

Dissociation of the distance effect and size effect in one-digit numbers

TOM VERGUTS and FILIP VAN OPSTAL
Ghent University, Ghent, Belgium

Magnitude comparison of single digits is robustly characterized by a **distance effect** (close numbers are more difficult to compare than numbers further apart) and a **size effect** (for a given distance, comparison difficulty increases with increasing size). The distance effect indicates access to the mental number line (Dehaene, 1997), and the size effect is usually interpreted as indicating that the mental number line represents larger numbers more vaguely than smaller ones. In contrast, we have argued earlier (Verguts, Fias, & Stevens, 2005) that for symbolic numbers (Arabic or verbal notation), the size effect does not originate from the mental number line but, instead, originates from mappings to relevant output components that are specific for magnitude comparison. If the latter is true, it should be possible to dissociate the distance effect from the size effect in tasks other than magnitude comparison. In two experiments, we observed a robust distance effect in *same/different* judgments, which implies access to the mental number line. Yet the size effect was absent. Consistent with our prediction, this finding establishes a dissociation between the size effect and the distance effect.

Numbers in Arabic and verbal notations are highly abstract: They bear no relation to their magnitudes whatsoever. Yet comparing two such numbers is subject to a distance effect, meaning that it is more difficult to compare two numbers if they are close (e.g., 1 and 2) than if they are far apart (e.g., 1 and 8). This is obtained both when the numerical magnitude is relevant (e.g., Moyer & Landauer, 1967; Schwarz & Stein, 1998) and when it is irrelevant (Dehaene & Akhavein, 1995). A distance effect is also robustly obtained in priming tasks, in the sense that smaller prime–target distances lead to shorter response times (RTs; e.g., Reynvoet & Brysbaert, 1999). The distance effect suggests that a number's magnitude is accessed immediately and that the number is placed on a mental number line (Dehaene, 1997), after which further processing can proceed.

In addition to a distance effect, magnitude comparison (which number is smaller/larger?) is subject to a size effect: Comparison of two numbers is easier for small than for large numbers (e.g., 1 and 2 vs. 8 and 9). The latter finding has led some authors to posit that the number line represents large numbers more vaguely than small numbers, so that discriminating between larger numbers is more difficult. Different implementations of this general idea have been proposed (e.g., logarithmic compression, Dehaene, 1992; scalar variability, Gallistel & Gel-

man, 1992; see also recent discussions in Carey, 2001, and Dehaene, 2001). Arithmetical operations (e.g., addition or multiplication) typically also exhibit a size effect, but it is controversial whether this size effect has the same origin as that observed in number comparison (Dehaene, Piazza, Pinel, & Cohen, 2003; Gallistel & Gelman, 1992). In the following, we will therefore ignore size effects in arithmetic.

We (Verguts, Fias, & Stevens, 2005) have recently argued that, at least in the range of one-digit numbers, large numbers (e.g., 8 or 9) are represented as exactly (or as vaguely) as small numbers (e.g., 1 or 2). One argument was that with symbolic numbers¹—that is, numbers in Arabic or verbal notation—the size effect appears in magnitude comparison, but not in number naming or parity judgment. A second argument was that, in masked priming studies with number naming and parity judgment (e.g., Reynvoet & Brysbaert, 1999; Reynvoet, Caessens, & Brysbaert, 2002), there is a distance effect between the prime and the target, but no size effect (of either the prime or the target). We argued that the size effect in magnitude comparison originates from mappings (connections) from the number line to the task-relevant output component.

To substantiate this, we trained a neural network on magnitude comparison, number naming, and parity judgment. The input part of the model consisted of either a single number line (in number naming and parity judgment) or two number lines (in magnitude comparison). The output component for each task consisted of a number of nodes, one for each possible response: In number naming, the relevant response nodes corresponded to the (spoken) responses “one,” “two,” and so on; in parity judgment, there were two response nodes, corresponding to the responses “odd” and “even”; and in magnitude comparison, there were also two response nodes, corresponding to “left

Our work was supported by Grant G.0188.04 from the Fund of Scientific Research–Flanders. The second author's work was also supported by Grant 12053101 of the Special Research Fund of Ghent University to André Vandierendonck and Wim Fias. Correspondence concerning this article should be addressed to T. Verguts, Department of Experimental Psychology, Ghent University, H. Dunantlaan 2, 9000 Ghent, Belgium (e-mail: tom.verguts@ugent.be).

number larger” and “right number larger.” Crucially, the model was trained using the number frequencies observed in daily life, as reported by Dehaene and Mehler (1992); that is, small numbers were presented more frequently than large numbers.

