BB paper.

Proceedings of the Twenty First Annual Conference of the Cognitive Science Society

Martin Hahn and Scott C. Stoness
Editors

August 19-21, 1999
Simon Fraser University
Vancouver, British Columbia

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1999 Mahwah, New Jersey London

A Computational Model of Number Comparison

Marco Zorzi (mzorzi@psico.unipd.it)
Dipartimento di Psicologia Generale, Universita` di Padova
via Venezia 8, 35131 Padova, Italy

Brian Butterworth (b.butterworth@ucl.ac.uk)
Institute of Cognitive Neuroscience, University College London
17, Queen Square, London WC1N 3AR, UK

Abstract

Number comparison is a task that has been widely used to investigate the mental representation of number magnitudes. It is frequently assumed that the mapping from numerals to a "mental number line" is compressive (i.e., logarithmic) or that magnitude representations have the property of scalar variability. In this study, we simulate the process of selecting the larger of two numbers in a neural network model. We show that it is possible to account for the main experimental effects (e.g., the distance effect and the number size effect) with a simple architecture using a linear representation of numerical magnitudes. The compressive effects that are found in the reaction times emerge from the non-linear interactions that are intrinsic to the decision process.

Introduction

Number comparison is one of the fundamental numerical abilities. McCloskey (1992) takes the ability to select the larger of two numbers to be the criterion of understanding of numbers. Neurological patients who perform abnormally on this task, turn out to be profoundly acalculic (Delazer & Butterworth, 1997). Recent research on infant numerosity discrimination is consistent with the idea that infants can recognise which of two visual arrays contains the more objects (e.g., Wynn, 1992). It is also known that some primates, when presented with two visual arrays, can reliably select the array with more objects, for small numbers (Brannon & Terrace, 1998; Washburn & Rumbaugh, 1991).

As with many stimulus dimensions, number size shows a symbolic *distance effect.* That is, it is easier and quicker to select the larger of two numbers when they are numerically dissimilar than when they are similar (Moyer & Landauer, 1967). This "distance effect" has been found in young children (Sekuler & Mierkiewicz, 1977) and even in primates (Washburn & Rumbaugh, 1991; Brannon & Terrace, 1998). Thus, the ability to compare numerosities could be ontogenetically and phylogenetically basic.

According to McCloskey, Caramazza, and Basili's (1985) model, comparing the magnitude of two numerals requires the generation of an abstract representation corresponding to each numeral. However, the way in which the magnitudes of these abstract representations are compared is an unspecified process. In Dehaene's (1992) triple-code model, number comparison is performed on the basis of an *analogue magnitude* code (Dehaene & Cohen, 1995). Again, the details of the comparison process operating on these codes is largely unspecified. In effect, number comparison has not been simulated to our knowledge within a computational framework (but see Dehaene & Changeaux, 1993, for a simulation of the preverbal elementary ability to compare small sets of up to 3-4 objects). The serious limit of any verbal model of number comparison is that the comparison process itself and nature of the input-output representations on which this processes operates are not made explicit; the corollary to this fact is that the origin of the classic effects found in magnitude comparison tasks, such as the symbolic-distance and the number size effects, are still poorly understood.

The symbolic-distance effect refers to the finding that the latency of the comparative judgement is an inverse function of the numerical distance between the two numerals. That is to say, it is easier (i.e., faster) to compare 3 and 5 than 3 and 4. This classic result, indexed by the "split" or difference between the two numbers, has been found with arabic numerals (Banks, Fujii, & Kayra-Stuart, 1976; Buckley & Gillman, 1974; Duncan & MacFarland, 1980; Moyer & Landauer, 1976; Parkman, 1971; Sekuler & Mierkiewicz, 1977), patterns of dots (Buckley & Gillman, 1974), and written-word numerals (Foltz, Poltrock, & Potts, 1984). The number size (or serial-position) effect refers to the fact that, for a given symbolic distance, pairs of small numbers are compared faster than pairs of large numbers. Again, this is a classic result which has been observed with arabic numerals (Buckley & Gillman, 1974; Parkman, 1971), with patterns of dots (Buckley & Gillman, 1974), and with written-word numerals (Foltz et al., 1984).

