite quantities? One reason is that an extended linear array of spatially represented numbers would run out of space in the brain. A deeper reason is that such a linear array exhibits a Weber law property (Figs. 4 and 6), wherein larger numbers have an ever coarser resolution, thereby leading to increasingly inaccurate operations with them.

In reality, people can deal with arbitrarily large numbers without losing much accuracy. Therefore, additional means for creating an adequate representation for numerical information are required. We suggest herein how the higher-level cognitive process of categorical perception is

closely related to the symbolic structure of any number system. Our model posits learned interactions between number categories (that are themselves learned within the brain's What processing stream for language acquisition) and the spatial representation of numbers that was modeled in Part I (that is part of the brain's Where processing stream for spatial representation and action). The model hereby predicts that symbolic numerical abilities arise through a What–Where interstream interaction. It is through learned interactions between cognitive number categories and the primal spatial number representation that the SpaN model can show why and how the open-ended nature of numerical representation, and thus the edifice of human mathematics, arises. In order to conceptually distinguish the Where SpaN model mechanisms from the model's What-to-Where mechanisms, we call the latter model the Extended SpaN, or ESpaN, model.

The ESpaN model uses the same parameters of the SpaN model, so can simulate all the data described in Part I of the article. ESpaN also provides a quantitative fit to both error rate and reaction time data for multi-digit numbers. It simulates the reaction times and suggests an explanation of the paradoxical reversed Numerical Distance effect observed in Bryzbaert (1995) and partly indicated in Hinrichs et al. (1981). The model also simulates the numerical comparison results for two-digit numbers in the asynchronous digit presentation paradigm and points out some differences between two-digit number comparisons for different language structures, such as English (24 is pronounced as *twenty-four*) versus Dutch (24 is pronounced as *four-and-twenty*). This aspect of the model is restricted to humans, because as far as we know, animals do not have names for numerical categories. Section 7 describes the structure of the ESpaN model, focusing on the interaction between number-name categories and the spatial number map through learning. The model is then used in Section 8 to simulate the reaction time and error rate data in a multi-digit number comparison task as well as the example with asynchronous digit presentation. In Section 9, we discuss the evolutionary implications of the proposed model, its relationship to other multi-modal fusion phenomena in the brain, and its limitations.

7. The ESpaN model

The essence of the ESpaN model is the learned association of verbal categories for number-words and spatial analog numerical representation, which may be considered as a fusion of What and Where information streams (Fig. 15). Recent neurophysiological data have begun to demonstrate the existence of such What-to-Where interactions in the primate brain (Rainer, Asaad, & Miller, 1998). Behavioral and functional imaging data also support this hypothesis (Dehaene et al., 1999). In the present example, the Where stream is represented by the spatial

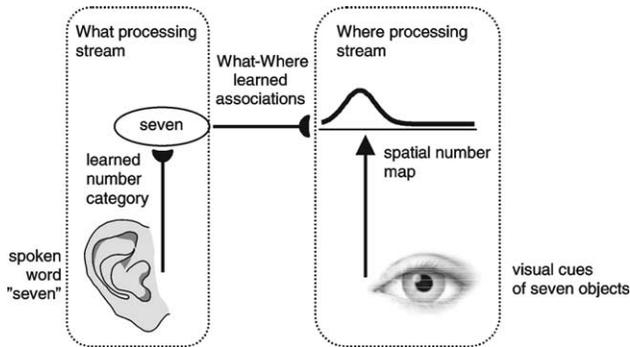


Fig. 15. What–Where information fusion is a key hypothesis of the ESpaN model. Previously learned phonetic categories in the What stream become associated with corresponding locations of the spatial number map in the Where stream. These learned What–Where associations are essential for building a number-system based on the place-value principle.

number map which has a brain correlate in the IPC, as summarized in Part I. Part I also noted that this spatial number map could be activated by sensory inputs coming from visual, auditory or other modalities. Here, we acknowledge that it can also be activated by cognitive categorical inputs originating in other cortical areas, notably areas that process spoken language. The What stream is represented by a set of verbal categories that are learned from speech representations in the temporal and prefrontal cortical areas that have projections from the auditory cortex (Grabowski, Damasio, & Damasio, 1998; Gruber, Kleinschmidt, Binkofski, Steinmetz, & von Cramon, 2000). The verbal categories are connected to the spatial map by adaptive memory weights via associative learning, and are activated by phonetic number-names.

In the SpaN model, the spatial numerical representation was given by a one-dimensional map. This one-dimensional map is extended to a two-dimensional spatial map in the ESpaN model. The original one-dimensional map structure is unchanged. It is extended by spatial gradients of activation that decrease in a direct perpendicular to the original one-dimensional map. Thus the original one-dimensional map is extended to *strips* of cells that represent the same number in the two-dimensional map, albeit with

different levels of activation. Strips of feature-specific cells have been posited to exist in models of other Where stream processes, notably to explain how humans can process multiple auditory streams during auditory scene analysis in order to solve the Cocktail Party Problem (Grossberg, 1999b). Such redundant strip representations may thus reflect a more general Where stream design.

How the two-dimensional number map is used to represent multi-digit numbers may be understood through the following example: The number *two-hundred* requires categories *two* and *hundred* in order to activate the corresponding region in the spatial number map. Before What-to-Where learned associations form, the basic spatial number map has a one-dimensional structure that is represented by extended strips across the map (Fig. 16A). What-to-Where learning converts these strips into localized regions that represent multi-digit numbers (Fig. 16B). Using this learned representation, the ESpaN model can explain numerical abilities for the number system based on both additive and multiplicative principles, as long as there are learned categories for number names. In order to demonstrate and test the ESpaN What-to-Where learning hypothesis, we have simulated the model using examples of English number naming and the decimal place-value number system.

7.1. Number categories and spatial organization

The ESpaN model preserves the primary one-dimensional spatial number map, with activations of the smaller numbers towards the left side of the map, and of larger numbers towards the right side. The ESpaN model extends the one-dimensional map to a two-dimensional map by using a Gaussian gradient along the strips perpendicular to the primary number map (Eq. (B4)). Neural connections from linguistic number categories can be tuned through learning to any locations on this two-dimensional spatial number map. In English, these categories include the single digits from *one* to *nine*, group categories such as *hundred* or *thousand*, and specific phonetic markers such as *ty*, that denotes tens in *twenty* or *thirty*. The schematic

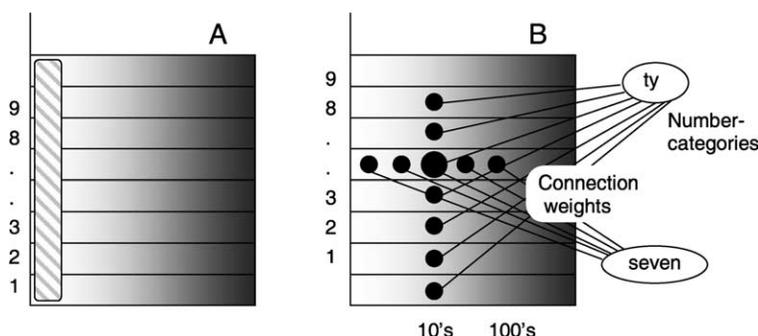


Fig. 16. Schematic representation of the spatial number map and learned What–Where associations. (A) The striped area on the left shows the location of the primary (units) weights strip. (B) An example of where the association for *seven-ty* is formed in the spatial map. The size of the solid circles encodes weight magnitude; the strongest association for *seventy* is arises at the spatial location where both the associations for categories *seven* and *ty* are present.

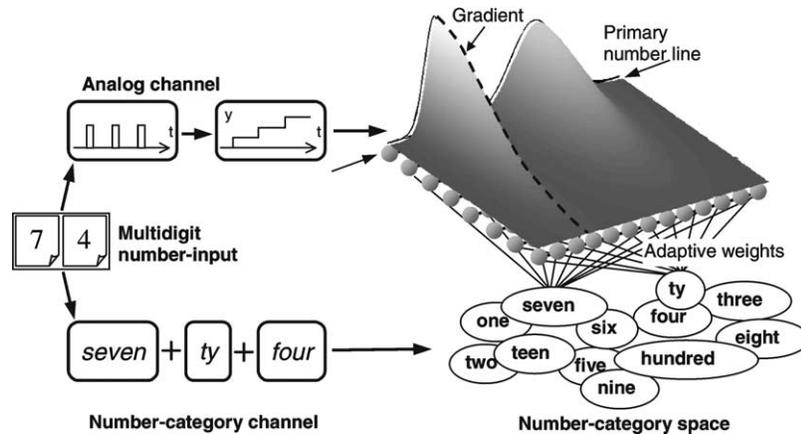


Fig. 17. ESpaN model in the learning mode. During initial development the analog input channel provides the necessary numerical input that drives the formation of the primary strip in the spatial number map. Later in development, with the acquisition of language, number-words reflecting learned number-categories provide the input for the other channel: The activation of the primary strip due to the analog channel propagates down the gradient across the spatial map. When both input channels are active, the weights are learned in the locations of the spatial map activated simultaneously by both the number-category input and the gradient of activation.

operation of the model during the learning phase is shown in Fig. 17. Numerical inputs may activate both the analog and the categorical input channels. A number input like *seventy-four* is, for example, assumed to activate such categories as *seven*, *ty* (for the tens), and *four*. The number categories may be more complex, such as in French or Basque, reflecting the mixture of base-10 and base-20 systems. Phonetic number-categories may also exactly reflect the decimal structure of the number system with the exception of *eleven* and *twelve* in the teens and awkward structures such as *twen-ty* instead of *two-ten* as in modern Chinese.

Human infants who are only a few months old are capable of distinguishing small quantities (Wynn, 1998), suggesting early development of a basic spatial number map. With the acquisition of language, the number words that denote the small single-digit numbers begin to influence the number representation process. During learning within the model, these learned number categories are associated with the area of the spatial number map that has the highest activation level at the time when the number category input is active. For one-digit numbers, this area represents the original, or the primary, number line. Later in map development, when learning more complex two- or three-digit numbers, new categories such as *teen*, *ty*, and *hundred* are learned and become associated with areas of the number map that have smaller activation than the primary number line. Learning of categories *ty* or *hundred* occurs in the presence of a single-digit category; for example, *seven* for *seven-ty* or *seven hundred*. This means that weights for *ty* or *hundred* in the case of *seven-ty* or *seven hundred* will be learned within that portion of the map gradient, where *seven* is active (Fig. 16). If a gradual exposure to more complex numerical structures is assumed, and tens tend to be learned before hundreds, then the categories corresponding to tens will be learned at approximately the same values of

the gradient orthogonal to the primary number line. Thus, the structure of the spatial number map after learning both one- and two-digit numbers will represent a strip of a primary number line corresponding to the learned one-digit numbers and another strip, somewhat parallel to it, corresponding to the tens, or the *ty*, category. Such a strip structure of a spatial number map may not be very regular, but it tends to have a topology where one dimension represents the analog quantitative scale and the other dimension spans a certain number of categories that expand the numerical system with progressively larger number of digits.

