A model of exact small-number representation

TOM VERGUTS, WIM FIAS, and MICHAËL STEVENS
Ghent University, Ghent, Belgium

To account for the size effect in numerical comparison, three assumptions about the internal structure of the mental number line (e.g., Dehaene, 1992) have been proposed. These are magnitude coding (e.g., Zorzi & Butterworth, 1999), compressed scaling (e.g., Dehaene, 1992), and increasing variability (e.g., Gallistel & Gelman, 1992). However, there are other tasks besides numerical comparison for which there is clear evidence that the mental number line is accessed, and no size effect has been observed in these tasks. This is contrary to the predictions of these three assumptions. Moreover, all three assumptions have difficulties explaining certain symmetries in priming studies of number naming and parity judgment. We propose a neural network model that avoids these three assumptions but, instead, uses place coding, linear scaling, and constant variability on the mental number line. We train the model on naming, parity judgment, and comparison and show that the size effect appears in comparison, but not in naming or parity judgment. Moreover, no asymmetries appear in primed naming or primed parity judgment with this model, in line with empirical data. Implications of our findings are discussed.

How are numerical values mentally represented? The obvious way in which to investigate this issue behaviorally is to use tasks that rely critically on number magnitude, the most straightforward task being numerical comparison. This approach dates back at least to Moyer and Landauer's (1967) seminal article that described two critical factors determining number comparison performance: the distance effect and the size effect. Most current theories are rooted in this work.

The distance effect reflects the fact that comparison times are shorter for a larger numerical distance between two numbers that have to be compared. For example, comparison is faster for 2 and 6 than for 2 and 3. The second factor is the size of the numbers: For a given distance, comparison is faster for smaller numbers (e.g., 2 and 4) than for larger numbers (e.g., 7 and 9). The distance and size effects affect not only processing speed but also accuracy. The size and distance effects are robust phenomena and occur with numbers presented in various formats—for example, Arabic notation (Banks, Fujii, & Kayra-Stuart, 1976; Dehaene, Dupoux, & Mehler, 1990; Sekuler, Rubin, & Armstrong, 1971), verbal notation (Koechlin, Naccache, Block, & Dehaene, 1999), and nonsymbolic notation, such as collections of dots (Buckley & Gillman, 1974).

Together, the two effects put crucial constraints on how number magnitude is represented and processed in various tasks. A generally accepted idea is that mental number representations can be seen as organized along a mental number line—that is, a set of units in which close

numbers are represented with overlapping distributions of activation. This number line assumption can explain the distance effect, because numbers close to each other (e.g., 1 and 2) have more distributional overlap than do numbers that are more distant (e.g., 1 and 4), and it will be more difficult to discriminate the closer numbers.

To account for the size effect, additional assumptions are needed. One assumption that makes it possible to explain the size effect is *magnitude coding*. This means that the mental code for a number is analogous to the magnitude it represents. For instance, if a given number activates a set of units on the mental number line, this set of activated units is a subset of the units activated for a larger number (e.g., Zorzi & Butterworth, 1999; see Figure 1A for a graphical representation). With magnitude coding, number comparison is similar to physical magnitude discrimination, which is robustly characterized by distance and size effects, as has long been known from psychophysics (e.g., Weber's law; Festinger, 1943). Another possible assumption that makes it possible to account for the size effect is *compressed scaling* (e.g., logarithmic) of numbers. With compressed scaling, the distances between numerical representations on the number line are smaller for larger numbers (e.g., Brysbaert, 1995; Dehaene, 1992; see Figure 1B for a graphical illustration). Compressed scaling can explain the size effect, because small numbers are represented as being further apart from each other than larger numbers (for a fixed numerical distance). Therefore, small numbers are easier to disambiguate than large numbers. A third possible assumption regarding the source of the size effect is *increasing variability*. Consider the Gaussian curves depicted in Figure 1C: Units close to the maximally activated unit (the unit over which the curve is centered) are also activated, and the amount of activation depends on the distance between the two units. This may be interpreted in
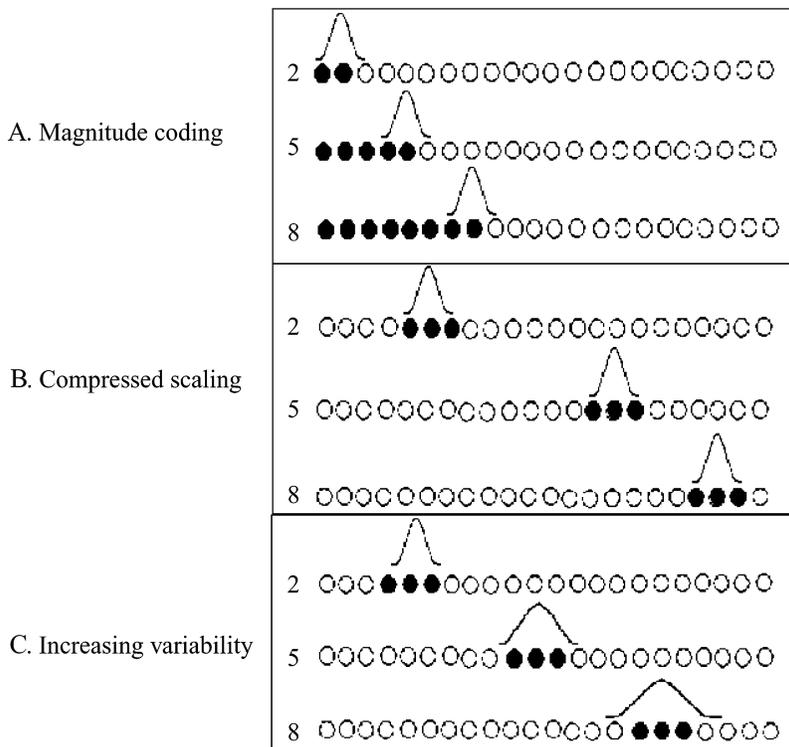
Figure 1. Graphical representation of (A) magnitude coding, (B) compressed scaling, and (C) increasing variability.

either a stochastic sense (units further away are activated with smaller probability) or a deterministic sense (units further away are activated less strongly). The assumption of increasing variability entails that the standard deviation of these Gaussian curves increases with increases in the size of the number. This also leads to a size effect, since larger numbers have larger variability and, hence, the activation distributions overlap more than for smaller numbers. In existing models and theories of numerical cognition, one or more of these principles are used to account for the size effect—for example, magnitude coding (Zorzi & Butterworth, 1999), compressed scaling (Dehaene, 1992), and magnitude coding and increasing variability (Gallistel & Gelman, 1992).

Importantly, these three assumptions would seem to imply an asymmetry between small and large numbers not only in the case of number comparison (as reflected in the size effect), but also whenever the number line is involved in the task. With compressed coding and increasing variability, this is attributable to inferior discrimination of large numbers relative to small numbers. With magnitude coding, there is also an asymmetry on the number line, in that a representation of a larger number implies the representations of all the smaller numbers, whereas a smaller number does not imply the full representation of a larger number (see Figure 1A).

However, the prediction of generalized asymmetry between processing small and large numbers is inconsistent with two empirical findings. The first finding relates to

distance-related priming, which reflects prime–target interactions at the level of the number line (Dehaene, 2004). The finding is that equal prime–target distances lead to equal priming effects. In particular, a prime that is *n* units smaller than a target leads to the same amount of priming as a prime that is *n* units larger than the target (e.g., prime 3 with target 5 vs. prime 7 with target 5). In this sense, priming is symmetric. This means that, for example, the response time (RT) to name the target number 5 is the same whether it has been primed by 3 or by 7 (Reynvoet & Brysbaert, 1999, 2004; Reynvoet, Brysbaert, & Fias, 2002). Besides number naming, a similar phenomenon is observed in parity judgment (judging 5 to be odd takes the same amount of time whether it has been primed by 3 or 7; Reynvoet & Brysbaert, 1999, 2004; Reynvoet, Caessens, & Brysbaert, 2002). Importantly, symmetric priming has been observed in various combinations of prime–target modalities (Arabic–Arabic, Arabic–verbal, verbal–Arabic, and verbal–verbal; see, e.g., Reynvoet, Brysbaert, & Fias, 2002; Reynvoet, Caessens, & Brysbaert, 2002). The cross-modal priming pattern refutes the interpretation that the priming effects derive from a format-specific processing stage and provides additional evidence that distance-related priming originates from a modality-independent mental number line (Dehaene, 2004).