We present below a connectionist model of number comparison, which can account for the main findings with a very simple processing architecture and a very limited set of basic assumptions.

Representation of number magnitudes

What representations are used in the comparison task to select the larger? The discoverers of the distance effect in numerical judgments, Moyer and Landauer (1967) state that "the decision process ... is one in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line." (p. 1520). They carefully separate the process of *conversion* of symbols to analogue representations from the process of *deciding* which is the larger. It is frequently assumed, however, that the key parametric findings should attributed to the process of conversion alone. The rationale for this position is very clearly stated by Dehaene who, in a series of papers, argues that the mapping from numerals to the number line is non-linear. This is because the line is held to be compressive, that is obeying Weber-Fechner logarithmic law. Accordingly, the subjective difference between two numbers will depend on their positions on the line, that is, the subjective difference between N and N+1 will be smaller as N increases (Dehaene, 1992; Dehaene, Dupoux & Mehler, 1992).

The similarity between numerical judgments and physical judgments has struck many other authors including Gallistel and Gelman (1992), who argue that the *mechanism* of comparing the magnitudes, again conceptualised as analogue, of two numbers is equivalent to comparing the lengths of two lines. However, their conception of analogue magnitude is subtly but importantly different from Dehaene's. Gallistel and Gelman (1992) propose that the mapping from number symbol (word or numeral) to the magnitude representation is linear, not compressive, but the variability of the mapping increases in proportion to the magnitude. For this reason, "the discriminability of the two numbers decreases as their mean numerical value increases, not because they are subjectively closer together, but because the variability (noise) in the mapping is scalar." (p. 57).

Analogue representations, however, fail to capture our intuitive notion of whole numbers, and whole-number arithmetic. Perhaps because of our early experience with counting, we intuitively think of whole numbers as meaning not approximate analogue magnitudes, but discrete numerosities.

In particular, we think of a whole number as denoting the numerosity (or cardinality) of a set with discrete members. Intuitively, we think of arithmetical operations on whole numbers in terms of sets and numerosities. For example, we think of the addition of x and y as being the numerosity of the union of two disjoint sets whose numerosities are x and y (Giaquinto, 1995).

Our working hypothesis is, then, that number representations are ordered by numerosity: smaller numbers denote proper subsets of the sets denoted by bigger numbers. Notice that this hypothesis is not trivial. If we conceptualise numbers essentially as words, then they will not be intrinsically ordered by numerosity magnitude; they will be instead be intrinsically ordered along some verbal dimension, such as the alphabet. Even ordinal numbers are not ordered by magnitude - the first past the post is not smaller than the second past the post, even though 1 is smaller than 2.

Our principal question is which aspects of the comparison phenomena should be attributed to the representation of numerical magnitudes and which to the implementation of a decision process. The reaction time data for the judgement of physical magnitudes across a wide range of domains (e.g., line length, pitch, weight) are well represented by the equation proposed by Welford (1960):

$$RT=a+k \log[L/(L-S)]$$

where L and S are the larger and the smaller physical magnitudes, respectively, and a and k are constants. The same equation has been found to be the best predictor of number comparison reaction time data, accounting for about 50% of the variance (e.g., Moyer & Landauer, 1973; Dehaene, 1989). Given this striking similarity, it is unclear why the experimental effects found in number comparison should be attributed to the *representation* of numerical magnitudes rather than to the decision process *per se*.