Fig. 18 presents an example of the dynamics of the simulated two-dimensional weight pattern connecting number categories with the spatial number map during the learning phase. To roughly portray the first steps of how the child may learn numbers, only one-digit numbers were presented to the network at first. Fig. 18(top row) shows how the weights for number-categories from one to nine evolve with presentation of more and more inputs. The growing strip of large weights connect single-digit categories to the locations of the spatial number map along the primary number line. This strip will be referred as the *units strip* later on. As with a child learning numbers from simple to more complex, the two-digit numbers were then added to the training set of the ESpaN network.

Fig. 18 (rows two and three) shows that once the category weights near the primary number line have saturated, weights corresponding to the new inputs, the tens, are learned within the strip that is parallel to the primary number line and located to the right of the units strip. This *tens strip* represents how the weights from single-digit categories from 1 to 9 are associated with the decades digit of the two-digit numbers.

In our simulation example, we used the English language, which does not possess the most optimal structure of number naming, and bears odd artifacts of

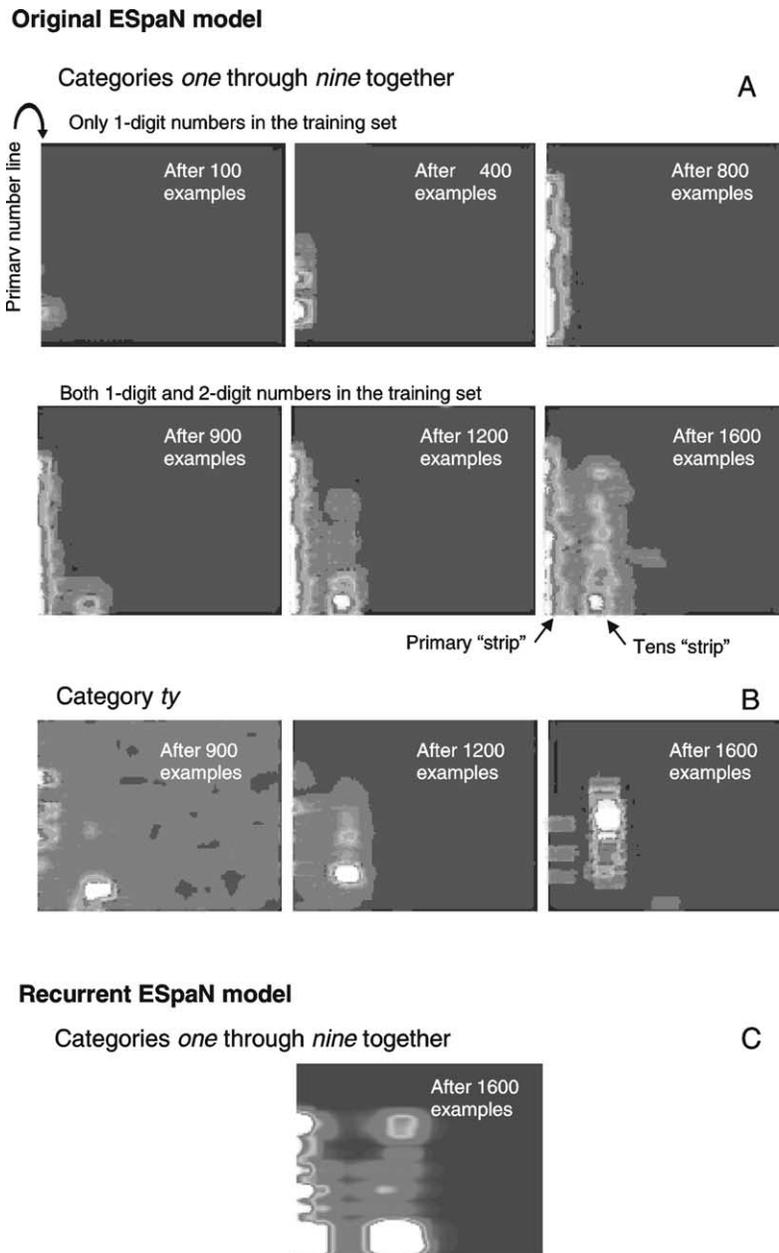


Fig. 18. Computer simulation of learned weight amplitudes for different categories: magnitude is coded by the shades of gray from dark (small) to light (large). (A) Learning progress for weights connecting categories *one* through *nine* to the spatial number map; weight patterns for categories *one* through *nine* are plotted together on each panel. (B) Learning progress for only the weights connecting category *ty* to the spatial map; no learning for category *ty* occurs when only single-digit numbers are present in the training set. C: learned weight pattern for recurrent ESpaN formulation for categories *one* through *nine* plotted together; this weight pattern represents the same stage of network learning as the bottom right panel of part A of this figure.

the past such as *eleven* and *twelve*, and to a lesser extent, the whole structure of *teens*. One may argue that numbers 11 and 12 fall out of the teen linguistic structure. In this case, these two numbers may be assigned individual categories *eleven* and *twelve* that would be learned further to the right from the basic numbers (1–9) on the primary number line. The latter case does not contradict the model hypothesis, as we know that primary number line does not have to end at 9, but may extend as high as 50 for some animal species (Rilling & McDiarmid, 1965).

The formation of the strip structure of the weights to the spatial number map is clarified in Fig. 19 in the example of how the weights are learned from each category. It illustrates the change of the weights connecting the number-category *five* to the spatial map. When only one-digit numbers are presented, the weight patterns with each number line resemble the activation produced by the analog representation of the input corresponding to five items or events as a result of processing by basic mechanisms of the SpaN model (Fig. 19, top three panels).

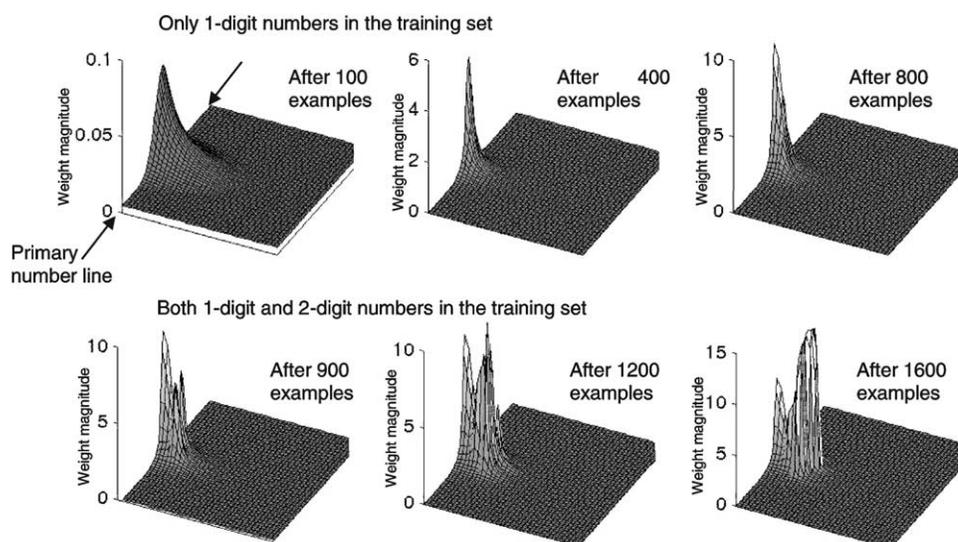


Fig. 19. Computer simulation of the learning process for category *five*. When only 1-digit numbers are present in the training set (top three panels), weights in the primary strip are learned. Presence of 2-digit numbers results in weights growing further from the primary number line, in the *ty* strip (bottom three panels). The irregularities of the weight patterns observed in the figure are due to the initialization of weights to small random numbers and random order of training examples in the training set.

Incorporation of two-digit numbers into the learning process leads to formation of the part of the tens strip that corresponds to the number-category *five* (Fig. 19, bottom three panels). The structure of the weight pattern deviates from the regular bell-shaped activations obtained from in the SpaN model due to competitive interstrip interactions during the learning process and the stochastic nature of weight initialization and ordering of the training set.

In addition to the weights from single-digit categories, the weights connecting *ty* (or the tens) category to the spatial number map are also learned. No learning for the *ty* category occurs when only the single-digit numbers are presented. When the two-digit numbers are included in the training set, a strip of weights spanning the numbers from 2 to 9 along the primary number line dimension is learned in approximately the same location as the strip that corresponds to the category *ty* (Fig. 18B). Note that numbers from 11 to 19 do not contribute to the learning of the tens strip, since in English they are formed by a separate phonetic structure such as *teen*, as in *four-teen*, and represent a separate category. Model equations for the map learning are given at the end of the article. In order to make the map learning simulations manageable, the map was simulated with many fewer cells than would occur in the corresponding map in the brain. Due to this coarse spatial resolution, the learning simulations exhibit larger random fluctuations than would be expected in vivo. With the simulated spatial resolution, each learning trial still took 3–5 h on a PII 300 MHz computer.

7.2. Multi-digit number comparison

The second part of the ESpaN model embodies a mechanism by which the proposed spatial number map

gets incorporated into simple operations with numbers. This mechanism extends the comparison wave proposed in the SpaN model of Part I. The original comparison wave was able to model successfully such properties of human and animal numerical comparison as reaction times and error rates. In the ESpaN model, *multiple* comparison waves exist to represent the redistribution of activation patterns across the two-dimensional spatial number map in a direction *parallel* to the primary number line. In other words, if one looks at the spatial number map as a set of number lines that are parallel to the primary number line, then multiple comparison waves occur within the individual number lines. The ensemble of all these waves determines judgements in the manner described below.