The second finding that raises problems for the assumption of generalized asymmetry is that a size effect is predicted whenever the number line is involved. Again,

this conflicts with empirical observations. Whereas a size effect is routinely observed in number comparison, this is not the case for number naming (at least for numbers up to 20; see Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Reynvoet, Brysbaert, & Fias, 2002) or parity judgment (Dehaene, Bossini, & Giraux, 1993; Fias, Brysbaert, Geypens, & d'Ydewalle, 1996; Reynvoet, Caessens, & Brysbaert, 2002). Despite the fact that these tasks are subject to distance-related priming effects, indicating the involvement of the mental number line, processing time does not increase with increases in number size. This raises problems for magnitude coding, compressed scaling, and increasing variability, because all of these predict a size effect whenever the mental number line is involved. Indeed, the size effect is the rationale behind these assumptions.

Hence, magnitude coding, compressed scaling, and scalar variability have difficulty accounting for symmetric priming in number naming and parity judgment, in conjunction with a size effect in comparison, but not in naming or parity judgment. Four arguments could be raised against this view. First, one could argue that naming and parity judgment are not semantically mediated. However, as was noted by Dehaene (2004), the distance dependency of the priming effect (primes further away from the target gradually lead to smaller priming) is in itself convincing evidence that these two tasks are semantically mediated. Second, one could argue that there is not only a semantic route to name Arabic numbers, but also an asemantic route, and that the asemantic route is dominant. As a consequence, the distance-related priming would be attributed to automatic activation of the semantic route without the critical transcoding processes explicitly relying on it. However, it is unclear why the distance effect, but not the size effect, from this asemantic route would influence naming times. Moreover, it then remains to be explained why the priming effects observed in number comparison (e.g., Koechlin et al., 1999), where the semantic route is the only one available, are of the same order of magnitude as the effects observed in naming and parity judgment. Moreover, this still does not solve the issue involved in the fact that the priming pattern is symmetric. As a third counterargument, one could invoke additional processing mechanisms that straighten out the asymmetric priming effects. However, such additional mechanisms can be considered valid only if they do not lead to a size effect. For instance, from the perspective of magnitude counting, one could postulate that larger numbers take more time to *settle* on the number line. The settling times could then be chosen so that the inherently smaller priming effects for smaller numbers (as predicted by magnitude coding) are compensated by the fact that settling is faster for smaller numbers. As a result, the priming effects in number naming and parity judgment would be (approximately) symmetric. However, it then has to be demonstrated that this delicate balancing can work for all targets simultaneously, since the

symmetry is observed in all the targets (Reynvoet, Brysbaert, & Fias, 2002; see below). Furthermore, a problem that is introduced with this procedure is that one has to predict that RTs are longer for larger targets (i.e., a size effect), and, as was discussed above, this has not been observed empirically. It is precisely the combination of symmetry in distance-related priming and the absence of a size effect in the same task that poses a serious problem for magnitude coding. And finally, a fourth counterargument consistent with increasing variability and/or compressed scaling is to assume that the increase in variability or the compression factor is negligibly low for small numbers and that, therefore, (approximate) symmetries are obtained in priming studies. However, if the increase in variability or in the amount of compression is too small to be detected, why then is the size effect obtained so robustly in number comparison? After all, increasing (scalar) variability and compression were introduced precisely to account for the size effect.

In sum, none of the four counterarguments explains why number priming is symmetric and why no size effect is observed in tasks that nevertheless show involvement of the mental number line. Therefore, in order to accommodate these findings in a unified framework, one needs a model that lacks number line compression, magnitude coding, and increasing variability. Below, we will propose such a model. It assumes a linear (i.e., noncompressed) *place-coding* representation with fixed variability. It will be shown that this model generates symmetric patterns of distance-related priming. We will further argue that the size effect is a consequence of *nonlinear mappings* from the number line to the comparison output system and that these nonlinear mappings derive from the lower frequency of larger numbers, observed in daily life (Dehaene & Mehler, 1992).

In the following section, we first will outline the structure of our model. We then will train the model in number naming, parity judgment, and number comparison and will show that nonlinear mappings (which can generate a size effect) develop from the mental number line to the comparison output system, but not to the naming or the parity output system. Thereafter, performance of the model will be described in primed number naming, in primed parity judgment, and in number comparison.

A MODEL OF SMALL-NUMBER REPRESENTATION

From the Input to the Number Line Field

The input field consists of units each of which represents an Arabic number; unit 1 in the input field responds to presentation of the (Arabic) number 1, unit 2 to number 2, and so on (see Figures 2A, 2B, and 2C for a graphical illustration). It is important to note that there is no further internal organization (e.g., lateral connections) at this level. Activation in the input field is propagated to the number line field.

The number line field consists of a set of units, one for each number, and is characterized by the following principles. It uses a *place-coding* system (as opposed to magnitude coding), so each number activates an equal amount of units on the number line (see the activation curves in Figures 2A, 2B, and 2C). There is also *linear scaling* in the mapping from the input field to the number line, meaning that the distance between numbers on the number line is equal for all the numbers with a fixed distance. The model is further characterized by *constant variability* in the mapping: When a number is presented, it activates its corresponding unit on the number line with maximal strength, but it also activates neighboring units with gradually decreasing strength as a function of distance. Importantly, the activation functions expressing this spreading of activation are constant over numbers. In other words, the activation curves have the same width for each number. This is called a model of *exact* number representation, because it preserves distances of the integer number line (e.g., after mapping onto the mental number line, the distance between 8 and 7 is the same as the distance between 2 and 1).

The input and number line properties, as described above, were formally implemented as follows. The activation of the ith unit in the input field (with subscript A for *Arabic*) is denoted $x_{Ai}(t)$, and it obeys the following equation:

$$\frac{d}{dt} x_{Ai}(t) = -x_{Ai}(t) + I_{Ai}(t), \qquad (1)$$

in which $I_{Ai}(t)$ is a function equal to 1 if Arabic number i is presented at time point t and is 0 otherwise. Equation 1 implies that if Arabic number i is not presented (so $I_{Ai} = 0$), the activation x_{Ai} decreases, because then $d/(dt) x = -x$, which implies exponential decrease toward zero. Hence $-x$ in the right-hand side of Equation 1 is a decay term. On the other hand, if the number i is presented ($I_{Ai} = 1$), it follows that $d/(dt) x = -x + 1$, and x tends toward its asymptotic value of 1.

The amount of activation received by unit j in the number field from a unit i in the input field depends on the connection strength c_{ij} between these two units. Note that each input number activates exactly one of the number line units with maximal strength: In this sense, this number line unit may be considered as the unit corresponding to that particular number.

We assume that processing is *thresholded* in the sense that a unit sends activation to another unit only if the activation of the first unit exceeds a fixed threshold θ. The relevant equation for a number line unit j with activation $x_{Nj}(t)$ is as follows:

$$\frac{d}{dt} x_{Nj}(t) = -x_{Nj}(t) + \sum_{i=1}^{n} \exp(-|i-j|) \lceil x_{Ai}(t) - \theta \rceil^{+}, \quad (2)$$

where $\lceil A \rceil^{+}$ denotes the maximum of zero and A (this is thresholded processing). The reason for this assumption

will be explained later. As in Equation 1, $-x_{Nj}$ in the right-hand side of Equation 2 functions as a decay term. Concerning the summation in the right-hand side of Equation 2, note that only the term in this summation corresponding to the active input unit i can become positive. For other input units, the factor $\lceil x - \theta \rceil^{+}$ equals zero. The factor $\exp(-|i-j|)$ formalizes the activation curves depicted in the number line fields of Figures 2A, 2B, and 2C. Due to this factor, (active) input units i have a strong influence on number field units j that are of similar size (e.g., if $j = i$) but have much less influence if i and j are far apart. For example, the activation of Arabic input 3 (i.e., x_{A3}) has a large impact on number field unit 3 (which has activation $x_{Nj} = x_{N3}$), because $\lceil x_{A3}(t) - \theta \rceil^{+}$ in Equation 2 is multiplied by the (large) factor $\exp(-|3-3|) = 1$. On the other hand, x_{A3} has much less influence on number field unit 6, because in this case $\lceil x_{A3}(t) - \theta \rceil^{+}$ in Equation 2 is multiplied by the (small) factor $\exp(-|3-6|) \approx 0.05$. This difference leads to the activation curves in Figure 2. The fact that $\exp(-|i-j|)$ depends only on the absolute distance $|i-j|$, and not on the individual numbers i or j, leads to constant variability. The parameter θ was arbitrarily set at 0.08.