This lead us to the issue of *implementation*. Within a connectionist framework, a two-choice decision process can be implemented by two nodes that compete with each other for responding to the input (e.g., Zorzi & Umiltà, 1995). What is less straightforward, however, is how to represent numbers as activation patterns over a set of processing units. The analogue **"number line" hypothesis** represents number magnitudes as points or regions on a continuous psychological dimension. In one of the few attempts to model numerical processes in a neural network, that is McCloskey and Lindemann's (1989) model of multiplication facts retrieval, numbers were encoded over an ordered sequence of input nodes, where each node stood for a particular number. Moreover, the two immediate neighbours of the number were activated as well: thus 5 was represented as the activation of the node labelled "5" plus (lesser) activation of "4" and "6". Although this provides some ordering of numbers, "8" and "4", with no overlapping neighbours, would activate orthogonal representations (i.e., nodes 7-8-9 for "8" and nodes 3-4-5 for "4"). McCloskey and Lindeman (1989) did not however attempt to model number comparison, and it is not clear how it would succeed in capturing the distance effect. In any event, this representation is very different from, and incompatible with, a numerosity representation (see Figure 1).

Our approach is to represent **numerosity magnitude** straightforwardly as the number of units activated, such that bigger numbers include smaller numbers; therefore, for N>M, a set with M members can be put in 1-1 correspondence with a proper subset of the set with N members. This representational scheme is also known as a "thermometer" representation (see Figure 1, right panel).

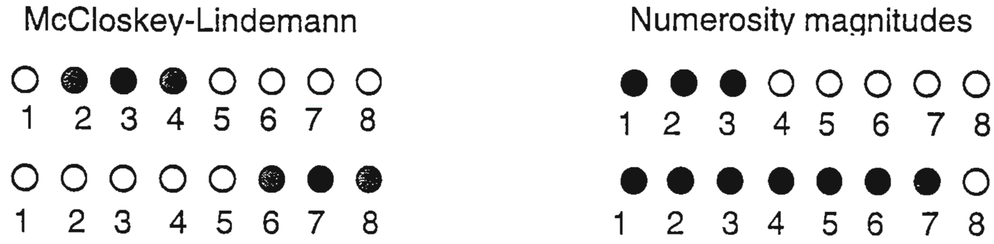

Figure 1: Alternative schemes for representing numbers (top row: 3; bottom row: 7). On the left is the McCloskey-Lindemann scheme in which an input number activates its own representation and, to a lesser extent, its immediate neighbours. On the right is our numerosity (i.e. the cardinality) representation, where each number is represented by a set of activated units corresponding to its numerosity.

The numerosity representation just described has several advantages. First, it readily maps onto lower level perceptual processes (e.g., object identification) and enumeration procedures (e.g., subitizing, counting). That is, each magnitude increment in our numerosity representation corresponds to the enumeration of a further element in the to-be-counted set. Second, it entails that larger numbers are more similar to each other than smaller numbers, without assuming a logarithmic compression, since large numbers share more active nodes. For example, 9 and 8 would share 8 nodes, whereas 1 and 2 would share only 1 node. This can also be formalized in terms of the cosine of the angle formed by the vectors coding the two numbers. Finally, we do not assume that the variability of the mapping from symbols to magnitude representation increases with size as Gallistel and Gelman (1992) proposed. Rather, the mapping in our scheme is linear and not noisy.

Model of number comparison

Architecture

The model is implemented in a network, in which each node is associated with an activation value. Nodes are connected by weighted links, which may be excitatory (positive) or inhibitory (negative).

We assume that number comparison is performed on the basis of magnitude, "semantic" codes (e.g., McCloskey et al., 1985; Dehaene, 1992). Therefore, this level of representation is used as input data for the model. The numbers to be compared are each represented by a set of 9 nodes, which are activated according to the "numerosity magnitude" scheme discussed above. The representation of the two possible responses (left or right button-press, to indicate which of the numbers is the larger) consists of two nodes, which we call the "response system" (see Zorzi & Umiltà, 1995). Activation values of the input nodes (number magnitudes) are in the [0,1] range, whereas in the response system suppressed states (modelled as negative activations) are permitted. In this case the range of activation is [-1,1], but only positive activations propagate through the connections. At stimulus onset, the relevant input nodes are clamped to the "1" value.