The operation of the ESpaN model in the comparison mode is described by the diagram in Fig. 20. Here, the number comparison process is assumed to take place after a sufficient amount of learning has been accomplished, and a weight pattern from number-categories to the two-dimensional spatial number map (similar to the example in Fig. 18) has been formed. The learning stage of the ESpaN model (and the original SpaN model) assumes that a signal whose amplitude is proportional to the numerical input is generated in the analog input channel. After the weight pattern connecting number categories to the spatial map has already been learned, the input to the analog channel does not have to be present. For example, as children learn more, they become less dependent on the primitive process of counting on fingers. They rely more on numbers that are expressed in their symbolic form with the help of spoken number-names or written number-symbols. Similarly, the ESpaN model assumes that category-based input channel takes a major role in number processing after learning has been completed. It is then sufficient for

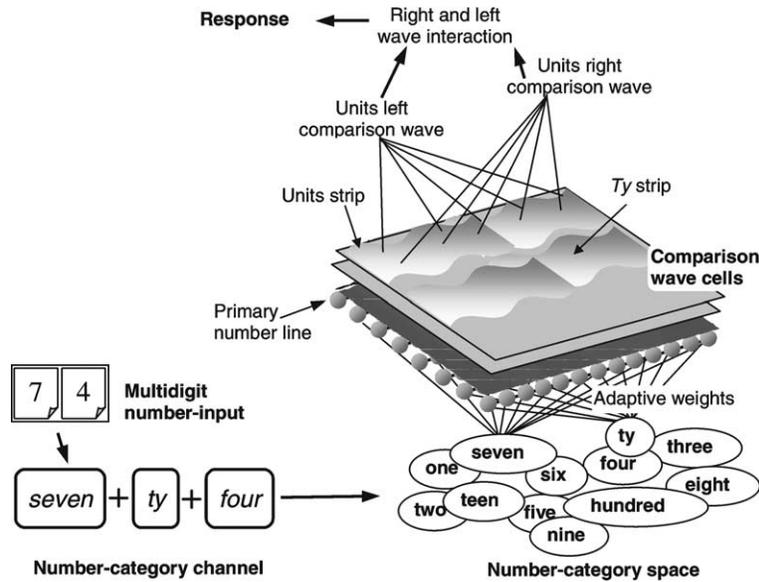


Fig. 20. ESpaN model in the comparison mode. Multi-digit numerical input activates categories in the number-category space. These categories activate specific regions within corresponding strips of the spatial number map via connection weights formed during learning. Dynamic redistribution of activation across the spatial number map is detected with the help of the comparison wave cell layers. The direction of comparison (right or left) is determined from the interaction of comparison waves that occur in different strips.

the number-inputs to directly activate the number-categories that project through the weights onto the spatial number map, producing activation patterns similar to the ones that would have been produced through the analog input channel alone. As in the case of development of mathematical skills, when children start to operate without difficulty with both multiplication tables learned in verbal format and Arabic number-symbols, the model assumes that numerical information in both visual and auditory modalities can activate the category input channel and thus the corresponding number maps and comparison waves.

Simulated comparison waves for number pairs (32, 55) and (38, 55) are shown in Fig. 21. In both examples, the processing

of the decades digit starts before the units digit as reflected in the input temporal structure. Separate comparison waves occur within the tens and units strips of the spatial number map. In case of 32 and 55, waves to the right in the tens strip ($3 < 5$) and in the units strip ($2 < 5$) are larger than waves to the left. When these waves are added up, the cumulative comparison wave moving to the right is significantly larger than the one moving to the left (bottom portion of Fig. 21). This result corresponds to the judgment that 55 is larger than 32. Such an interaction between the comparison waves is assumed to occur during the output processing stage or shortly before the response is produced. In case of 38 and 55, the right wave in the tens strip ($3 < 5$) is again larger than the left one,

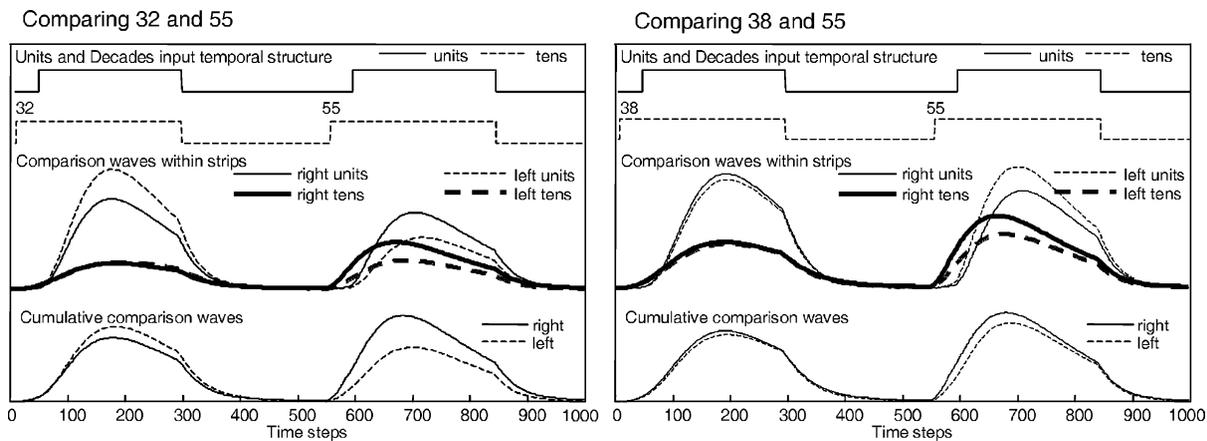


Fig. 21. Simulation of the comparison wave for number pairs $32 < 55$ (left) and $38 < 55$ (right) presented in the numerical comparison task. Two top lines on each panel illustrate the time course of a two-digit number presentation: units digit follows the decades digit after a short delay. Middle graph shows that comparison waves occur in both tens and units strips in both directions. Note that only the comparison wave that occurs after the onset of the second input contributes to the response. The comparison wave before 500 time steps is a by-product of the growth and partial decay of the activation of the first input. The inter-strip interaction results in a cumulative left and right comparison wave (bottom graph). The comparison wave with a larger magnitude wins and determines the response: if the larger wave to the right, then the second number is larger; if the larger wave is to the left, then the second number is smaller.

but the opposite situation occurs in the units strip ($8 > 5$), where the wave moving to the left wins. The cumulative right comparison wave is still larger for this pair of inputs because of the assumption that attention more heavily weights the tens strip; see Eqs. (B10) and (B11), where the weighting coefficients modulate the level of attention. The difference between the left and the right is, however, smaller than for the number-pair 32 and 55, thereby leading to more errors and slower RTs. Model equations for the comparison wave process are given at the end of the article.

8. Data simulations

The ESpaN model allows simulation of the two-digit number comparison experiments and thereby offers an explanation of reaction time data (Bryzbaert, 1995) about the reversed distance effect. A decades-units interaction mechanism based on the cumulative properties of multiple comparison waves is predicted to underlie this paradoxical effect. In the simulated paradigm, a pair of two-digit numerical stimuli were presented to the subject with a stimulus onset asynchrony (SOA) between the first and the second ranging from zero to several seconds. Even for the zero SOA, serial processing of the stimuli assures that processing of the second number starts after the first number has already begun to be processed. Processing of a composite stimulus such as a two-digit number was treated as a mixture of parallel and serial mechanisms in the following fashion: the input signal corresponding to the tens digit started a few milliseconds before the input signal corresponding to the units digit. After this brief delay (denoted UTA, Units-Tens Asynchrony), both tens and units inputs were present simultaneously.

The simulations were implemented in MATLAB environment and run on a 300 MHz PentiumII PC. An array of 120×50 cells (along number lines \times across number lines) was used for both the extended spatial number map and the comparison wave direction-sensitive cells. Throughout the simulations, all parameters in the model equations were fixed. The presence of a stochastic component due to initialization of the spatial map weights to small random numbers at the beginning of the learning process resulted in learned weight patterns that deviate from an ideal bell-shaped profile across the map. Therefore, the simulation results of the both error rates and the reaction times do not always exhibit an entirely regular structure. As noted above, the amount of irregularity in the simulations is greater than that expected in the brain due to the smaller number of cells in the simulations than in the corresponding brain regions. All experimental data were plotted as dashed lines, and all model results were plotted as solid lines.

8.1. Error rates

In the simulations, two-digit numbers from 21 to 89 were compared to a fixed standard of 55. As in the SpaN model,

the error rate was determined by the comparison wave amplitude. The ESpaN results reported here thus include and extend the results simulated by the SpaN model. It is assumed that the larger the relative amplitude of the comparison wave in the left or right direction, the more reliable and accurate the response. The cumulative comparison wave was generated from the waves that occur in the units and the tens strips. The cumulative wave is assumed to be a linear combination of the waves that occur in the tens and units strips. The contribution of the wave in the units strip is thus the same for each decade, e.g. $X_1 < 55, X_2 < 55, \dots, X_8 > 55, X_9 > 55$. Thus, if the error rates are averaged within each decade, the difference in the comparison wave magnitudes across the decades is determined solely by the decades digit. ESpaN simulations are shown in Fig. 22A, along with the experimental data. In these experiments (Hinrichs et al., 1981), the subjects were simultaneously presented with a pair of two-digit numbers (one was always a standard of 55) on a projection screen. Subjects were instructed to respond as quickly as possible by pressing the button associated with either smaller or larger response. Both data and simulations demonstrate the Numerical Distance effect; that is, an increase in the error rate as the stimuli get closer to the standard. In the ESpaN model simulations, the larger distance between the two decades digits resulted in a greater spatial separation of the corresponding activations of the spatial number map *along the number line dimension*. Larger spatial separation of the activations caused a more pronounced redistribution of the activation *between* the spatially separated positions along the number lines within the tens strip, thus producing a larger cumulative comparison wave.

The Numerical Distance effect was the only reliable effect related to the error rates that was reported in the experimental data known to us. The regular intra-decade pattern of error rates (error increase at the end of the decade for numbers smaller than 55, error increase at the beginning of the decade for numbers larger than 55) was generated by the ESpaN model as a result of the decades and units comparison wave interactions. This fine structure has not yet been reported in the experimental data, unlike the data about reaction times that are discussed in the next section. Experiments with more subjects and more trials may be necessary to clarify this issue. To illustrate the argument about the fine structure of the error rates, ESpaN simulation data are shown along with Dehaene et al. (1990) data in Fig. 22B. In the study by Dehaene et al. (1990), the same experimental paradigm as in Hinrichs et al. (1981) was employed with the exception of a cathode-ray tube used instead of a projection screen.