The number line started from 1, rather than from 0, because it has been argued on empirical grounds that 0 has a special status and is not represented on the number line (Brysbaert, 1995; Butterworth et al., 2001). The endpoint was set at 15, for the following reasons. Butterworth et al. (2001) explicitly demonstrated that there is no size effect in the naming of numbers from 1 to 18, even after partialing out possible noise-inducing variables. For larger numbers, on the other hand, there does appear to be a size effect in number naming (Brysbaert, 1995). Hence, only small numbers are modeled. Another relevant finding is that of Reynvoet and Brysbaert (1999), who showed that distance-related priming extends over the 10s decade (e.g., with target 11, the effect of prime 9 is the same as the effect of prime 13). On the basis of the above, we decided to put the boundary somewhere between 10 and 20. The choice of 15 was then arbitrary, but it should be noted that the precise extension does not matter very much beyond a certain point. Due to the exponential function used in Equation 2, if a small number is presented (e.g., 5), number field units corresponding to large numbers will be activated extremely weakly. If only numbers up to 12 are used (as in the experiments we describe in this article), it is immaterial whether the number line is extended to 15, 16, or some other, higher number. The issue of processing large numbers is taken up in the General Discussion section.

From Number Lines to Output Fields

The basic model must be extended for each task, depending on the set of output components that are needed. These components will be described next.

Naming. For number naming, there is one output unit for each possible number (corresponding to phonologi-

A. Naming

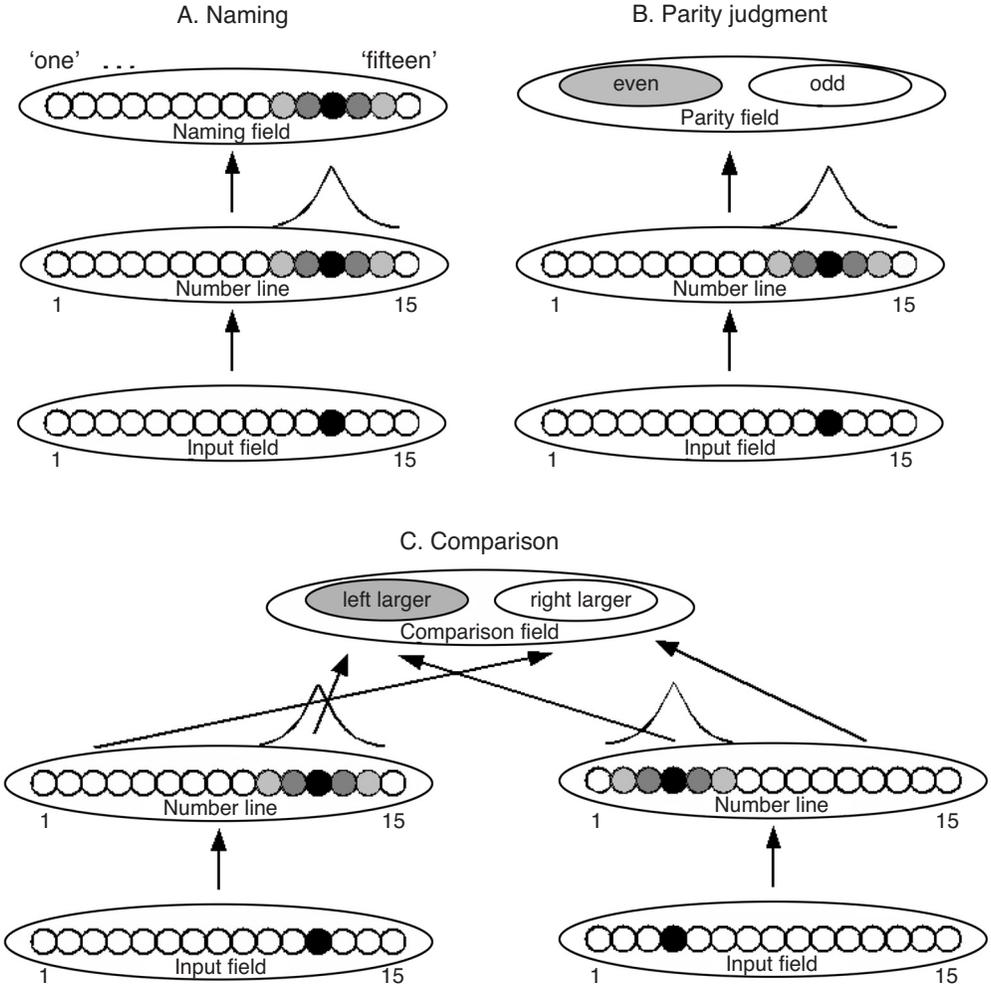## B. Parity judgment

C. Comparison

Figure 2. Input fields and number line fields with the output fields for (A) naming, (B) parity judgment, and (C) number comparison.

cal lexical units *one*, *two*, and so on; see Figure 2A). The equation for the activation $x_{Sj}(t)$ of unit j (ranging from 1 to 15) in the naming output field, is

$$\frac{d}{dt}x_{Sj}(t) = -x_{Sj}(t) + \sum_{i=1}^{15} w_{Sij}\lceil x_{Ni}(t) - \theta \rceil^{+}, \qquad (3)$$

where we use the index S (for *spoken*) to index both naming field units and weights going to the naming field. The right-hand side of Equation 3 should be understood in a way similar to Equations 1 and 2. For example, if the connection between number line unit i and naming unit j is strong (w_{Sij} is large), the term $\lceil x_{Ni}(t) - \theta \rceil^{+}$ is multiplied by a large factor, and hence, the activation of number line unit i has a strong influence on naming unit j. An appropriate choice of the weights w_{Sij} ensures that the correct response will be given by the model. The connection values were determined by training the mappings from the number line to the naming output field. The training procedure and resulting weight values are described

in the Training Connection Weights From Number Lines to Output Fields section.

Parity judgment. For parity judgment, there is an output unit for the *even* response and an output unit for the *odd* response (see Figure 2B). Otherwise, the reasoning is exactly the same as that for number naming. The equation for activation $x_{Even}(t)$, the activation of the *even* unit in the parity field, is

$$\frac{d}{dt}x_{Even}(t) = -x_{Even}(t) + \sum_{i=1}^{15} w_{Pi,Even}\lceil x_{Ni}(t) - \theta \rceil^{+}.$$

The equation for the *odd* response is similar. As in naming, the weights $w_{Pi,Even}$ and $w_{Pi,Odd}$ ensure that a correct response will be given, and they are determined by training (see the Training Connection Weights From Number Lines to Output Fields section).

Comparison. In the case of number comparison, there are two number fields, one for each number (see Figure 2C). The output components depend on the task in-

structions (*choose larger* or *choose smaller*). For definiteness, we assume the *choose larger* instruction. Results are similar for other numerical comparison designs. Hence, the relevant units are for *left-larger* and *right-larger* responses, and the equation for the left-larger unit is

$$\frac{d}{dt} x_{\text{Left}}(t) = -x_{\text{Left}}(t) + \sum_{i=1}^{15} w_{\text{M}i,\text{Left}} \left\lceil x_{\text{M}i}(t) - \theta \right\rceil^{+}$$

$$+ \sum_{i=1}^{15} w_{\text{N}i,\text{Left}} \left\lceil x_{\text{N}i}(t) - \theta \right\rceil^{+}. \qquad (4)$$

In this equation, a summation is taken over the two copies of the number line (indicated by M and N, respectively). A similar equation applies to the right-larger unit. Appropriate weight values (e.g., $w_{\text{M}i,\text{Left}}$, $w_{\text{M}i,\text{Right}}$) were again determined by training the model on this task (see the Training Connection Weights From Number Lines to Output Fields section).

Model Dynamics

Upon presentation of a stimulus, all the activation values $x(t)$ of the network start changing according to the relevant equation (e.g., Equations 1 and 2). When the activation value of one of the output field units reaches a fixed threshold, the corresponding response is assumed to be chosen by the model. This threshold parameter was set at 0.5. The response chosen and the time taken to reach that choice are recorded. The procedure for training the model on the three tasks and the weights obtained from this training will be described in the next section.