The response system incorporates a competitive mechanism, that, via lateral inhibition, implements response competition (e.g., Zorzi & Umiltà, 1995). Thus, the response system can be represented as a dipole, where two mutually exclusive responses compete: each response node has an inhibitory connection to the other node. The state of each response node changes smoothly over time in response to influence (of both excitatory and inhibitory kind) from the other nodes of the network (magnitude nodes for the two numbers, and the other response node). For simulation purposes, continuous time units can be approximated with discrete time units, in which time is discretized into ticks of some duration τ. Therefore, the new state of each response node at time $t+\tau$ is a weighted average of its current state at time t and the state dictated by its external input, according to the following equations:

$$a_j^{[t+\tau]} = \tau\sigma(net_j) + (1-\tau)a_j^{[t]} + \eta \quad (1)$$

where a_j is the activation level of the response node j, and net_j is the external input to j, and τ is a parameter defining the weighting proportion that determines how gradually the state of the node changes over time. Note that η represents a small random noise, which gives the model non-deterministic (stochastic) properties.

To bound the activation values in the range [-1,1], the function $\sigma(x)$ in (2) is a S-shaped squashing function (hyperbolic tangent):

$$\sigma(x) = \frac{2}{1+e^{-\lambda x}} - 1 \quad (2)$$

where λ is a "temperature" parameter defining the sigmoidal shape of the function. The net input (external input) to the response nodes is given by:

$$net_j = (\sum_i w_{ij} o_i) - w^- a_k \quad (3)$$

where w_{ij} is the weight of the connection from the input node i to the response node j, o_i is the output of the input node i, a_k is the activation of the other response node, and

w^- is the weight of the inhibitory link from the other response node. Free parameters for all simulations reported in this paper: λ=4, τ=0.01, η=random gaussian noise (mean = 0, standard deviation = 0.01).

After stimulus onset (i.e., activation of the relevant numerical magnitudes), the system is allowed to cycle until there is a winning node. We simply assume that a response occurs when the difference ~~between~~ the activations of the two response nodes exceeds a certain threshold. The number of cycles required by the system to settle represents a measure of the reaction time (RT) in responding to the stimulus.

The connections linking the magnitude nodes for the two numbers with the response system are learnt in the model by simple association of the input with the required response. This is done in the model by simple "hebbian learning": a connection is strengthened if the activation of the nodes that it connects are correlated (e.g., both nodes are active). Formally:

$$w_{ij} = \varepsilon \, a_i a_j \tag{4}$$

where w_{ij} is the weight of the connection between the nodes i and j, and a_i and a_j are the activation values of the two nodes, and ε is a small learning rate.

Learning is done in a "one-shot" fashion, in the sense that the 72 possible input patterns (i.e., all combinations of two 1-9 digits, excluding the ties) are presented just once, simultaneously with the required response. The connections involved in the learning phase are those from the magnitude nodes to the response nodes. The model's architecture, with its nodes and connections, is depicted in Figure 2. Not surprisingly, the connections linking uncorrelated nodes are not strengthened.