8.2. Reaction times

The reaction times (RT) for a two-digit number comparison task were simulated for the same set of two-digit number pairs from 21 to 89 that were compared to a fixed standard of 55. The RTs were computed according to Eq. (B14). Fig. 23 shows the reaction times simulated with the ESpaN model

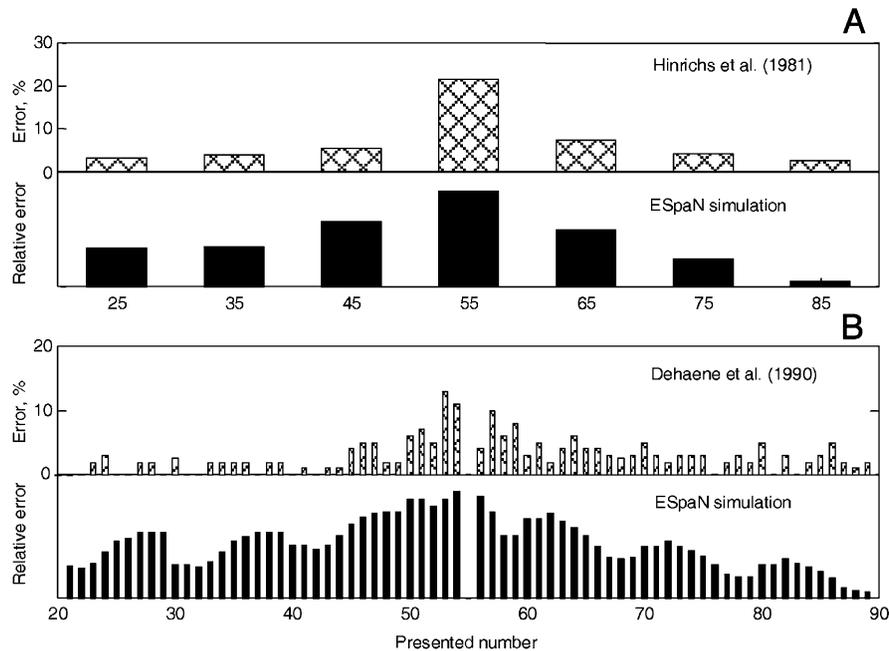


Fig. 22. Simulated and experimental error rates. In the experimental paradigm, a pair of two-digit numbers was presented (one of them 55) and a key-press response to a larger (or smaller) number on the right (or left) was required (see text for more details). Top: error rate data averaged across decades; both experimental and simulated data demonstrate a general decrease of the number of errors with increasing distance from the standard of 55. Bottom: error rate data for all numbers presented; experimental data demonstrate no regular pattern besides the general decrease away from the fixed standard.

(panels A and B) and psychophysical data (Hinrichs et al., 1981, panel C) and (Dehaene et al., 1990, panel D).

The general trend of the RT curve reflects the temporal side of the Numerical Distance effect; namely, the RT

increases for the inputs closer to the standard. In the model, the larger distance between the two numbers results in a greater spatial separation of the corresponding activations in a single number line, and the redistribution of the activation

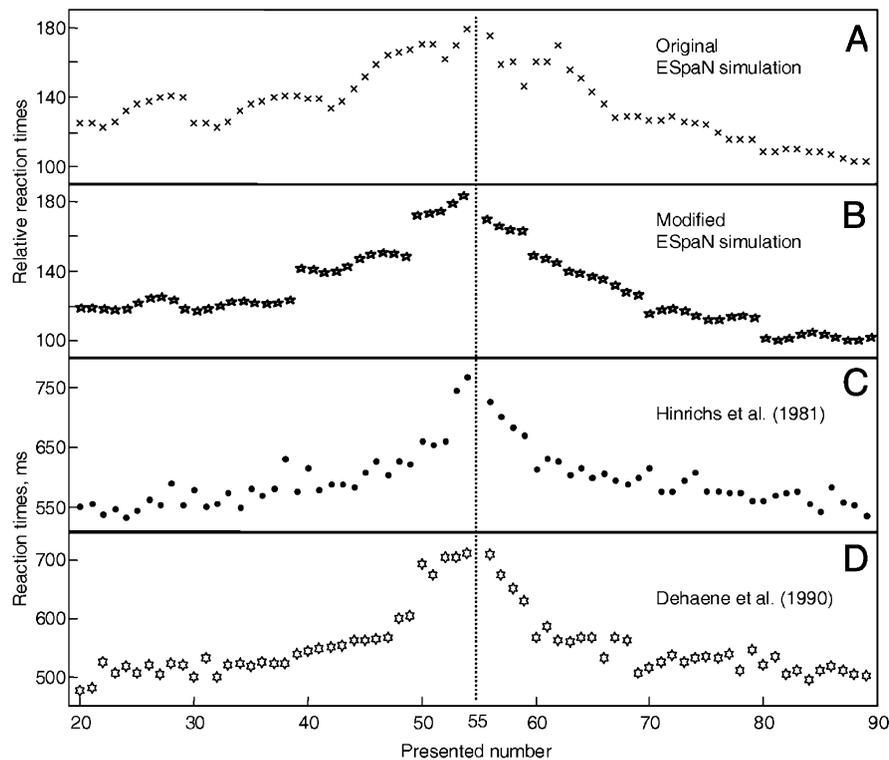


Fig. 23. Reaction times for two-digit numbers compared to 55. (A) ESpaN simulations, learning as described by Eq. (B1); (B) ESpaN simulations, learning as described by Eqs. (B1a) and (B1b); (C) Hinrichs et al. (1981) data; (D) Dehaene et al. (1990) data. [Reprinted with permission from Hinrichs et al. (1981), Dehaene et al. (1990).] Simulation parameter: $T_h = 125$.

occurs *between* the spatially separated positions along the number line as opposed to activation decay and rise at almost the same position in the number line. Presence of a substantial along-number-line component of activation redistribution produces a larger amplitude of the comparison wave.

In order to provide evidence in support of the ESpaN hypothesis about the decades–units interaction within the comparison wave, also referred to as *the interference model* in Dehaene et al. (1990), one must analyse the intra-decade structure of the simulated RT data that extend beyond the conventional distance effect. As already mentioned, most experiments reported no fine structure within the individual decades in the reaction time data in number comparison experiments. Some evidence for the RT discontinuity on the decade boundaries appeared in the study by Hinrichs et al. (1981), who mentioned a statistically significant increase in the reaction time change for the two boundaries between the decades (49–50 and 59–60) with respect to the adjacent intervals. The most reliable experimental results have been obtained by Bryzbaert (1995), who found a pattern of reaction time that increased for smaller numerical distance for two-digit numbers across the decades boundaries, therefore exhibiting a reversed Numerical Distance effect. In these experiments, two-digit numbers were presented side-by-side in a computer screen. The two numbers appeared asynchronously, with an SOA of 0, 200, 400, and 600 ms. Subjects were required to respond by pressing a button on the side of the smaller number. The reversed distance effect was observed for all SOAs, and was the most pronounced for the SOA of 200 ms.

In the ESpaN model simulations (Fig. 23, panel A or B), a regular structure of the intra-decade reaction times is observed. For the inputs smaller than the standard, an additional RT increase occurs towards the end of the decade with the peak at X_8 or X_9 . For the inputs larger than the standard, a similar increase is present towards the beginning of the decade, peaking at X_1 or X_2 . This intra-decade effect is explained by the dynamics of the interaction of the comparison waves between the tens and units strips. For numbers of the same decade, say 40–49, compared to a standard of 55, the largest comparison wave occurs within the tens strip and goes from left to right ($4X < 5X$). The right comparison wave (which is larger than the left) that occurs in the units strip propagates in the same direction as the tens wave for the numbers from 40 to 44 ($4X < 5X$ and $40, \dots, 44 < 45$), adding more to the cumulative right wave. The opposite happens for the numbers from 46 to 49 ($4X < 5X$, but $46, \dots, 49 > 45$), when the units right wave is smaller than the left wave, thus adding less to the cumulative tens and units right wave. The observed intra-decade pattern of reaction times that increase at the end of the decade for numbers smaller than 55, and at the beginning of the decade for numbers larger than 55, is produced due to the contribution of the units digit. In other words, the comparison wave in the units strip moving to

the right is larger (so it reaches a fixed threshold Th faster) at the beginning of the decade than at its end ($40, \dots, 44 < 45$ vs. $46, \dots, 49 > 45$). When the cumulative comparison wave is generated by adding both decades and units waves together, the difference of the units waves for different units digits affects how fast the total wave builds up, which is translated into a regular intra-decade pattern of the reaction times.

8.3. Asynchronous presentation

The experiments with asynchronous presentation of decades and units digits were used by Dehaene et al. (1990) as the main argument in favor of the holistic model of multi-digit number comparison. The ESpaN model assumes that the two-digit number input is fully processed after the categories corresponding to both units and decades digits are activated. Therefore, in order to simulate the asynchronous input presentation, the time-delay between the onset of the decades and units digits was varied. In this paradigm, decades category input always preceded the units, thereby reflecting the structure of number naming in contemporary English. Fig. 24 shows the reaction time patterns generated for decades leading units with large delay (100 time steps) between them (top panel), medium delay (60 time steps) that roughly reflects a synchronous presentation (middle panel), and a zero delay implying units leading decades in the presentation order (bottom panel). A zero delay in the latter simulation reflects the case when the units digit has been already processed, and only awaits the remaining decades part of the two-digit to activate its corresponding category and trigger the comparison process. Both the experimental (Dehaene et al., 1990) and the simulated reaction time data demonstrate similar trends. Based on the fact that the response pattern did not depend on the digit order and the delay between digit presentation, Dehaene et al. (1990) proposed that they had disproved the possibility that any mechanism involving interaction between decades and units digits (interference hypothesis) controls the number comparison process. Based on these ESpaN simulation results, we suggest that the presence such a mechanism (Section 7.2) does not contradict the experimental data, and that the interference hypothesis may thus remain as a plausible explanation for the reaction time patterns observed.