TRAINING CONNECTION WEIGHTS FROM NUMBER LINES TO OUTPUT FIELDS

Procedure

Initial weights were sampled from a uniform distribution between zero and one. They were adjusted with a supervised learning rule. For this purpose, the Widrow–Hoff learning rule (Widrow & Lehr, 1990) can be used if all the equations are in a stable state (i.e., if all derivatives dx/dt are zero), which is the case if time t is sufficiently large. Hence, we assumed that during learning, participants first respond quickly with the response whose unit reaches threshold first and then wait until all the activation values are stable, at which point the weights are changed. The learning rate parameter was set at 0.02.

For each task, there were 30,000 learning trials in every simulation round. Ten simulation rounds were performed for each task. Importantly, the training regime took into account the frequencies with which numbers occur in daily life. Dehaene and Mehler (1992) observed that the frequencies of numbers decrease strongly with increases in number. Depending on language and input format (Arabic or verbal), they found either an exponential decrease (frequency of number i proportional to $e^{-\alpha i}$) or a power decrease (frequency of number i proportional to $1/i^{\alpha}$), with α larger than zero in both cases. On the basis of frequencies plotted in Dehaene and Mehler, we used exponential decrease and estimated the coefficient α to be equal to about 0.2. However, the results were very similar with different values of α or with a power decrease function.

A squared error function on the difference between teacher output values and actual values was used. The teacher value was 1 for the correct response unit and 0 for the other response units. However, it is not necessary that activation values be exactly zero for the other response units, as long as they are below the value of the correct unit. We therefore made the difference between teacher and actual response values zero, if this difference was smaller than .5. For example, if the teacher value was zero and the actual output value was .3 ($<.5$), this was considered "good enough," and the deviation between teacher and output values was set to zero.

Results

Figure 3 shows the weights from number line to output, for each task separately (naming, parity, and comparison), with weight states at three points during training (after 100, 1,000, and 30,000 trials).

Number naming. We plot the direct connection weights from each number line unit to its corresponding naming output field unit (e.g., from number line unit 2 to output unit 2), averaged over the 10 simulation rounds (Figures 3A, 3B, and 3C). As can be seen, the connections from the smaller numbers to their output units are adapted first by the learning algorithm (Figure 3B). This is because smaller numbers are more frequent in the training regime. Nevertheless, after 30,000 trials, all the weights have reached asymptote, and the connection values are the same for each number (see Figure 3C).

Parity judgment. Connection weights from each number line unit to the *even* response unit, averaged over the 10 simulation rounds, are plotted in the second row of Figure 3 (Figures 3D, 3E, and 3F). At the end of training (Figure 3F), there is an alternating pattern, so that all the number line units corresponding to even numbers have a large weight to the *even* response unit and all the other number line units have a small weight. This ensures that the task will be solved correctly. It is again the case that the weights for small numbers are adapted first (see Figure 3E). Most important, however, at the end of training, all the weights for even number line units are the same. There is also a small end effect, in the sense that number line units for the end points, 1 and 15, have a slightly different weight than do the other odd numbers. This is due to the fact that these end points have fewer immediate neighbors than the other units have. However, this endpoint discrepancy does not substantially influence any of the results. The results are similar for the *odd* response unit.

Number comparison. Weights from the number line corresponding to the left number, going to each response unit (*left larger* and *right larger*), are plotted in the third row of Figure 3 (Figures 3G, 3H, and 3I). Values are again averaged over the 10 simulation rounds. The two patterns of weights gradually diverge over trials, and at the end of training, the weights to the left-larger response unit are increasing and the weights to the right-larger re-
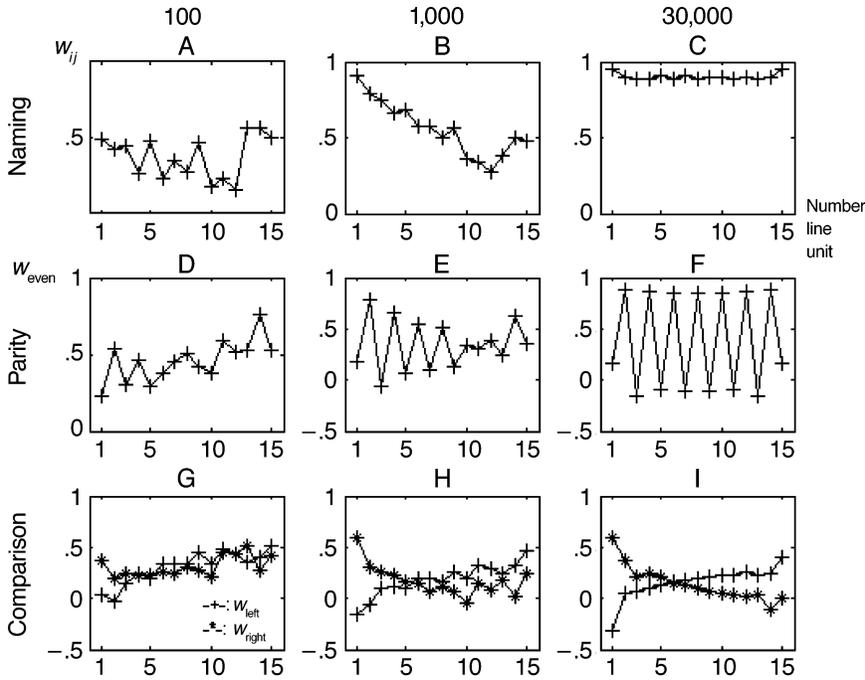
Figure 3. (A, B, and C) Evolution of weights for number naming. (D, E, and F) Evolution of weights for parity judgment. (G, H, and I) Evolution of weights for number comparison.

sponse unit are decreasing. Weights to the left-larger unit are increasing because large numbers presented on the left should activate the left-larger unit more strongly than small numbers presented on the left do. A similar reasoning holds for the right-larger weights and for the weights coming from the right number line (not shown). These increasing and decreasing trends ensure that the correct response will be given by the model.

It is important to note that the pattern of weights is *compressive*: The differences between weights are smaller for larger numbers (Figure 3I; e.g., the difference in connection weight from number line units 1 and 2 to *left larger* is larger than the difference in connection weight from number line units 6 and 7 to *left larger*). This compressive pattern of weights (and increasing or decreasing weights, depending on the specific number line and output unit) is able to generate a size effect, as will be shown below. Moreover, the pattern is obtained for a large number of settings of the training parameters: For example, it does not matter whether or not there is variability on the number line(s), thresholding is not necessary, and the value of the learning rate parameter is also unimportant. The only crucial factor appears to be the decreasing frequencies for increasing numbers during learning: If this assumption is dropped, no compressive pattern occurs (but rather, linear increases and decreases), and the model still generates a correct response and still generates a distance effect, but the size effect disappears.

To conclude, the differential frequencies of numbers in the training regime cause nonlinearities in the connection weights between the number line and the comparison

field (Figure 3I), but not in the connection weights between the number line and the naming and parity output fields (Figures 3C and 3F). As will be demonstrated below, these differential number-line-to-output mappings, in conjunction with the model's basic assumptions of place coding, linear scaling, and constant variability, make it possible to account for the symmetric pattern of distance-related priming and the lack of a size effect in naming and parity judgment and, at the same time, account for the presence of a size effect in number comparison.

In the following, we first will show statistically that priming is symmetric and that there is no size effect in the primed number-naming data in Reynvoet, Brysbaert, and Fias (2002). We then will show that the model simulation yields the same pattern of results. Subsequently, we will apply the same analyses to the primed parity judgment data in Reynvoet, Caessens, and Brysbaert (2002) and will show that here, also, priming is symmetric and there is no size effect. This is again in line with the performance of our model. Finally, we will describe how the model fares in accounting for the distance and size effects in number comparison by comparing its performance with behavioral data obtained by Schwarz and Stein (1998).

NUMBER NAMING

For number-naming data to bear on the issue of number line representations, it is, of course, crucial that naming be semantically mediated (i.e., mediated by the mental number line). Although semantic mediation in general

word naming is a controversial issue, semantic mediation in number naming is attested by the presence of distance-related priming effects in overt number naming (Brysbaert, 1995; Reynvoet & Brysbaert, 1999; Reynvoet, Brysbaert, & Fias, 2002). There are some neuropsychological data suggesting that an asemantic route might exist (Cipolotti & Butterworth, 1995) and can be damaged selectively. However, the experimental evidence in healthy participants suggests that such an asemantic route is slow and not influential in normal number naming.

In verbal number naming, it is less clear whether naming is predominantly semantically mediated (Fias, 2001; Fias, Reynvoet, & Brysbaert, 2001). Therefore, in the following, we will address only the processing of Arabic numerals and will make no explicit claims regarding the processing of verbal numerals (although for prime–target combinations involving the verbal modality, the results are highly similar; see Reynvoet, Brysbaert, & Fias, 2002).