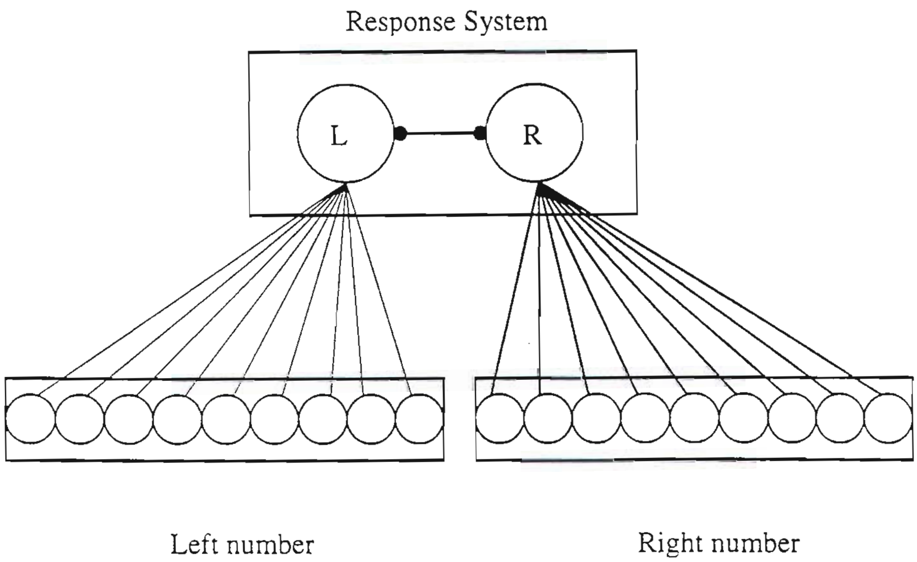

Figure 2: Architecture of the model. The connection linking the two response nodes is an inhibitory dipole implementing response competition. All connections from magnitude representations to response system are excitatory.

Results

The model was presented with the numerosity representations of all possible pairs of single digit numbers (1-9). Activation of the magnitude nodes propagates gradually to the response nodes, and the model is allowed to cycle until response criterion is reached, which consists in the difference of the activations of the two response nodes; we assume that a response can be unambiguously selected when this difference becomes equal to [0.5] or bigger. At that point, the number of cycles needed by the system to reach response threshold is taken as a measure of the reaction time. Crucially, this will in turn depend on the amount of competition between the two nodes. Note that response competition is what accounts in general for the relevant part of empirical RTs, across domains as different as attention (e.g., Cohen & Huston, 1994; Houghton & Tipper, 1994; Zorzi & Umiltà, 1995) and reading aloud (e.g., Zorzi, Houghton, & Butterworth, 1998).

The presence of noise in the activation function of the response nodes implies that the model can exhibit a relative variability in the response times. Therefore, each pair of numbers is presented 100 times to the model, and a mean RT is computed for each pair. The mean RTs produced by the model are analysed by regressing the standard structural variables (the magnitudes of the two numbers, and their difference) onto them. The two main effects that are usually found in a number comparison task are the distance or "split" effect (i.e., RTs increase as the difference between the two numbers becomes smaller) and the number size effect (i.e., RTs increase as the size of the two numbers increases). A variable that is standardly used to index the latter effect is the sum of the two numbers. A linear regression onto the model's RTs showed that the split accounts for 40.3% of the variance (p<.001) and the sum accounts for 57.6% of the variance. A different variables that has been often reported as a good predictor of comparison times is the Welford function, i.e. Log (Larger-Smaller)/Larger)). Used as a predictor of the model's RTs, the Welford functions accounts for 88.3% of the variance. As with human performance, the reaction times produced by the model are sensitive to the difference between the two numbers (i.e., split) and to the overall size of the two numbers. The effect of the split can be seen more clearly in Figure 3.

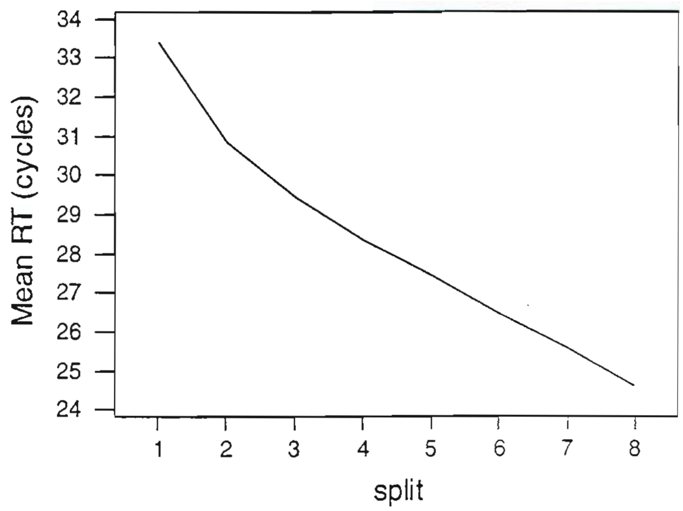

Figure 3: Graph shows the effect of the split (difference between the two numbers) on the number of cycles that the system needs to select a response.