9. Discussion of multi-digit data and simulations

At present, no brain imaging or single-cell recording data are available to either support or disprove the hypothesis that a two-dimensional spatial map with a learned strip organization underlies the representation of multi-digit numbers in the brain. On the other hand, other experimental and theoretical evidence for the existence of mechanisms that link numerical competence to spatial attention and motion detection abilities have been described in various sources, and reviewed in Part I above. The simplest

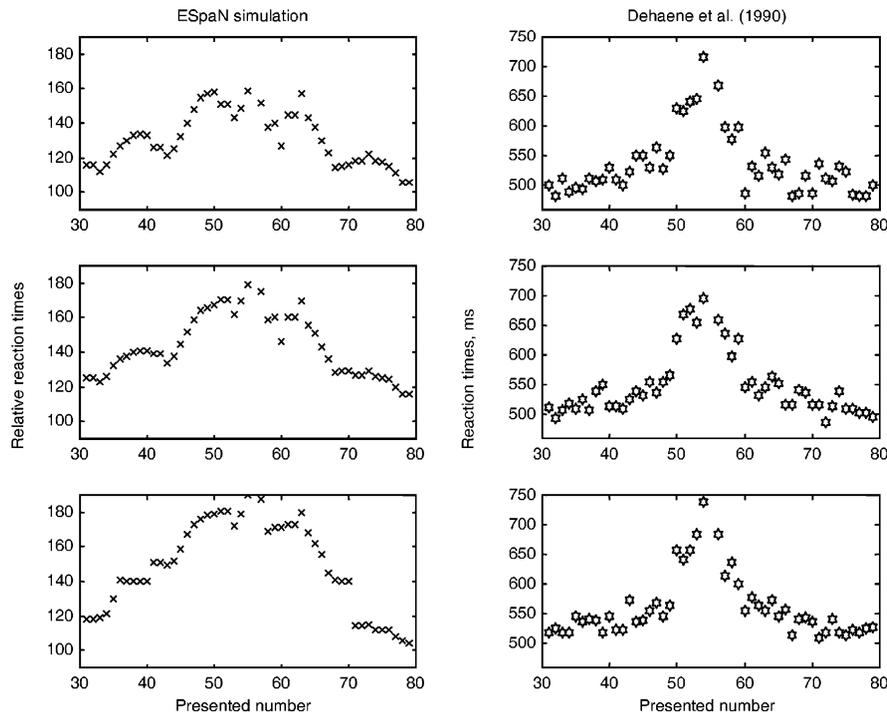


Fig. 24. Reaction times for two-digit numbers compared to 55. Left: ESpaN simulations for zero, medium (60 time steps) and large (100 time steps) delay between the presentation of units digit following the decades digit. Right: experimental data from Dehaene et al. (1990), where two-digit numbers were presented with units leading decades by 50 ms, synchronously, and with decades leading units by 50 ms. [Reprinted with permission from Dehaene et al. (1990).]

version of spatial organization, a one-dimensional one, cannot suffice for large numbers, if only because the accuracy of such a representation would drop dramatically because of the Weber law property. In reality, people are able to perform various mental operations with large numbers without a significant decrease in the performance they demonstrate with smaller numbers. Combining the concept of a spatial map organization with the categorical representation of large numbers in language, leads to the extended spatial number map structure that is modeled herein. In this framework, number category labels supply the additional information that allows learned formation of a compressed and open-ended representation of numbers through interactions with the spatial map. These interactions suggest how units, teens, tys, hundreds, thousands, etc. may be organized in a natural map representation that accommodates the order-of-magnitude increase in numerosity with each successive place value.

In earlier discussions of the role of lexicographic and holistic approaches for the explanation of the number processing (Dehaene et al., 1990), the emphasis was always made on the interaction between the processing of decades and units digits. In the ESpaN approach, this interaction may be controlled by an attentional mechanism that determines how much attention is paid to the decades comparison wave relative to the units comparison wave. In one version of this concept, the interaction between decades and units occurs just before response generation, or during a *post-processing* stage with respect to the comparison stage. Another possibility is

that decades and units interact in the *pre-processing* stage. The post-processing stage gives more attention to one of the comparison wave outputs (see the attentional weighting coefficients R_{units} and R_{tens} in Eqs. (B10) and (B11)). The pre-processing stage gives more attention to the stage where the category input is fed into the spatial number map. Attention could, for example, increase the decades input (I_k) in Eq. (B7) relative to the single digit category input, and the attentional weights in Eq. (B10) and (B11) would be equal ($R_{\text{units}} = R_{\text{tens}}$). Moving the attentional mechanism from post- to pre-processing stage produces almost identical simulation results in Fig. 25. Additional experimental studies are required in order to dissociate the two possibilities and choose which mechanism, pre- or post-processing attentional modulation, is responsible for the behavioral patterns observed in the data.

A possible experimental paradigm that may clarify this issue may exploit the number-naming differences in different languages, such as Dutch versus English. Different pronunciation of two-digit number names such as *four-and-twenty* in Dutch and *twenty-four* in English may lead to dissociation of processing the number-stimuli during the input or output stages. A study performed by Bryzbaert (1995) with Dutch subjects suggested a possible interaction of phonetic and spatial representations in the post-processing stage. In that study, the subjects were asked to name the target number after being presented with both prime and target numbers in a visual Arabic format. The main effects observed were the Number Size (RT increase with the number absolute value increase), the Numerical Distance, and the SNARC

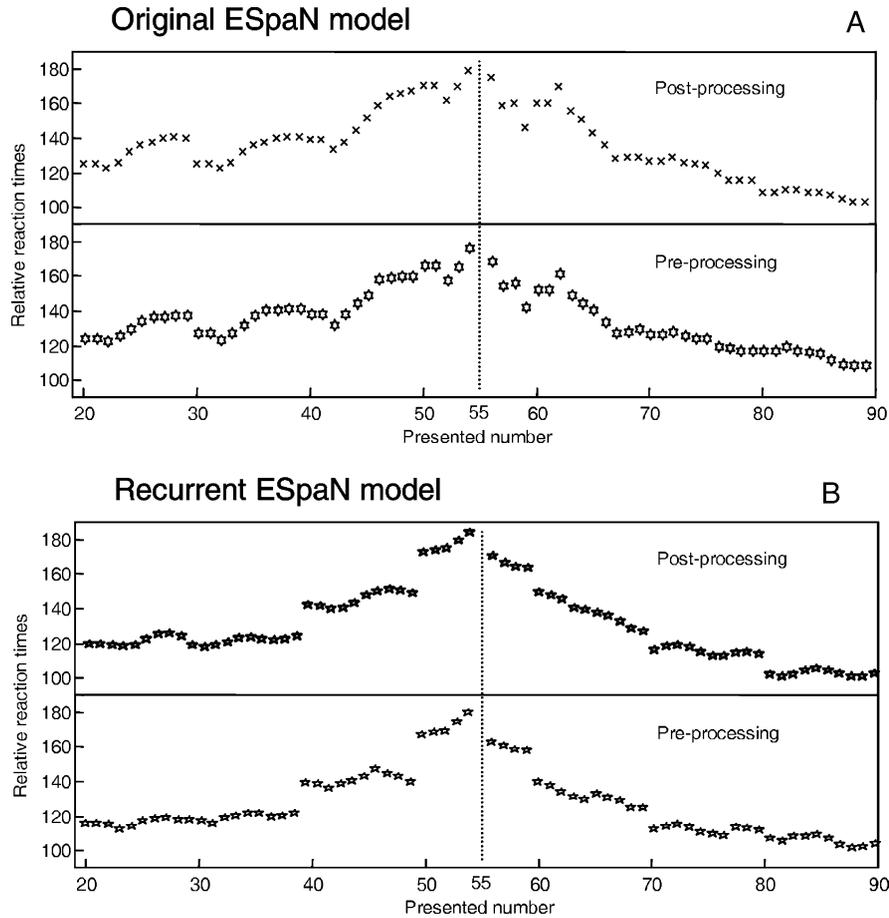


Fig. 25. ESpaN simulation of reaction times for two-digit numbers compared to 55. (A) Original ESpaN model, top: Interaction between category and spatial representation occurs at post-processing stage; bottom: Interaction between category and spatial representation occurs at pre-processing stage. (B) Same as panel A for recurrent ESpaN formulation.

(Spatial-Numeric Association of Response Codes) effects. The SNARC effect (Dehaene et al., 1993) demonstrated that left-hand responses were faster than right-hand for the smaller numbers within a given set of numbers, and conversely for the larger numbers. These effects in the reaction time data were interpreted as evidence for the interaction between a spatial representation and number-names at the response stage. Therefore, a psychophysical experiment similar to the Bryzbaert (1995) paradigm applied to both, say, Dutch and English subjects (to account for the decades-units order in the number naming structure) may help to determine the actual place where the interaction between phonetic and spatial representation occurs in vivo.

Acknowledgements

The authors wish to thank Robin Amos and Diana Meyers for their valuable assistance in the preparation of the manuscript. Stephen Grossberg was supported in part by the National Science Foundation (NSF IRI-97-20333 and the Office of Naval Research CONR N00014-01-1-0624). Dmitry V. Repin was supported in part by the Defense

Advanced Research Projects Agency and the Office of Naval Research (ONR N00014-95-1-0409) and the National Science Foundation (NSF IRI-97-20333).

Appendix A. Span model equations

A.1. Preprocessor

The transient cell response is calculated as the product xz , where the cell activity x increases as a function of input I , and the transmitter gate z habituates, or depresses, as a function of x (Abbott, Varela, Sen, & Nelson, 1997; Baloch et al., 1999; Francis & Grossberg, 1996; Francis, Grossberg, & Mingolla, 1994; Grossberg, 1972, 1980; Markram & Tsodyks, 1996; Ögmen & Gagné, 1990). In particular, the time-averaged activity x is computed by a leaky integrator with a time constant A , where the input I takes the form of a rectangular pulse and α is a constant tonic level:

$$\frac{dx}{dt} = -Ax + I + \alpha. \quad (\text{A1})$$

The habituating transmitter gate z accumulates at rate B to a target level 1, and is inactivated (released, or depressed) by the mass action coupling $-C[x]^+z$ with activity x :

$$\frac{dz}{dt} = B(1 - z) - C[x]^+z. \quad (\text{A2})$$

In Eq. (A2), rates B and C are constant, and the value of x is thresholded, or rectified, at zero: $[x]^+ = \max(x, 0)$. The activity y , which is the final output of the preprocessor, integrates (or sums, in the discrete time formulation) the output signals xz over a threshold value Y

$$y = \sum_{t=0}^t [xz - Y]^+, \quad (\text{A3})$$

where $Y = x(0)z(0)$. The amplitude of this integrated signal is roughly proportional to the number of items or events in a sequence, so that the output reflects numerical properties of the input; see Fig. 5. Parameter t in Eq. (A3) denotes the current time. The initial conditions for Eqs. (A1)–(A3) are $x(0) = \alpha/A$, $z(0) = B/(B + C\alpha/A)$, and $y(0) = x(0)z(0) = (\alpha/A)B/(B + C/A)$. The threshold Y in Eq. (A3) is set equal to $x(0)z(0)$ in order to eliminate the DC component in the integrator final output.

The linearity of the preprocessor does not have to be imposed exactly. In a more general case, a monotonously increasing function of the number of transients corresponding to the number of items or events is sufficient. Having, say, a logarithmic relation between the number of transients and the preprocessor output would result in a different distribution of thresholds and slopes in the spatial number map in order to preserve the properties of the number retention and comparison processes.