After training the model, the connections from the number line to the naming output component were such that each location on the number line was connected only to its corresponding response: The location for the number 1 was connected to the response “one,” the location for the number 2 to the response “two,” and so on. Importantly, the connection weights were approximately equal for all the numbers, the result being that no size effect was obtained. Similarly, for parity judgment, each number line location was connected to its corresponding response (e.g., locations 1, 3, 5, . . . to the response “odd”). The connection weights were about the same across all numbers, such that, again, no size effect was obtained. In magnitude comparison, connections after training were such that, for example, larger numbers on the left-hand number line had a larger connection to the response “left larger.” Importantly, however, the model had adapted the pattern of weights across numbers in a *compressive* manner, in the sense that the difference between weights was smaller for larger numbers. For example, the difference between weights originating from the locations for the numbers 8 and 9 was smaller than the difference between weights coming from the locations for the numbers 1 and 2. Because of this compressive pattern, a size effect was obtained. To summarize, although all three tasks (naming, parity judgment, and magnitude comparison) were trained with a skewed frequency distribution, the size effect emerged only in the mapping from the number line(s) to the comparison-relevant output module.

If the size effect does not originate from the mental number line but, rather, from task-specific connections to an output component, it follows that the size effect should be dissociable from the distance effect. A stringent test of this hypothesis would be obtained with a task in which there are two task-relevant numbers (as in magnitude comparison, but unlike number naming or parity judgment), so that a distance effect can be calculated between the two (task-relevant) numbers. If there is a distance effect, this implies access to the mental number line, so if the distance and the size effects originate from the same source (e.g., the number line), a size effect should be obtained as well. On the contrary, if our argument is correct, the size effect should not be obtained in tasks other than magnitude comparison (at least with one-digit numbers). A task that is ideally suited for these purposes is the *same/different* judgment, in which people judge whether two numbers are (numerically) equal. Contrary to magnitude comparison, which is a very frequent task in daily life, it is extremely unlikely that people have connections to the task-relevant responses (“same” and “different”) available for *same/different* judgments. Hence, some general-purpose mechanism must be used to read off the informa-

tion on the mental number line. Because, by hypothesis, this mechanism is not specific for numerical cognition, the skewed frequency argument outlined in the previous paragraph cannot apply here, and it follows that there should be no size effect in *same/different* judgments. This was tested in the present study.

Besides testing these two general views, the present study is also informative with respect to yet a third possibility as to the origin of the size effect, which is that the size effect could obtain in any task in which a comparison has to be made between two numbers. Indeed, given that a size effect appears in magnitude comparison, but not in number naming or in parity judgment, this may be the case, because magnitude comparison requires an explicit comparison between two numbers, whereas naming and parity judgment do not. According to this hypothesis, then, a size effect should be obtained in *same/different* judgments as well.

Dehaene and Akhavein (1995) already have used a *same/different* judgment task in which they mixed Arabic (A) and verbal (V) notations (conditions A–A, V–A, A–V, and V–V; e.g., “5–eight” is a stimulus in the A–V condition). The authors used the numbers 1, 2, 8, and 9. A size effect was reported in the sense that, for a fixed distance, RTs with small numbers (pairs 1–1, 2–2, and 1–2) were shorter than those with larger numbers (pairs 8–8, 9–9, and 8–9). However, this size effect may have been artifactual, because with the numbers 1, 2, 8, and 9, physical similarity is confounded with size. To verify this, we used the similarity ratings for single digits collected by Campbell and Clark (1988). The Arabic numbers 1 and 2 were not physically similar to any of the three other digits (on a 7-point scale ranging from 1 to 7; mean rated similarity, 1.91; percentile, 28), whereas the Arabic numbers 8 and 9 were judged to be physically similar to each other (mean rated similarity, 5.75; percentile, 95). Hence, the size effect could be attributed to the physical similarity between 8 and 9. Physical similarity is most problematic in the A–A condition, but physical similarity may also play a role in the A–V and V–A conditions. For example, encoding the Arabic number 8 in a stimulus such as “8–eight” in the A–V condition will be more difficult than encoding the Arabic number 1 in the stimulus “1–one.” This is because the number 8 is physically similar to one of the other three Arabic digits that are repeatedly presented in the experiment (9), whereas the digit 1 is not.