General Discussion

We have modelled number comparison for the first time, and have done it in such a way that seems to capture out intuitive understanding of whole numbers. This shows that analogue representations of number magnitudes are not necessary to fit the data from comparison tasks, as has been often claimed. The crucial point, however, is that magnitude representations need not be compressed in order to observe a Weber-Fechner logarithmic effect in number comparison, contrary to the claims of Dehaene and his colleagues. In our simulation, numerals were mapped linearly on to magnitude representations, and the compressive effect on the comparison times emerges by virtue of the non-linear interactions that are intrinsic to the decision process itself. The nonlinear squashing function in the *response units* produces a compression of the input magnitudes which is larger for bigger numbers. It is also not necessary to postulate that magnitude representations have the property of scalar variability, that is, that the standard deviation of mapping from numerals to magnitudes increases with the mean magnitudes of the numbers, as claimed by Gallistel and Gelman (1992).

There are two main psychological advantages of linear mapping from number symbols to number magnitudes. For one thing, it corresponds to our intuitive notion that each counting increment when enumerating a set of objects is equivalent, regardless of the size of the set. Secondly, these magnitudes are appropriate for arithmetical operations on whole numbers, which are linear, whereas compressive representations would not be. As Gallistel and Gelman note, "the concatenation of mental magnitudes is isomorphic to addition of the corresponding values." (1992, p. 57).

Acknowledgments

This research was supported by Wellcome Biomedical Research Collaboration Grant 048004 and Wellcome Grant 045013.

References

Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgements of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 435-447.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282*, 746-749.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103*. 1131-1136.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In: C. Umiltà & M. Moscovitch (Eds.), *Attention & Performance XV* (pp. 453-475). Cambridge (MA): MIT Press.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics, 45,* 557-566

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1-42

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

Dehaene, S., & Cohen, L. (1995). Towards and anatomical and functional model of number processing. *Mathematical Cognition, 1,* 83-120.

Dehaene, S., Dupoux, E., & Mehler, J. (1992). Is numerical comparison digital: Analogical and symbolic effects in two-digit number comparison. *Jorrnal of Experimental Psychology: Human Perception and Performance, 16,* 626-641

Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology, 14,* 613-636.

Duncan, E. M., & McFarland, C. E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: an additive factors analysis. *Memory and Cognition, 8,* 612-622.

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 442-453.

Houghton, G., & Tipper, S. P. (1994). A model of inhibitory mechanisms in selective attention. In D. Dagenbach & T. Carr (Eds*), Inhibitory Mechanisms in Attention, Memory and Language*. San Diego, CA: Academic Press.

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

Giaquinto, M. (1995). Concepts and calculation. *Mathematical Cognition, 1,* 61-81.

McCloskey, M. (1992) Cognitive mechanisms in number processing: Evidence from acquired dyscalculia. *Cognition, 44,* 107-157 .

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4,* 171-196.

McCloskey, M., & Lindemann, A. (1992). Mathnet: preliminary results from a distributed model of arithmetic fact retrieval. In J.I.D. Campbell (Ed.), *The nature and origins of mathematical skill* (p. 365-409). Amsterdam: Elsevier.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215,* 1519-1520.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology, 91,* 191-205.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgements of numerical inequality. *Child Development, 48,* 630-633.

Washburn, D. A., & Rumbaugh, D. M. (1991). Ordinal judgments of numerical symbols by macaques (*Macaca Mulatta*). *Psychological Science, 2,* 190-193.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358,* 749-751.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131-1161.

Zorzi, M. & Umiltà, C. (1995). A computational model of the Simon effect. *Psychological Research, 58,* 193-205.