A.2. Spatial number map

The spatial number map is implemented as a linear array of nodes, where the input s_i to each node is the output of a sigmoid, or S-shaped, signal function of the integrator activity y :

$$s_i = \frac{([y - \Gamma_i]^+)^n}{\beta_i^n + ([y - \Gamma_i]^+)^n}. \quad (\text{A4})$$

Eq. (A4) begins to convert the nonspecific distribution of the shared input y into a location along a spatial map. In Eq. (A4), i is the node position along the map ($i = 1$ being the leftmost node), y is the integrator output in Eq. (A3), the Γ_i are signal thresholds that increase from left to right, parameters β_i control the slope increase from left to right, and parameter n determines how steep the slopes are; see Fig. 4 for signal function examples. In particular, the thresholds Γ_i increase linearly as a function of i , and the parameters β_i vary as a hyperbolic function of i that takes the form $a + b/(i - c)$; see the end of this section. Normalization of the inputs s_i preserves their relative sizes while also ensuring the same order of magnitude for

values on the left and right sides of the map:

$$S_i = \frac{s_i}{\sum_k s_k}. \quad (\text{A5})$$

In Eq. (A5), the summation over k spans all the map nodes. The normalized S_i input to the activities p_i of the map via an on-center off-surround network, whose cells obey the membrane, or shunting, properties familiar from cell recordings (Grossberg, 1973, 1980, 1988):

$$\frac{dp_i}{dt} = -Dp_i + (1 - p_i) \sum_k F_{ik} S_k - (p_i + E) \sum_k G_{ik} S_k. \quad (\text{A6})$$

In Eq. (A6), D is a constant decay rate, E is a constant hyperpolarization coefficient, and terms F_{ik} and G_{ik} are excitatory and inhibitory Gaussian kernels that define the on-center and off-surround, respectively. The excitatory and inhibitory input sums in Eq. (A6) are gated by the membrane equation, or shunting, gain control terms $(1 - p_i)$ and $(p_i + E)$, respectively, which keep each p_i within the interval $-E$ to 1. Summation over k spans all nodes where the kernels have nonzero values. The initial condition is $p_i(0) = 0$, since no spontaneous activity is assumed to be present. The Gaussian excitatory and inhibitory kernels for different locations on the map obey:

$$F_{ik} = \frac{F}{\gamma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{k-i}{\gamma}\right)^2\right\}$$

and (A7)

$$G_{ik} = \frac{G}{\delta\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{k-i}{\delta}\right)^2\right\}.$$

In Eq. (A7), parameters F , G , γ , and δ are constant. Fig. 6 shows the steady-state activations $p_i(\infty)$ for different values of the integrator output that correspond to increasing numerical magnitude of the input stimulus. In the number reading (Fig. 10) and number priming (Fig. 11) simulations, the reaction times were determined at the time (t_{model}) when $\max(p_i)$ reached a fixed threshold of $\text{Th} = 0.012$. The total reaction times shown in Figs. 10 and 11 were calculated as $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$. For number reading, $t_{\text{fixed}} = 195$ ms. For number priming $t_{\text{fixed}} = 360$ ms.

A.3. Comparison wave

Two arrays of nodes with activities q_i^{right} and q_i^{left} transform the outputs p_i of the spatial number map into left and right comparison waves. The direction-sensitive activities q_i^{right} and q_i^{left} detect the redistribution of activation patterns across the spatial number map through time

$$\frac{dq_i^{\text{right}}}{dt} = -Hq_i^{\text{right}} + [p_{l-m}(t) - p_{l-m}(t-1)]^+ p_l(t), \quad (\text{A8})$$

and

$$\frac{dq_l^{\text{left}}}{dt} = -Hq_l^{\text{left}} + [p_{l+m}(t) - p_{l+m}(t-1)]^+ p_l(t). \quad (\text{A9})$$

In Eqs. (A8) and (A9), index l reflects the cell position in the array, H is the time decay rate, m is a constant shift value, and (t) and $(t-1)$ denote current time and the time one integration step back. This direction-detection mechanism computes the product of activation $p_l(t)$ at current node l and the phasic change of activation $[p_{l\pm m}(t) - p_{l\pm m}(t-1)]^+$ (a derivative-like operation) at the node shifted to the m positions to the left (+) or to the right (-) from the node l . The activities q_i^{right} (q_i^{left}) are added to compute the right and left outputs g^{right} (g^{left}) from the Comparison Wave at any given time t , namely:

$$g^{\text{right}}(t) = J \sum_{k=1}^M q_k^{\text{right}}(t), \quad (\text{A10})$$

and

$$g^{\text{left}}(t) = J \sum_{k=1}^M q_k^{\text{left}}(t). \quad (\text{A11})$$

The comparison wave magnitude G^{max} at the response time T_r is calculated according to Eq. (A12), which integrates the values of g^{right} (g^{left}) up to time T_r , which corresponds to the moment when the larger wave (g^{right} or g^{left}) reaches its maximum amplitude level

$$G^{\text{max}} = \int_{t=0}^{T_r} (\max\{g^{\text{right}}(t), g^{\text{left}}(t)\}) dt. \quad (\text{A12})$$

In all simulations, T_r was fixed at 200 steps (≈ 100 ms) based on the EEG studies of numerical comparison discussed in Dehaene (1997). Parameter J in Eqs. (A10) and (A11) is a constant scaling factor, M is the total number of left or right direction-sensitive nodes, which in turn equals to the number of nodes in the spatial number map.

In the number comparison as a function of magnitude (Fig. 12) and number comparison as a function of distance (Fig. 13) simulations, the reaction times were determined at the time (t_{model}) when the comparison wave amplitude $\max\{g^{\text{left}}, g^{\text{right}}\}$ reached a fixed threshold of $\text{Th} = 6.2$. The total reaction times shown in Figs. 12 and 13 were calculated as $RT = t_{\text{fixed}} + (1/2)t_{\text{model}}$. For the magnitude simulations, $t_{\text{fixed}} = 320$ ms. For the distance simulation $t_{\text{fixed}} = 580$ ms. The following parameter values were fixed for all simulations in both the SpaN and ESpaN models: $A = 10$, $B = 0.05$, $C = 5$, $D = 0.7$, $E = 0.15$, $F = 1$, $G = 3$, $H = 2$, $J = 0.0004$, $M = 120$, $\alpha = 20$, $n = 4$, $m = 10$; for $k = 1, \dots, 120$: $\gamma = 5$, $\delta = 32$, $\Gamma_k = 0.23 + 0.17 \cdot k$, $\beta_k = 1.98 \cdot 10^4 + 2.87/(k - 282)$. Data fits using comparison wave estimates could not be optimized because they are based on nonlinear emergent properties of dynamically changing network comparisons. These fits were found by trial-and-error.

Appendix B

ESpaN model equations

The ESpaN model during the learning phase is described by Eqs. (B1)–(B6). These equations generalize the SpaN model one-dimensional spatial number map to a two-dimensional map that can be activated by number-categories through a learning process. For each training example (a single one- or two-digit number), the learning process is described by the system of Eqs. (B1) and (B6). Eq. (B1) describes the evolution through time of the activation p_{ij} of each cell of the spatial number map. The index i denotes the node position along each number line. The index j designates multiple copies of the number line, with $j = 1$ designating the primary number line; see Figs. 16 and 17.

B.1. Extended number map

$$\begin{aligned} \frac{dp_{ij}}{dt} = & -Dp_{ij} + (1 - p_{ij}) \left[\sum_n F_{in} S_{nj} + \sum_k I_k W_{kij} \right] \\ & - (p_{ij} + E) \left[\sum_n G_{in} S_{nj} + \sum_n P_{in} K_{nj} \right]. \end{aligned} \quad (\text{B1})$$

In Eq. (B1), parameter D is a constant decay rate, term $(1 - p_{ij})$ bounds p_{ij} to remain less than 1 in response to excitatory inputs $\sum_n F_{in} S_{nj} + \sum_k I_k W_{kij}$ from numerical inputs and learned categories, respectively; and term $(p_{ij} + E)$ bounds p_{ij} to remain greater than $-E$ in response to inhibitory inputs $\sum_n G_{in} S_{nj} + \sum_n P_{in} K_{nj}$ from numerical inputs and recurrent inhibitory feedback, respectively. The parameter E determines the maximal hyperpolarization level. Terms F_{ik} and G_{ik} are excitatory and inhibitory kernels that define the on-center and off-surround, respectively, that is activated within each strip j in response to the numerical input S_{nj} . These kernels are thus responsible for the *intra-strip* competition. The term $\sum_n P_{in} K_{nj}$ with kernels K_{nj} , controls the selection of which strip will respond after *inter-strip* competition takes place. Inter-strip competition allows localization of map activation by the individual number-categories and prevents learning from spreading uncontrollably across the strips. All kernels in Eq. (B1) represent Gaussians with constant scaling factors (F , G , and K) and constant variances (γ , δ , and ε). They are defined according to Eq. (B2):

$$\begin{aligned} F_{ik} &= \frac{F}{\gamma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{k-i}{\lambda}\right)^2\right\}, \\ G_{ik} &= \frac{G}{\delta\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{k-i}{\lambda}\right)^2\right\}, \\ K_{nj} &= \frac{K}{\varepsilon\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{n-j}{\varepsilon}\right)^2\right\}. \end{aligned} \quad (\text{B2})$$

The input from the k th number-category multiplies the category output signal I_k with a learned adaptive weight w_{kij} . The learned input from all categories are added via the term $\sum_k I_k w_{kij}$ at map location (i, j) . The analog input S_{nj} generalizes the input to the primary spatial map in Eq. (A5) as follows. An input s_i to location i of the primary number line is generated by the preprocessor of the SpaN model. The equations describing the preprocessor operations are given in Appendix A. After normalization to $s_i / \sum_k s_k$, the normalized input is projected onto a strip of the two-dimensional spatial map in the direction orthogonal to the primary number line via a spatial gradient L_j :

$$S_{ij} = \frac{s_i}{\sum_k s_k} L_j. \quad (\text{B3})$$

In Eq. (B3), the gradient L_j has a unit value at the primary number line ($j = 1$), and decays exponentially to the opposite side of the spatial number map:

$$L_j = \exp\left\{-\frac{j^2}{\zeta}\right\}. \quad (\text{B4})$$

In Eq. (B4), ζ is a constant parameter responsible for the slope of the gradient. The category input I_k takes binary values depending on the activation of a particular category through the category input channel:

$$I_k = \begin{cases} 1, & \text{if the category is present} \\ 0, & \text{if the category is absent} \end{cases}. \quad (\text{B5})$$

Parameters in Eq. (B1) were chosen with the same values as in Eq. (A6).