Also, Dehaene and Akhavein (1995) did not distinguish *same* and *different* responses in calculating the size effect. However, for *same* responses a size effect is difficult to interpret, because the distance effect cannot be calculated (distance is always trivially zero in these trials). Indeed, the distance effect, either between two task-relevant numbers or between a prime and a target, is a universally agreed upon signature of access to the mental number line (e.g., Dehaene & Akhavein, 1995; Reynvoet et al., 2002), and if the intention is to interpret the origin of the size effect, such a signature is required.

In Experiment 1 of the present study, we replicated Dehaene and Akhavan's (1995) design (with the numbers 1, 2, 8, and 9) but used mixed pairs (A–V and V–A) only to avoid a physical similarity matching strategy as much as possible. In addition, we calculated the size effect for *same* and *different* trials separately. In Experiment 2, we replicated this design, but with the numbers 1, 2, 7, and 8, so that the physical similarity problem could be addressed more directly.

EXPERIMENT 1

Method

Stimuli. The numbers 1, 2, 8, and 9 were used. Three types of trials were distinguished: *same* (e.g., 1–1), *close* (e.g., 1–2), and *far* (e.g., 1–8). Following Dehaene and Akhavan (1995), we equalized the number of same, close, and far trials. Since there are more individual far than close stimuli (8 and 4, respectively, ignoring notation), this implies that each individual close stimulus was presented twice as often as each individual far stimulus. However, as was noted by Dehaene and Akhavan, this can only work against the distance effect, because responses on far trials are predicted to be faster. Each individual same and close stimulus was presented twice in each block, and each individual far stimulus once. Since there are two notation orders (A–V and V–A), this made for a total of $3 \times 8 \times 2 = 48$ stimuli in each block, for a total of eight blocks

per participant. The verbal stimuli were the Dutch words for 1, 2, 8, and 9: “een,” “twee,” “acht,” and “negen,” respectively.

Procedure. A fixation cross was presented on each trial. The fixation cross disappeared after 200 msec, followed by a stimulus consisting of two numerals (font, Arial; font size, 22), one Arabic and one verbal (all uppercase letters). One numeral was presented left and the other right of fixation. The distance between the two numerals was 1.5 cm. The stimuli were presented as white characters on a black background. The stimulus remained on the screen until a button was pressed. Responses were recorded with an external response box attached to the computer. The intertrial interval was 200 msec. Response mappings (left/right button for *same/different* response) were counterbalanced across participants.

Participants. Twelve psychology students from Ghent University (about 20 years old; three of them male) participated for course credit. None was familiar with the purpose of the experiment.

Results

Different trials. Incorrect trials (4.7%) and trials with RTs larger than 2,000 msec (1%) were removed from the RT analysis. For the purpose of checking the size effect, close trials were further split up into close-small (e.g., 1–2) and close-large (e.g., 8–9) trials. The right-hand part of Figure 1A shows the mean RTs for close-small, close-large, and far stimuli separately. There was an overall difference between the three conditions [$F(2,22) = 12.714$,