Behavioral Data: Additional Analyses of Reynvoet, Brysbaert, and Fias (2002)

The experiment of Reynvoet, Brysbaert, and Fias (2002) will now be described in some detail, because new analyses of these data will be described. Targets ranged from 4 to 9. Primes were chosen from the target minus 3 to the target plus 3 (e.g., from 2 to 8 for target 5). On each trial, a prime was presented for 57 msec, preceded and followed by a mask of 57 msec. Immediately after the postmask, the target was presented until response. Ten persons participated in this condition. RTs aggregated over targets are plotted in Figure 4A (suggesting symmetric priming), and RTs aggregated over primes are plotted in Figure 4B (suggesting absence of size effect). The results are similar for individual targets, as the statistical analysis below will illustrate.

Symmetric priming. Figure 4A (full line) shows the mean RTs as a function of prime–target distance. Apart from a large identity effect, a reliable distance-related priming effect can be observed [difference between prime–target distance 0 and 1, $F(1,9) = 36.807$, $p < .001$; difference between distance 1 and 2, $F(1,9) = 15.662$, $p = .003$; difference between distance 2 and 3, $F(1,9) = 6.000$, $p = .037$].

To investigate the symmetry (V shape) of the curve in Figure 4A statistically, we performed a linear regression with RT as a dependent variable and predictor $X = -1$ for primes smaller than the target and $X = +1$ for primes larger than the target (Lorch & Myers, 1990). Targets were not aggregated in this analysis. A significant effect of X indicates that the smaller-than-target priming curves have a different intercept than the larger-than-target priming curves. A second predictor Y coded the distance between the prime and the target (ranging from -3 to $+3$). This predictor tests the difference in slope between primes smaller and larger than the targets. If there is no effect of either X or Y, this indicates that the two (optimal) regression lines through the observed data (left and right of the target) are equal. It turned out that the effects obtained were very small [$F(1,9) < 1$ for the intercept variable X, and $F(1,9) = 1.227$, $p = .297$, for the slope variable Y], which confirms symmetry. As was noted in the
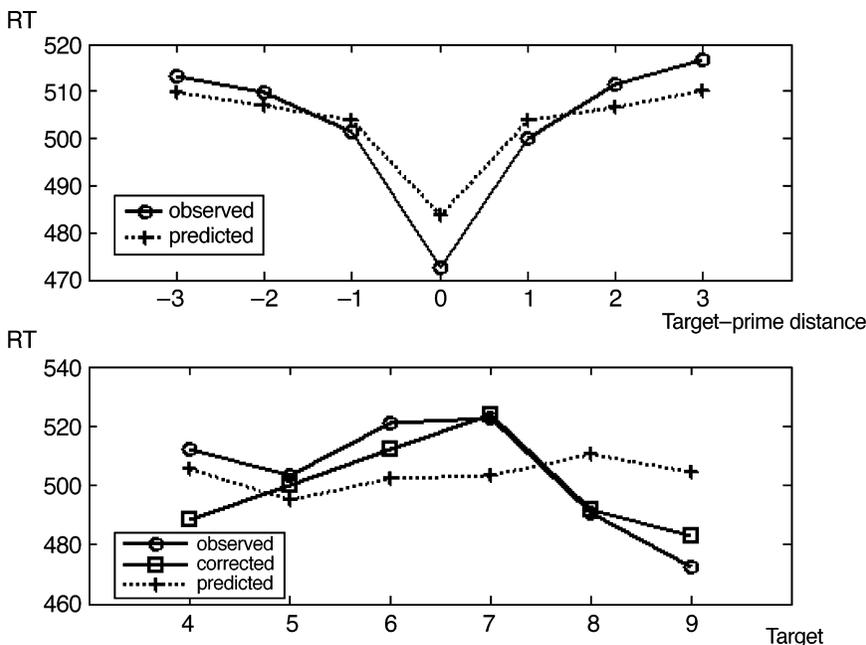

Figure 4. Number naming. (A) Response times (RTs) for fixed prime–target differences. (B) Mean RTs corrected for voice-onset time, and predicted RTs.

introduction, magnitude coding, compressed scaling, and increasing variability lead to difficulties in accounting for this finding.

Absence of a size effect. A second relevant feature of the naming data is that RTs for larger targets are not longer. Figure 4B (full line) shows mean RTs for each target separately. This figure suggests that if anything, RTs are *shorter* for larger targets. The decreasing (linear) trend is significant $[F(1,9) = 14.129, p = .004]$ and is due to the shorter RTs to target 9 (without target 9, $F < 1$). This again raises problems for the three assumptions, because all three entail that the size effect in comparison originates from less discriminability for larger numbers on the mental number line, and this should occur for any task that accesses it.

Delayed-naming experiment. One concern with the argument that RTs do not get longer over targets is that there could be a measurement artifact due to differential sensitivity of the voice key to different phonological onsets for the different number words. To correct for such effects, we performed a separate *delayed-naming* experiment. RTs from this experiment were then used to correct the Reynvoet, Brysbaert, and Fias's (2002) data for possible voice key artifacts, and the analyses were performed anew on the corrected RTs from Reynvoet, Brysbaert, and Fias.

In the delayed-naming experiment, a fixation cross was first presented for 1 sec, followed by a number, also presented for 1 sec. Then a blank appeared for a period of $200 + T$ msec, where T was a random number from an exponential distribution with a mean of 1,000 msec. In this way, anticipatory responses were avoided, because the end period of the blank was unpredictable. Then a question mark appeared, at which point the number had to be named. In this way, all the number line effects were presumably "faded out." For example, Gallistel and Gelman (1992) assumed that the standard deviation of noise in the mapping process decreases over time (Gallistel & Gelman, 1992, p. 57). If one waits long enough, the mapping should have become noise free, and all the differences between numbers should have largely disappeared. Seven research assistants participated, all naive as to the purpose of the experiment. Each participant received targets 1 to 9, each target presented 40 times in random order. The experiment took about 35 min.

No naming errors were made; 5.1% of the data were discarded due to voice key failures. A linear regression was performed with delayed-naming RT (from this experiment) as an independent variable and naming RT (from Reynvoet, Brysbaert, & Fias, 2002) as a dependent variable; the residual RT from this regression was then used as the dependent variable in the same analyses as those performed previously on the uncorrected RTs.

The corrected RT data were given the same mean as the original RT data, and the resulting RTs are plotted in Figure 4B. As can be seen, there is also no size effect in the corrected data. Statistically also, the results were similar. With respect to prime–target distances (analo-

gous to the results in Figure 4A), the distances 0 versus 1, 1 versus 2, and 2 versus 3 differed significantly from each other $[0$ vs.1, $F(1,9) = 36.809, p < .001; 1$ vs. 2, $F(1,9) = 10.443, p = .010; 2$ vs. 3, $F(1,9) = 5.982, p = .037]$. Primes smaller and larger than the target did not lead to a different absolute level of RT (variable X, discussed above; $F < 1$), and the absolute slope values of the priming curve for primes smaller and larger than the target were not significantly different [variable Y, discussed above; $F(1,9) = 1.340, p = .277$]. Most important, with respect to Figure 4B, the linear trend was not significant $(F < 1)$, indicating that RTs were not larger for larger numbers. This rules out the explanation that phonological artifacts obscure relevant effects—in particular, the size effect—in number naming.

Model Performance

The model was tested in primed number naming, using the connection weights from each of the 10 simulation rounds of the learning phase. For each set of connection weights, each prime–target pair was presented once. Each prime was presented from $t = 0$ to 0.3 (expressed in arbitrary model time units); then nothing was presented for another time interval of size 0.3 (so the prime and the postmask had the same length, as in the procedure of Reynvoet, Brysbaert, & Fias, 2002). To model the disruptive effect of the postmask, activation in the input field was decreased by a factor of 12 at the end of the prime interval. Because RTs were about 3 time units, prime duration was about 10% of the total simulated RT, in line with the data from these authors. After the postmask, the target was presented, and RT and correctness of the response were recorded.

The model achieved 100% accuracy. We performed a linear regression from the simulated RTs on the observed RTs (six targets with seven distances = 42 data points). The intercept may be considered a motor constant in which all peripheral processing is aggregated. The slope scales the arbitrary model time units with the observed RTs.