The weights w_{kij} from number category k to the spatial number map locations (i, j) were learned according to the equation.

B.2. Category-to-map learning

$$\frac{dw_{kij}}{dt} = \eta I_k p_{ij} w_{kij} \left[Q - \sum_l w_{lij} \right]. \quad (\text{B6})$$

Weight w_{kij} changes at a rate proportional to the product $I_k p_{ij} w_{kij}$ of the current category input I_k , cell activity p_{ij} , and

the current weight w_{kij} until the sum of the weights $\sum_l w_{lij}$ associated with the node (i, j) attains the maximum level Q . Product $I_k p_{ij}$ defines an associative learning constraint in that learning occurs only if the category and the map cell are simultaneously active. Parameter η is the fixed learning rate parameter, which equals 0.07 in all the simulations. The Q term defines a competition for a limited weight resource at each map cell. The maximum total weight $Q = 15$. Before the learning process starts, all weights are assigned normally distributed (μ_w, σ_w) small positive values.

The learning paradigm described by Eqs. (B1)–(B6) assumes that a single-digit number *seven* related to the units (*seven*), decades (*seventy*), or even hundreds (*seven hundred*) may be processed through the analog input channel via a serial counting-like mechanism. For example, when children learn basic numbers, counting on fingers is one way that an analog representation of the number of visually presented items may be created. For tens or hundreds, analog input may also be generated through the auditory modality as a result of counting by increments of ten or hundred. The process of silent counting often leads to the activation of auditory categories (*ten*, *hundred*), which may be reflected by lip movement. Those categories are activated serially, and then every instance of encountering the category gives rise to a transient signal, or an activity burst. The accumulation of such bursts over time gives an analog signal whose amplitude is proportional to the number of times any of the categories got activated.

As the process of learning numbers continues, the analog input is required less and less often, as number-words become associated with number categories that develop strong connections to the spatial representation. The latter is especially true for the basic numbers from 1 to 9. In order to model the possibility that number categories become the *major input* for the learning process, even supplanting the role of the analog input channel, we have also studied the following modification of the learning process. In this recurrent model, we assume that inputs from both the analog and the category channels are fed to the cells of the primary number line only. Other strips receive recurrent signals from the primary number line and learned category inputs (Fig. B1).

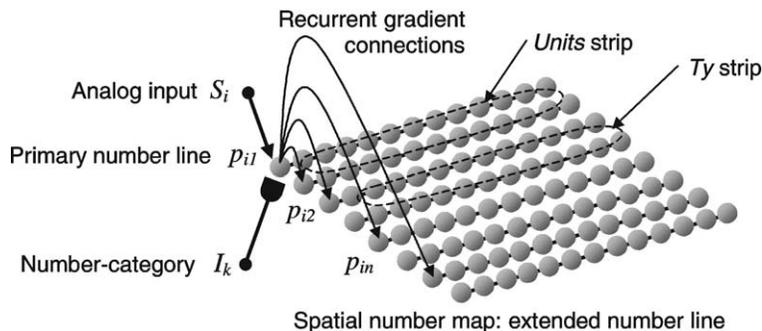


Fig. B1. Recurrent version of the spatial map learning. The input from either of the analog or category input channel activates the primary number line. The activation of the primary number line is then extended onto the whole spatial number map through recurrent gradient connections.

This circuitry enables both an analog input and a learned category input to activate the primary number line *and* the corresponding strips of the extended number line via the recurrent connections, which decrease in strength via a gradient L_j , much as in Eq. (B4). Thus, as the primary number line activation p_{i1} grows, whether due to analog or category input, it extends onto the whole spatial number map through a recurrent gradient of activation L_j ; see term $p_{i1} L_j$ in Eq. (B1b). Eqs. (B1a) and (B1b) below now replace Eq. (B1), since the primary number line (B1a) now requires a separate treatment as the only place which receives an analog input S .

B.3. Recurrent extended number map

$$\begin{aligned} \frac{dp_{i1}}{dt} = & -Dp_{i1} + (1 - p_{i1}) \left[\sum_n F_{in} S_{n1} + \sum_k I_k W_{ki1} \right] \\ & - (p_{i1} + E) \left[\sum_n G_{in} S_{n1} + \sum_n p_{in} K_{n1} \right], \end{aligned} \quad (\text{B1a})$$

$$\begin{aligned} \frac{dp_{ij}}{dt} = & -Dp_{ij} + (1 - p_{ij}) \left[p_{i1} L_j + \sum_k I_k W_{kij} \right] \\ & - (p_{ij} + E) \left[\sum_n p_{in} K_{nj} \right], \quad j \neq 1. \end{aligned} \quad (\text{B1b})$$

In Eq. (B1a), the input S_{n1} is defined by Eq. (B3a):

$$S_{i1} = \frac{s_i}{\sum_k s_k}. \quad (\text{B3a})$$

The two learning paradigms, described by Eq. (B1) and Eqs. (B1a) and (B1b), differ in the amount of learning based on mostly analog estimation of number of items or quantity versus learning based on acquired cognitive categories. We believe that both mechanisms may be mixed in a certain proportion during the initial stages of development of numerical abilities in humans. The reaction time simulations shown in Fig. 23 (panels A and B), demonstrate that both learning circuits lead to similar results.

B.4. ESpaN comparison wave equations

During the number comparison process, inputs come only through the number category channel. The spatial number map Eq. (B1) is simplified by setting inputs S_{nj} coming through the analog channel equal to zero ($S_{nj} = 0$).

B.5. Learned read-out from number-categories

$$\frac{dp_{ij}}{dt} = -Dp_{ij} + (1 - p_{ij}) \sum_k I_k W_{kij} - (p_{ij} + E) \sum_n p_{in} U_{nj}. \quad (\text{B7})$$

In Eq. (B7), parameters D and E are the same as in Eq. (B1), and the category inputs I_k are defined according to

Eq. (B5). Interstrip competition is also present, as in the learning mode.

The comparison wave operates along each number line, denoted by a fixed index j , in the same way as in Eqs. (A8) and (A9), according to Eqs. (B8) and (B9).

B.6. Comparison waves

$$\frac{dq_{lj}^{\text{right}}(t)}{dt} = -Hq_{lj}^{\text{right}}(t) + [p_{l-m,j}(t) - p_{l-m,j}(t-1)]^+ p_{lj}(t), \quad (\text{B8})$$

$$\frac{dq_{lj}^{\text{left}}(t)}{dt} = -Hq_{lj}^{\text{left}}(t) + [p_{l+m,j}(t) - p_{l+m,j}(t-1)]^+ p_{lj}(t). \quad (\text{B9})$$

When $j = 1$, the comparison wave responds to the primary number line, where it detects redistribution of activation along that one-dimensional array of nodes. For each $j > 1$, there is a separate comparison wave. These waves are combined as follows. The activities q_{ij}^{right} (q_{ij}^{left}) are added to compute the right and left outputs g^{right} (g^{left}) from the comparison wave at any given time t . The summation spans both dimensions, along number line positions ($l = 1, \dots, M$) and across number lines for each strip (units: $j = 1, \dots, P_{\text{units}}$; tens: $j = P_{\text{units}}, \dots, P_{\text{tens}}$, etc.):

$$\begin{aligned} g^{\text{right}}(t) = & R_{\text{units}} \sum_{j=1}^{P_{\text{units}}} \sum_{l=1}^M q_{lj}^{\text{right}}(t) + R_{\text{tens}} \sum_{j=P_{\text{units}}}^{P_{\text{tens}}} \sum_{l=1}^M q_{lj}^{\text{right}}(t) \\ & + R_{\text{hundreds}} \sum \dots \end{aligned} \quad (\text{B10})$$

$$\begin{aligned} g^{\text{left}}(t) = & R_{\text{units}} \sum_{j=1}^{P_{\text{units}}} \sum_{l=1}^M q_{lj}^{\text{left}}(t) + R_{\text{tens}} \sum_{j=P_{\text{units}}}^{P_{\text{tens}}} \sum_{l=1}^M q_{lj}^{\text{left}}(t) \\ & + R_{\text{hundreds}} \sum \dots \end{aligned} \quad (\text{B11})$$

In Eqs. (B10) and (B11), R_{units} and R_{tens} are fixed weighting coefficients that may depend on attentional factors when generating a response. Due to the presence of the gradient in Eq. (B4) during the learning stage, the resultant activation of the units strip may become larger than that of the tens strip. The ESpaN model assumes that different levels of attention may exist for different strips in the output, or response generating, stream. According to this hypothesis, more attention is typically paid to the strips that correspond to numerical categories that are acquired later in the learning process. Thus the weighting coefficients obey the following pattern: $R_{\text{units}} < R_{\text{tens}} < R_{\text{hundreds}}$. In particular, $R_{\text{units}} = 1$ and $R_{\text{tens}} = 3$. The number of number lines that comprise each strip (P_{units} and P_{tens}) is not chosen a priori. It arises from the self-organizing structure of the spatial map that is created during learning, where the separation of strips is determined by the kernel K_{nj} in Eq. (B2). Model parameters were chosen such that each strip consists of at least 4–5 number lines in order to

simulate the properties that emerge from the two-dimensional spatial map. In particular, $P_{\text{units}} = 5$, $P_{\text{tens}} = 14$. These simulations do not use three digit numbers, so R_{hundreds} is not defined.

Two types of response were simulated based on the properties of the comparison wave: error rates and reaction times. The error rate was determined as the inverse of G^{max} in Eq. (B13), namely:

$$G^{\text{max}} = \int_{t=0}^{T_r} (\max\{g^{\text{right}}(t), g^{\text{left}}(t)\}) dt, \quad (\text{B12})$$

and

$$\text{Error} = \frac{1}{G^{\text{max}}}, \quad (\text{B13})$$

where g^{right} and g^{left} are defined in Eqs. (B10) and (B11), and the value of T_r (time of response) was fixed at 200 steps (≈ 100 ms) for all pairs of number inputs based on the EEG studies of numerical comparison discussed in Dehaene (1997). The reaction time for each pair of inputs was determined from Eqs. (B10) and (B11) as the moment T when the comparison wave magnitude $\max\{g^{\text{left}}(T), g^{\text{right}}(T)\}$ reached a fixed threshold value Th for all pairs of stimuli presented during the session:

$$RT = \min(T), \quad \text{when} \quad (\text{B14})$$

$$\max\{g^{\text{right}}(T), g^{\text{left}}(T)\} \geq \text{Th}.$$

The fixed value of Th implies that the energy of the wave has to reach a certain level in order to generate the response. We assumed the simplest hypothesis where this level is equal for all stimuli. In addition to parameters defined in the end of Appendix A, the following parameter values were used in the ESpaN model simulations: $Q = 15$, $P_{\text{units}} = 5$, $P_{\text{tens}} = 14$, $R_{\text{units}} = 1$, $R_{\text{tens}} = 3$, $\varepsilon = 5$, $\zeta = 550$, $\eta = 0.07$, $\mu_w = 0.1$, $\sigma_w = 0.05$. Data fits based on multiple comparison wave estimates are even harder to achieve than those based only on the primary number line.