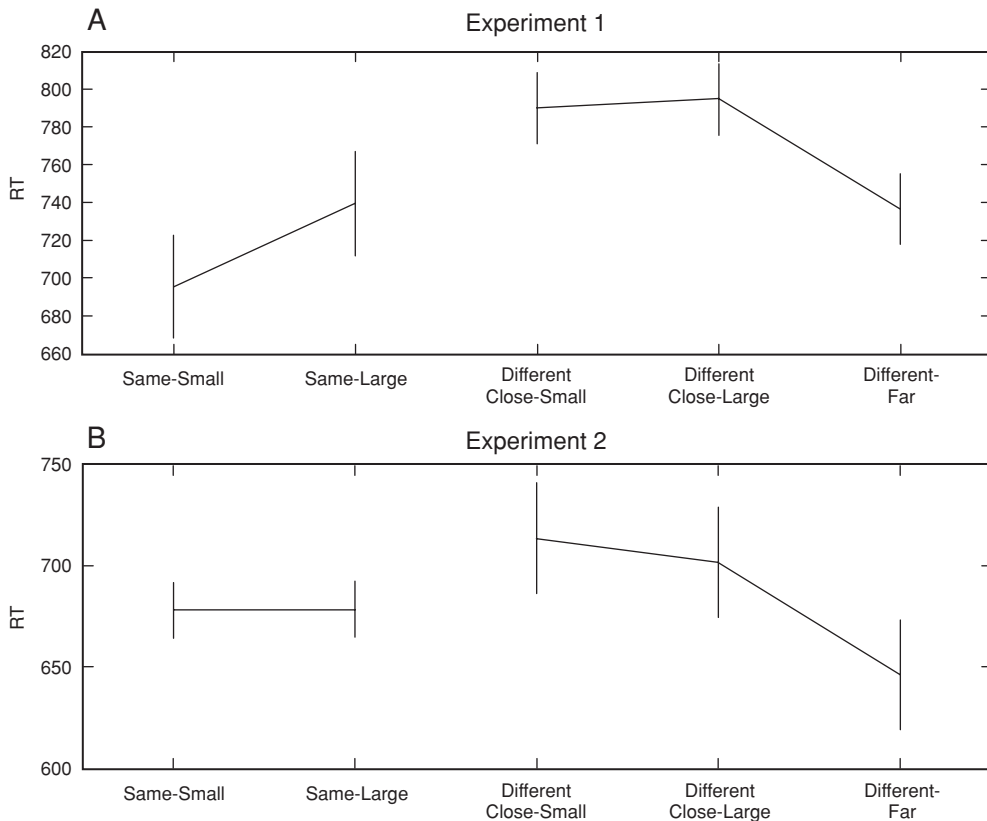


Figure 1. Response times (RTs) for same-small, same-large, different-close-small, different-close-large, and far trials. (A) Experiment 1: numbers 1, 2, 8, and 9. (B) Experiment 2: numbers 1, 2, 7, and 8. We constructed 95% confidence intervals according to the guidelines of Loftus and Masson (1994).

$MS_e = 992.085$, $p < .001$]. Paired t tests revealed that there was no size effect, because the RTs for close-small and close-large trials were approximately equal [close-small vs. close-large, $t(11) = -0.405$, $p = .693$]. On the other hand, there was a distance effect, because there was a difference between close-small and far [$t(11) = 3.701$, $p = .003$] and also between close-large and far [$t(11) = 4.848$, $p = .001$].

The accuracies paralleled the RT data. For close-small, close-large, and close-far trials, accuracies were 91%, 94%, and 96%, respectively. There was a significant difference between these three conditions [ANOVA on arcsine transformed proportions, $F(2,22) = 3.584$, $MS_e = 0.010$, $p = .045$]; paired t tests revealed that the close-small and the close-large conditions did not differ [$t(11) = -1.161$, $p = .270$], the close-small and the far trials did differ [$t(11) = -3.150$, $p = .009$], and the close-large and the far trials also did not differ [$t(11) = -1.357$, $p = .202$]. The fact that the accuracy data point in the same direction as the RT data indicates that there was no speed-accuracy trade-off at work.

Same trials. Incorrect trials (13.1%) and trials with RTs above 2,000 msec (1%) were removed from the RT analysis. The error rate is high, presumably because there were fewer *same* than *different* trials, in line with Dehaene and Akhavein (1995); our error rate was of a magnitude similar to that in Dehaene and Akhavein. The left-hand part of Figure 1A shows that there was a size effect in *same* trials [$F(1,11) = 6.373$, $MS_e = 1,838.532$, $p = .028$]. Accuracies did not differ significantly (*same*-small and *same*-large, 85% and 87%, respectively; $F < 1$).