Simulated RTs for each prime–target distance are shown in Figure 4A (dotted line). RT profiles are similar for individual targets. As can be seen, the symmetry holds in the model, even though the sampling frequencies in the learning phase were from an exponential distribution and, hence, asymmetric over numbers. The correlation between the observed and the predicted data points in Figure 4A was .99 ($p < .01$).

Figure 4B (dotted line) shows the mean RTs for each target separately. There is no indication of a size effect: The differences that occur are random deviations due to the inherent stochasticity in the learning algorithm.

The priming effect is obtained in the model because a prime is presented before a target and this prime already generates some activation in the network. The priming effect is distance related because primes closer to the target lead to more activation in the target number field units (e.g., prime 3 leads to more activation in number line unit 4 than prime does 2). The priming effect is sym-

metric because the activation curves are symmetric (see Equation 2; Figure 2).

There is no size effect because each number line unit maps to its output (naming) unit with approximately equal strength (see Figure 3C). However, Figure 3B shows that this absence of a size effect only holds asymptotically: With a small number of training trials, a size effect is obtained. Hence, we make the prediction that there should be a naming size effect in children (i.e., in participants with fewer learning trials). This prediction remains to be tested.

To conclude, the model naturally captures both trends (symmetric priming and absence of a size effect). Note that the connection weights were not adjusted to fit the data, because the training procedure only tries to make the model generate the correct response. In the next section, we will analyze primed parity judgment data with respect to the same aspects and, again, will see whether the model accounts for the findings.

PARITY JUDGMENT

Behavioral Data: Additional Analyses of Reynvoet, Caessens, and Brysbaert (2002)

Here, we will use the data in Reynvoet, Caessens, and Brysbaert (2002, Experiment 1). Targets ranged from 5 to 10, and primes ranged from target -4 to target $+4$. The participants were required to judge the parity of a target number as quickly and accurately as possible by pressing one of two buttons. Otherwise, the design was similar to that for the number-naming experiment described in the previous section.

Symmetric priming. Figure 5A (full line) shows mean RTs aggregated over targets, as a function of prime–target distance. Primes that were incongruent with the targets (e.g., prime 2, target 3) resulted in longer RTs than did the congruent pairs (e.g., prime 1, target 3) and are not plotted.[1] Again, the distance-related priming is significant [distance 0 vs. 2, $F(1,15) = 109.082$; distance 2 vs. 4, $F(1,15) = 16.061$; both $ps < .01$]. The priming curves are again symmetrical [slope difference Y, $F(1,15) = 1.388$, $p = .257$; intercept difference X, $F(1,15) < 1$].

Absence of size effect. Figure 5B (full line) shows mean RTs for each target separately. The effect of target on RT was reliable [$F(1,15) = 16.191$, $p = .001$], but in the direction opposite to that predicted by a size effect (larger targets lead to shorter RTs). This was again due mainly to the largest targets [without targets 9 and 10, $F(1,15) = 3.75$, $p > .05$].

Model Performance

The same sequence of steps (prime, postmask, and target) was presented to the model as before, in line with the procedure followed by Reynvoet, Caessens, and Brysbaert (2002). Accuracy of the model was 100%. Again, we performed a linear regression from the simulated RTs on the observed RTs. For this purpose, only congruent prime–target pairs were used, making six targets and five distances, or 30 data points.

Predicted RTs appear in Figures 5A and 5B (dotted lines). In Figure 5A, symmetric priming is again apparent. The correlation between the data and the predictions in Figure 5A is .97 ($p < .01$). Figure 5B shows that the model exhibits no size effect in parity judgment, in line
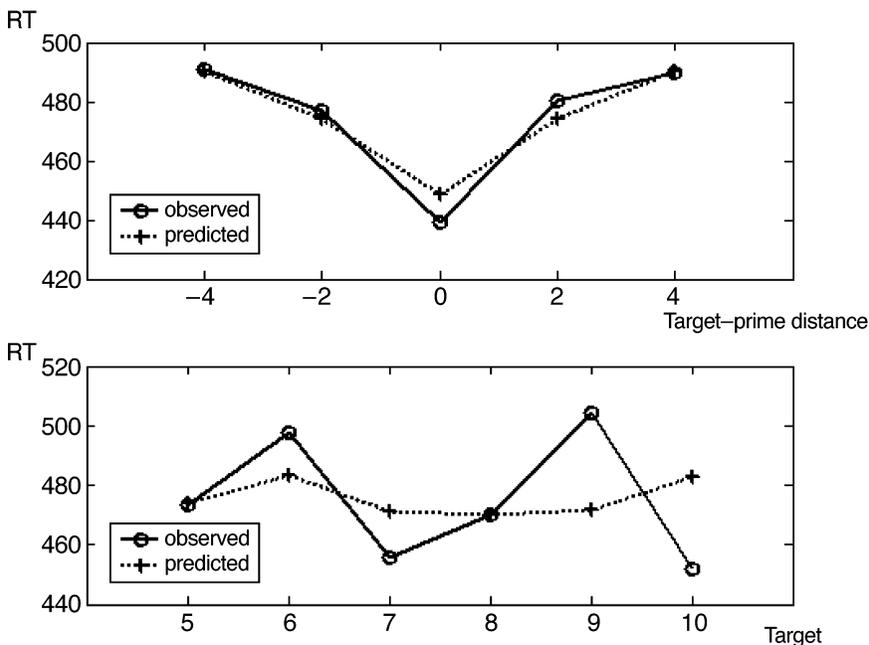

Figure 5. Parity judgment. (A) Response times (RTs) for fixed prime–target differences. (B) RTs for fixed target.

with the data, and despite asymmetric sampling of numbers in the learning phase. Hence, the same trends are observed as in number naming.

As with number naming, there is symmetric, distance-related priming, because the activation curves on the number line are symmetric, and there is no size effect, because each number line unit maps to its output unit with approximately equal strength (Figure 3F). However, just as for number naming, we do predict a size effect for parity judgments in children (see Figure 3E). This prediction also is not yet tested. In the next section, we will check whether the model can account for the distance and size effects in number comparison.

NUMBER COMPARISON

In the most common number comparison design, two numbers are presented simultaneously on each trial, and participants have to select the larger (or smaller) of the two. As elaborated by Schwarz and Stein (1998), the problem with this design is that it is difficult to distinguish size effects from bias. Bias occurs if a response is already prepared after processing only one of the two numbers. For example, if the numbers from 1 to 9 are used, the left number is 2, and only this number is processed, the right-larger response may already be partially prepared, because the probability that the other (right) number is larger (than 2) is high. In their Experiment 2, Schwarz and Stein used a design in which such bias ef-

fects were effectively removed by presenting the two numbers sequentially and making sure that the probability of the second number's being larger than the first was .5. The authors reported size and distance effects with this design. Interestingly, they found that, if simultaneous trials were intermixed with *bias-free* sequential trials, the bias effect also disappeared in the simultaneous trials.

In this article, we are not interested in such bias effects, and we wish to investigate the distance and size effects without ambiguity with respect to bias. Hence, to evaluate how the model fares in generating the size effect, the data for the simultaneous trials in Schwarz and Stein's (1988) Experiment 2 (reported in their Figure 8), were used. These data are plotted in Figure 6 (full lines). From this plot, the number pairs corresponding to a data point in Figure 6 can be unambiguously determined: For example, the second data point on the dist = 1 curve is the mean RT for the number pair 4 and 5 (aggregated over both orders). The distance effect is apparent from the fact that RTs are longer for a smaller distance between the members of a pair (and a fixed size, as indexed by the smaller number). The size effect is apparent from the fact that RTs are longer for a larger size (and a fixed distance; for statistical details, see Schwarz & Stein, 1998).