References

- Abbott, L. F., Varela, K., Sen, K., & Nelson, S. D. (1997). Synaptic depression and cortical gain control. *Science*, 275, 220–223.
- Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Enclosing of spatial location by posterior parietal neurons. *Science*, 230, 456–458.
- Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. J. Brainerd, & R. Kali (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research*. New York, NY: Springer.
- Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition*, 44, 75–106.
- Baloch, A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: line motion and for motion dynamics. *Vision Research*, 37, 3037–3059.
- Baloch, A. A., Grossberg, S., Mingolla, E., & Nogueira, C. A. M. (1999). A neural model of first-order and second-order motion perception and magnocellular dynamics. *Journal of the Optical Society of America*, 16, 953–978.
- Brybaert, M. (1995). Arabic number reading: on the nature of the numerical scale and the origin of phonological recording. *Journal of Experimental Psychology: General*, 124(4), 434–452.
- Buchanan, G. M., & Bitterman, M. E. (1998). Learning in honeybees as a function of amount and frequency of reward. *Animal Learning and Behavior*, 16(3), 247–255.
- Buck, B. H., Black, S. E., Behrmann, M., Caldwell, C., & Bronskill, M. J. (1997). Spatial- and object-based attentional deficits in Alzheimer's disease. Relationship to HMPAO-SPECT measures of parietal perfusion. *Brain*, 120(7), 1229–1244.
- Cajory, F. (1928). *A history of mathematical notation*. Chicago, IL: Open Court Publishing Company.
- Campbell, J. I. D., & Clark, J. M. (1992). Cognitive number processing: An encoding-complex perspective. In J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills*. Amsterdam: Elsevier.
- Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769–2786.
- Church, R. M., & Meck, W. H. (1984). The numerical attribute of stimuli. In H. L. Roitblat, T. G. Bever, & H. J. Terrace (Eds.), *Animal cognition*. Hillsdale, NJ: Erlbaum.
- Clark, J. M., & Campbell, J. I. D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition*, 17, 204–239.
- Cohen, M., & Grossberg, S. (1997). Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. *IEEE Transactions on Speech and Audio Processing*, 5, 301–318.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex*, 6(1), 39–49.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1–42.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396.
- Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: a neuronal model. *Journal of Cognitive Neuroscience*, 5(4), 390–407.
- Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 958–975.
- Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, 33(2), 219–250.
- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 626–641.
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science*, 284, 970–974.
- Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., & Mazoyer, B. (1996). Cerebral activations during number multiplication and comparison: a PET study. *Neuropsychologia*, 34(11), 1097–1106.
- Deibert, E., Kraut, M., Kremen, S., & Hart, J., Jr (1999). Neural pathways in tactile object recognition. *Neurology*, 52(7), 1413–1417.
- Emmerton, J., Lohmann, A., & Niemann, J. (1997). Pigeons' serial ordering of numerosity with visual arrays. *Animal Learning and Behavior*, 25(2), 234–244.
- Fairbank, B. A., Jr (1969). Experiments on the temporal aspects of number perception. *Dissertation Abstracts International*, 30(1B), 403.
- Fiala, J. C., Grossberg, S., & Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye blink response. *Journal of Neuroscience*, 16, 3760–3774.

- Francis, G., & Grossberg, S. (1996). Cortical dynamics of form and motion integration: persistence, apparent motion, and illusory contours. *Vision Research*, *36*, 149–173.
- Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: control of visual persistence. *Vision Research*, *34*, 1089–1104.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.
- Gancarz, G., & Grossberg, S. (1999). A neural model of saccadic eye movement control explains task-specific adaptation. *Vision Research*, *39*, 3123–3143.
- Gielen, I., Bryzbaert, M., & Dhont, A. (1991). The syllable-length effect in number processing is task-dependent. *Perception and Psychophysics*, *50*(5), 449–458.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.
- Grabowski, T. J., Damasio, H., & Damasio, A. R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *Neuroimage*, *7*(3), 232–243.
- Grossberg, S. (1970). Neural pattern discrimination. *Journal of Theoretical Biology*, *27*, 291–337.
- Grossberg, S. (1972). A neural theory of punishment and avoidance. II. Quantitative theory. *Mathematical Biosciences*, *15*, 253–285.
- Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, *52*, 217–257.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87*, 1–51.
- Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, *1*, 17–61.
- Grossberg, S. (1999a). How is a moving target continuously tracked behind occluding cover? In T. Watanabe (Ed.), *High level motion processing: Computational, neurobiological, and psychophysical perspectives* (pp. 3–52). Cambridge, MA: MIT Press.
- Grossberg, S. (1999b). Pitch-based streaming in auditory perception. In N. Griffith, & P. Todd (Eds.), *Musical networks: Parallel distributed perception and performance* (pp. 117–140). Cambridge, MA: MIT Press.
- Grossberg, S., & Kelly, F. J. (1999). Neural dynamics of binocular brightness perception. *Vision Research*, *39*, 3796–3816.
- Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control*. Elmsford, NY: Pergamon Press.
- Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*, 2521–2553.
- Grossberg, S., & Rudd, M. E. (1989). A neural architecture for visual motion perception: group and element apparent motion. *Neural Networks*, *2*, 421–450.
- Grossberg, S., & Rudd, M. E. (1992). Cortical dynamics of visual motion perception: short-range and long-range apparent motion. *Psychological Review*, *99*(1), 78–121.
- Gruber, O., Kleinschmidt, A., Binkofski, F., Steinmetz, H., & von Cramon, D. Y. (2000). Cerebral correlates of working memory for temporal information. *Neuroreport*, *11*(8), 1689–1693.
- den Heyer, K., & Briand, K. (1986). Priming single digit numbers: automatic spreading activation dissipates as a function of semantic distance. *American Journal of Psychology*, *99*(3), 315–340.
- Hinrichs, J. V., Yurko, D. S., & Hu, J.-M. (1981). Two-digit number comparison: use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 890–901.
- Link, S. W. (1990). Modeling imageless thought: the relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, *34*(1), 2–41.
- Link, S. W. (1992). *The wave theory of difference and similarity (Scientific Psychology Series)*. Hillsdale, NJ: Erlbaum.
- Luschei, E. S., & Fuchs, A. F. (1972). Activity of brain stem neurons during eye-movement of alert monkeys. *Journal of Neurophysiology*, *35*, 445–461.
- Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science*, *280*, 921–924.
- Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, *111*(1), 1–22.
- Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, *382*, 807–810.
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. *Cognition*, *44*, 107–157.
- McCloskey, M., & Macaruso, P. (1995). Representing and using numerical information. *American Psychologist*, *50*(5), 351–363.
- Mechner, F. (1958). Sequential dependencies of the lengths of consecutive response runs. *Journal of the Experimental Analysis of Behavior*, *1*, 229–233.
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334.
- Menninger, K. (1969). *Number words and number symbols: A cultural history of numbers*. Cambridge, MA: MIT Press.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*(10), 414–417.
- Naccache, L., & Dehaene, S. (2001). The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex*, *11*, 966–974.
- Nishitani, N., Uutela, K., Shibasaki, H., & Hari, R. (1999). Cortical visuomotor integration during eye pursuit and eye-finger pursuit. *Journal of Neuroscience*, *19*(7), 2647–2657.
- Ögmen, H., & Gagné, S. (1990). Neural network architecture for motion perception and elementary motion detection in the fly visual system. *Neural Networks*, *3*, 487–506.
- Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, *91*(2), 191–205.
- Pesenti, M., Thioux, M., Seron, X., & De Volder, A. (2000). Neuroanatomical substrates of arabic number processing, numerical comparison, and simple addition: a PET study. *Journal of Cognitive Neuroscience*, *12*, 461–479.
- Poltrock, S. E., & Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 32–45.
- Pinel, P., Le Clec, G., van de Moortele, P.-F., Naccache, L., Le Bihan, D., & Dehaene, S. (1999). Event-related fMRI analysis of the cerebral circuit for number comparison. *NeuroReport*, *10*, 1473–1479.
- Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, *393*, 577–579.
- Ramachandran, V. S., & Hubbard, E. M. (2001a). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society of London B*, *268*, 979–983.
- Ramachandran, V. S., & Hubbard, E. M. (2001b). Synaesthesia—a window into perception, thought and language. *Journal of Consciousness Studies*, *8*, 3–34.
- Repin, D. V., & Grossberg, S. (1999). How does the brain represent numbers. *Paper presented at the third international conference on cognitive and neural systems*. Boston: Boston University.
- Repin, D. V., & Grossberg, S. (1999). A neural model of how the brain represents numbers. *Paper presented at the 21st international summer school of brain research*. Amsterdam.
- Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*, 274–278.
- Rickard, T. C., Romero, S. G., Basso, G., Wharton, C., Flitman, S., & Grafman, J. (2000). The calculating brain: an fMRI study. *Neuropsychologia*, *38*, 325–335.
- Rilling, M., & McDiarmid, C. (1965). Signal detection in fixed-ratio schedules. *Science*, *148*, 526–527.

- Robinson, D. A. (1970). Oculomotor unit behavior in the monkey. *Journal of Neurophysiology*, 35, 393–404.
- Schiller, P. H. (1970). The discharge characteristics of single units in the oculomotor and abducens nuclei of the unanesthetized monkey. *Experimental Brain Research*, 10, 347–362.
- Seron, X., Pesenti, M., Noel, M.-P., & Deloche, G. (1992). Images of numbers, or ‘When 98 is upper left and 6 is sky blue’. *Cognition*, 44, 159–196.
- Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, 222, 179–181.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- Ullsperger, P., & Grune, K. (1995). Processing of multi-dimensional stimuli: P300 component of the event-related brain potential during mental comparison of compound digits. *Biological Psychology*, 40, 17–31.
- Washburn, D. A., & Rumbaugh, D. M. (1991). Ordinal judgments of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science*, 2(3), 190–193.
- Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences*, 2(8), 296–303.