Model Performance

The two relevant output units were defined as *left larger* and *right larger*, in correspondence to the task in-
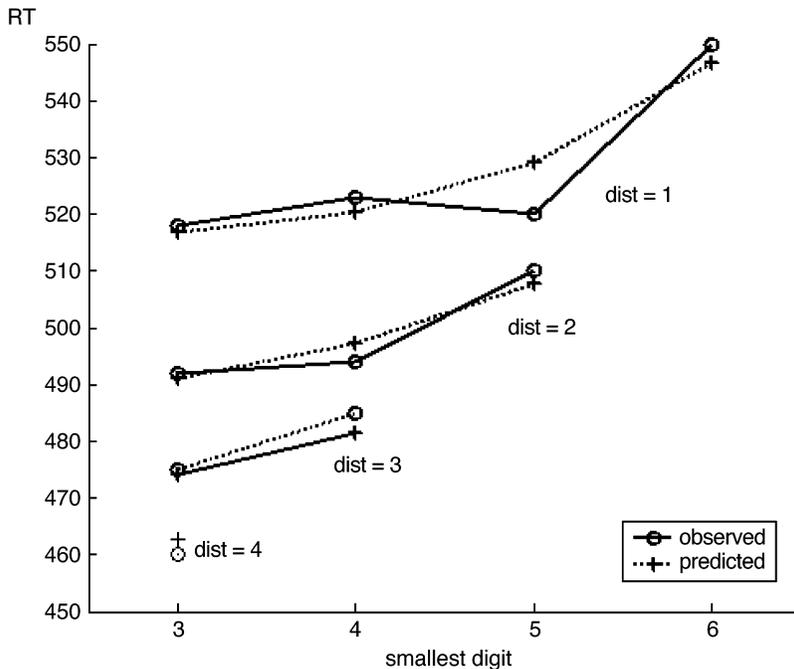

Figure 6. Number comparison: Response times (RTs) for different number pairs as a function of the smallest digit. A curve is depicted for each of the distances from 1 to 4 between digits.

structions. After learning, accuracy (over the 10 simulation rounds) was 98.5%. In one simulation round, an error was made on the pair 6 and 7, and in another simulation round, an error was made on the pair 5 and 6 and the pair 7 and 6. This was due to stochasticity in the learning rule, which does not guarantee a correct response after the learning phase.

As before, a linear regression was performed to put the model and the data RTs on the same scale. The predicted RTs are depicted with dotted lines in Figure 6. It is clear that both the distance and the size effects are present in the model simulations; the correlation between the observed and the model data is .99 ($p < .01$).

The model generates a correct response because the weight pattern is monotonic (either increasing or decreasing). For example, the weight from unit 7 on the left-hand number line to the left-larger response unit is larger than the same weight coming from unit 6, which is larger than the weight coming from unit 5, and so on (see Figure 3I). This pattern is required to generate a correct response for each pair of numbers, and, hence, the error correction learning algorithm generates such a pattern. A consequence of this monotonicity is that the distance effect is obtained. For example, suppose 3 is presented left and 7 right: Both number lines will send strong activation to the right-larger response. However, if 3 is presented left and 8 right, this tendency will be even stronger (right-hand number line unit 8 has a stronger connection to the right-larger response than number line unit 7 does), and hence, RT will be shorter with the number pair 3 and 8. Hence, as long as the model responds correctly, a distance effect will be obtained.

The size effect is obtained because of the *compressive* monotonic pattern of weights depicted in Figure 3I. Why do compressive weights lead to a size effect? To see this, let us suppose the pattern of weights from the left number line to the right-larger response is described by the decreasing and compressive exponential function e^{-i}; from the right number line to the right-larger response, weights are described by the increasing and compressive function $C - e^{-i}$. Suppose, then, that the numbers 1 and 2 are shown left and right, respectively; for simplicity, we will ignore the *spread* on the number line and will assume that activation of the corresponding number line units (1 and 2 for left and right, respectively) is 1. The input to the *right-larger* response unit is then $e^{-1} + (C - e^{-2})$. On the other hand, if the numbers 7 and 8 are shown left and right respectively, the input to right larger is $e^{-7} + (C - e^{-8})$. Due to compression, $e^{-1} - e^{-2} > e^{-7} - e^{-8}$, and so $C + (e^{-1} - e^{-2}) > C + (e^{-7} - e^{-8})$, and the RT to the pair 1 and 2 will be faster, leading to a size effect. It is clear that the argument is easily extended to an arbitrary compressive function $f(i)$. In contrast to what was the case in number naming or parity judgment, this is predicted to be an asymptotic effect. As long as numbers are presented with decreasing frequencies (as reported in Dehaene & Mehler, 1992, and used in our simulations), a size effect should remain.

GENERAL DISCUSSION

To account for the size effect in number comparison, various assumptions have been proposed in the literature, the most popular being magnitude coding, compressed scaling, and increasing variability on the mental number line. We have outlined difficulties with these assumptions when they are applied to other tasks, such as naming and parity judgment. Of primary concern are the symmetric pattern of distance-related priming and the absence of a size effect in these tasks.

Next, we developed a new model of numerical cognition that does not rely on these principles. Place coding, linear scaling, and fixed variability are the key features of this model. This leads to a model of exact small-number representation, in the sense that distances between numbers are preserved after mapping onto the mental number line. For example, the represented distance between the integers 1 and 2 is equal to the distance between 5 and 6. This model was applied to number naming, parity judgment, and number comparison. The weights from the number line to different output components were obtained by a training procedure. Note that the model and its training procedure have no knowledge of the observed RTs; the weights are chosen with the sole goal of making the model perform as accurately as possible. Nevertheless, the model successfully accounts for the distance and size effects in number comparison and, at the same time, accommodates the symmetric pattern of distance-related priming and semantic activation without a size effect in number naming and parity judgment.

A number of relevant issues have not been dealt with yet. First, we will discuss whether and how the model can account for error data. Second, the model can be extended to nonsymbolic input formats (Verguts & Fias, 2004), and this extension will be briefly reviewed. And finally, we will discuss how large approximate and large exact numbers would be treated in the present framework.

Accounting for Error Data

Size and distance effects are often observed not only in RTs, but also in error data. For example, in numerical comparison, errors decrease for larger distances and smaller numbers (distance and size effect, respectively). Although the absolute number of errors is generally low, these two effects in numerical comparison are established reliably (e.g., Link, 1990; Moyer & Landauer, 1967). If the model is extended with a stochastic component, it can handle these error patterns as well. Recall that in number comparison, the left-larger response is chosen if x_{Left} reaches a fixed threshold earlier than x_{Right}; if x_{Right} reaches the threshold first, the right-larger response is chosen. If the left number is larger than the right (e.g., 7 and 6, respectively), in the model it was almost always the case that x_{Left} was larger than x_{Right}, so x_{Left} reached the threshold first. However, for targets with smaller distance, the difference between activations ($x_{\text{Left}} - x_{\text{Right}}$) was smaller. In the present deterministic sit-

uation, this difference is immaterial. However, if one adds a stochastic component to the equations, this difference will become important, and larger differences will lead to higher probabilities of success. As a rough approximation, we have assumed that probability of success is not 1 but, rather, $p = x_{Left}/(x_{Left} + x_{Right})$ in the case in which *left larger* was the correct response and $1 - p = x_{Right}/(x_{Left} + x_{Right})$ otherwise. Size and distance effects were obtained with this extension. This suggests that error data do not pose a significant challenge to the model.

Nonsymbolic Input Formats

Recent single-cell recording data provide direct evidence on some characteristics of the mental number line (Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2003). Nieder et al. trained monkeys to respond to visual displays on a numerical basis: Two displays were shown consecutively, and the monkey was required to answer whether the second contained a smaller or a larger number of objects than the first. The authors recorded in the prefrontal cortex and found neurons that were selective to a specific number (presented as collections of dots). These number-selective neurons showed a typical number line characteristic, in the sense that if a neuron was most sensitive to number x, it was also somewhat (but less) sensitive to $x - 1$ and $x + 1$, still less sensitive to $x - 2$ and $x + 2$, and so on. Hence, numbers with a small distance (e.g., 2 and 3) lead to more overlapping activation distributions over the set of nodes than do numbers with a larger distance (e.g., 2 and 5). Note that this implies a place-coding system. Also, the neurons were found to have increasing variability: Activation distributions were more broadly tuned (larger variance) for larger numbers. These two properties are consistent either with linear scaling in combination with increasing variability on the number line or with compressed scaling in combination with constant variability on the number line (in the follow-up article, Nieder and Miller found evidence in favor of compressed scaling). In either case, these findings at first seem incompatible with the model postulated here. A crucial factor, however, is the input format of the stimuli. Numbers were presented in Arabic (symbolic) format in the data modeled in the present article, whereas numbers were presented as nonsymbolic numerosities in the studies of Nieder and colleagues (in particular, as collections of dots). In a recent article (Verguts & Fias, 2004), we showed how a number line can develop spontaneously under unsupervised learning conditions when a neural network is presented numbers as numerosities (e.g., numbers as collections of dots, as in Nieder et al.'s, 2002, study). After the neural network was trained, the properties of the resultant number line were strikingly similar to those observed by Nieder et al. and by Nieder and Miller. Moreover, we showed that if this network then learned to represent the meaning of numbers in a symbolic format (e.g., verbal or Arabic), the same number line that reacted to nonsymbolic stimuli also came to act as a number line for the symbolic format. Hence, this number line was sensitive to both input formats, symbolic and nonsymbolic. But importantly, we also showed that this number line exhibited different properties, depending on which kind of input it was presented. For both input formats, it was the case that if a certain node preferred digit x, it also reacted more weakly to $x - 1$ and $x + 1$, still more weakly to $x - 2$ and $x + 2$, and so on. However, there was no increasing variability for the symbolic format, in contrast to the nonsymbolic format. Hence, with symbolic input, the number line developed exactly those properties (place coding and constant variability) that have been postulated in the present article. On the other hand, the representation of nonsymbolic quantities is characterized by place coding and increasing variability. Note that this implies that with nonsymbolic stimuli, size effects are expected in tasks other than comparison. This prediction has been confirmed by results obtained in subitizing (e.g., Trick & Pylyshyn, 1994), estimation of the numerosity of a set of stimuli (e.g., van Oeffelen & Vos, 1982), and the production of an estimated number of keypresses (e.g., Whalen, Gallistel, & Gelman, 1999). Moreover, in comparison tasks, we predict larger size effects with nonsymbolic stimuli than with symbolic stimuli, because nonsymbolic stimuli bear a double generator of a size effect (increasing variability and nonlinear mappings), whereas with symbolic stimuli there is only one source (nonlinear mappings). This prediction is supported by a study from Buckley and Gillman (1974), who found larger size effects when collections of dots were compared than when Arabic digits were compared.

Small and Large Numbers

Our model incorporates numbers up to 15, but obviously, humans can work with numbers beyond that number. The most straightforward generalization of the model would be simply to extend the number line toward larger numbers. With a training regime in which small numbers are more frequent than large numbers, the resultant number line would be one in which small numbers are represented exactly but representations become fuzzier for larger numbers. A first reason why representations would become fuzzier for larger numbers is that frequency influences the mappings from number line to output. This is clear in Figure 3, which shows that weights of less frequent numbers are adapted more slowly (although all reach the same asymptote). If even less frequent numbers (beyond 15) had been used, the weights for those numbers would not yet have reached asymptote at the end of training (30,000 trials) and, hence, they would have been represented less accurately than smaller numbers. A second reason why large number representations would be fuzzier is that fewer number line units would be dedicated to larger numbers because of their low frequency (Verguts & Fias, 2004). Hence, one number line would represent both small and large numbers—small numbers in an exact manner, due to their high frequency, and large numbers increasingly fuzzier. Hence, the properties of compressed scaling and increasing variability may well hold for large numbers.

However, this one-number-line account cannot be the complete story of the processing of large numbers, for two reasons. First, a large number such as 4,653 may invoke the impression of a large number with an approximate value; but if one is to compare this number with 4,654, it makes little sense to use fuzzy number representations (Carey, 2001). Second, neuroscientific evidence suggests that the brain circuits involved in the processing of small numbers may (partly) differ from those involved in the processing of large numbers. For instance, disruption of the processing of small and of large numbers after transcranial magnetic stimulation occurs after stimulation of different brain areas (Göbel, Rushworth, & Walsh, 2001; Göbel, Walsh, & Rushworth, 2001). Furthermore, using brain imaging during the comparison of two-digit numbers and nonsymbolic magnitudes (line lengths and angle sizes), Fias, Lammertyn, Reynvoet, Dupont, and Orban (2003) observed common activation for quantitative comparison, irrespective of type of input. This was found in parietal regions that were more posterior than the horizontal segment of the intraparietal sulcus; the latter has been taken to be a crucial brain region for the processing of number magnitude, on the basis of studies predominantly involving single-digit numbers (Dehaene, Piazza, Pinel, & Cohen, 2003). Using event-related potentials measured during number comparison, Whalen and Morelli (2002) came to the same conclusion of different processing circuits for small and large numbers.

In light of this evidence, we propose that there might actually be two numerical systems; one corresponds to the system described in this article and is used for exact small and approximate large numbers. The other system represents multidigit numbers. It decomposes multidigit numbers in its base-10 representation on different number lines, so that one number line represents the unit value, one number line represents the decade value, and so on. Recent data support such a decomposition for multidigit numbers (Nürk, Weger, & Willmes, 2001; Ratinckx, Brysbaert, & Fias, 2005). The first system accepts both symbolic and nonsymbolic input, the second only symbolic input. The decomposition property of the exact system links numerical representations and language (see Spelke & Tsivkin, 2001, for a related two-system proposal). The explicit predictions laid out in the present article may provide a starting point for charting in detail the plausibility of such multiple-system views, the characteristics of and differences between the systems, how they cooperate, and how they interact.

REFERENCES

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

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, **124**, 434-452.

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

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *Quarterly Journal of Experimental Psychology*, **54A**, 1005-1029.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind & Language*, **16**, 37-55.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, **124**, 375-390.

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

Dehaene, S. (2004). The neural bases of subliminal priming. In N. Kanwisher & J. Duncan (Eds.), *Functional neuroimaging of visual cognition*: *Attention and performance XX* (pp. 205-224). Oxford: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, **122**, 371-396.

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 & Performance*, **16**, 626-641.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, **43**, 1-29.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, **20**, 487-506.

Festinger, L. (1943). Studies in decision: I. Decision time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, **32**, 291-306.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research*, **65**, 250-259.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, **2**, 95-110.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience*, **15**, 47-56.

Fias, W., Reynvoet, B., & Brysbaert, M. (2001). Are Arabic numerals processed as pictures in a Stroop interference task? *Psychological Research*, **65**, 242-249.

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

Göbel, S., Rushworth, M. F. S., & Walsh, V. (2001). rTMS disrupts the representation of small numbers in supramarginal gyrus [Abstract]. *NeuroImage*, **13**(6, Suppl. 1), 409.

Göbel, S., Walsh, V., & Rushworth, M. F. S. (2001). The mental number line and the human angular gyrus. *NeuroImage*, **14**, 1278-1289.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1882-1905.

Link, S. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, **34**, 2-41.

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 149-157.

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

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

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, **37**, 149-157.

Nürk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, **82**, B25-B33.

Ratinckx, E., Brysbaert, M., & Fias, W. (2005). *Naming two-digit*

Arabic numerals: Evidence from masked priming studies. Manuscript submitted for publication.

REYNVOET, B., & BRYSBAERT, M. (1999). Single-digit and two-digit numerals address the same semantic number line. *Cognition*, **72**, 191-201.

REYNVOET, B., & BRYSBAERT, M. (2004). Cross-notation number priming investigated at different stimulus onset asynchronies in parity and naming tasks. *Experimental Psychology*, **51**, 81-90.

REYNVOET, B., BRYSBAERT, M., & FIAS, W. (2002). Semantic priming in number naming. *Quarterly Journal of Experimental Psychology*, **55A**, 1127-1139.

REYNVOET, B., CAESSENS, B., & BRYSBAERT, M. (2002). Automatic stimulus–response associations may be semantically mediated. *Psychonomic Bulletin & Review*, **9**, 107-112.

SCHWARZ, W., & STEIN, F. (1998). On the temporal dynamics of digit comparison processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1275-1293.

SEKULER, R., RUBIN, E., & ARMSTRONG, R. (1971). Processing numerical information: A choice time analysis. *Journal of Experimental Psychology*, **90**, 75-80.

SPELKE, E. S., & TSIVKIN, S. (2001). Language and number: A bilingual training study. *Cognition*, **78**, 45-88.

TRICK, L. M., & PYLYSHYN, Z.W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, **101**, 80-102.

VAN OEFFELEN, M. P., & VOS, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, **32**, 163-170.

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

WHALEN, J., GALLISTEL, C. R., & GELMAN, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, **10**, 130-137.

WHALEN, J., & MORELLI, F. (2002, November). *Cognitive neuroscience of number*. Poster presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

WIDROW, B., & LEHR, M. A. (1990). 30 years of adaptive neural networks: Perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, **78**, 1415-1442.

ZORZI, M., & BUTTERWORTH, B. (1999). A computational model of number comparison. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 778-783). Mahwah, NJ: Erlbaum.

NOTE

1. In parity judgment, the model predicts that incongruent prime–target pairs are as slow as congruent prime–target pairs with large distance (e.g., 3 and 4, and 3 and 7). However, longer RTs for incongruent pairs can easily be accommodated in the model by incorporating lateral inhibition at the response level. Since this would add complexity to the model and is of no other use in the simulations described here, we chose not to incorporate lateral inhibition.