Discussion

We partially replicated a study of Dehaene and Akhavein (1995). There was a distance effect in the *different* trials, which implies access to the mental number line. Given that the number line represents larger numbers more vaguely, it follows that a size effect should also be obtained in these trials. On the other hand, if the size effect originates from a mapping to output that is specific for the magnitude comparison task, there should be no size effect in the present task. The results were in favor of the latter hypothesis.

One remaining concern is the size effect in *same* trials. It is unclear how either hypothesis would straightforwardly predict a size effect in this condition: If two numbers overlap completely on the mental number line, it should not matter whether their activation curves are broad or narrow. We propose that this remaining size effect might have been due to physical similarity. We used only A-V and V-A trials, in which this confounding problem is minimized, but as was discussed above, the problem may not disappear completely. To test this explanation, we replicated our design in Experiment 2, but with the numbers 1, 2, 7, and 8. On the basis of Campbell and Clark's (1988) physical similarity ratings discussed

earlier, none of these four digits is particularly similar to one of the others (mean similarity, 2.53; percentile, 44). If the size effect in *same* trials is a matter of physical similarity between 8 and 9, the effect should disappear when the physically dissimilar numbers 1, 2, 7, and 8 are used. Of course, physical similarity can never be completely controlled, and physical similarity between number words (e.g., *een* [1] and *twee* [2]) was ignored in this analysis. However, by removing the digit 9, we removed the largest similarity among the digits and, we suspect, the largest similarity in the whole stimulus set.

EXPERIMENT 2

Method

The procedure and stimuli were the same as those in Experiment 1, except that the number 9 was replaced by 7. Eleven students (about 20 years old; six of them male) from Ghent University participated for course credit. Response mappings were counterbalanced across persons.

Results and Discussion

One person inadvertently used the other response mapping rather than the one assigned. He was therefore classified as using the other response mapping. Excluding this person from the analyses did not significantly change any of the results.

Different trials. Trials with RTs above 2,000 msec (1.6%) and incorrect trials (2.8%) were removed from the RT data. As can be seen in the right-hand part of Figure 1B, the *different* trial data are similar in pattern to the data from Experiment 1. The three conditions differed significantly [$F(2,20) = 8.782$, $MS_e = 1,643.520$, $p = .002$]. Paired t tests showed that close-small and close-large did not differ significantly [$t(10) = 0.606$, $p = .508$], close-small was different from far [$t(10) = 5.455$, $p < .001$], and close-large was also different from far [$t(10) = 2.954$, $p = .014$]. There was no significant difference in accuracy between conditions (93%, 98%, and 98% correct for close-small, close-large, and far, respectively), but the data are in the same direction as those in Experiment 1.

Same trials. Trials with RTs above 2,000 msec (0.4%) and incorrect trials (10%) were removed. The left-hand part of Figure 1B shows that with physically dissimilar large stimuli (7 and 8), there is no size effect in *same* trials either ($F < 1$). Accuracies also were not significantly different for the two conditions (88% and 91% for same-small and same-large, respectively).

These data replicate the distance effect found in Experiment 1 and the absence of a size effect in *different* trials. Moreover, the experiment shows that using physically dissimilar large numbers leads to a complete disappearance of the size effect in *same* trials as well. It is well known that shorter RTs lead to smaller effect sizes (e.g., Hale & Jansen, 1994), and the RTs in Experiment 2 were shorter. However, this cannot explain the difference in size effect between the two experiments: If RT speedup solely

is responsible for the smaller effect size, the size effect expressed as a proportion of mean RT should be the same for the two experiments (Hale & Jansen, 1994). Instead, the size effect was 6% of the mean RT in Experiment 1 and 0.06% in Experiment 2, or 100 times smaller.

As an additional test of the physical similarity interpretation for the size effect in *same* trials obtained in Experiment 1, we compared RTs in *same* trials for the three numbers that were used in both experiments (the numbers 1, 2, and 8). If our interpretation is correct that the size effect in Experiment 1 originated from the fact that both 8 and 9 were used, RTs to *same* trials containing the number 8 (i.e., “8–eight” and “eight–8”) should be longer in Experiment 1 than in Experiment 2. To take into account overall RT differences between experiments, the increase in RT for the number 8 in Experiment 1 (relative to Experiment 2) should be more than the extra RT for the numbers 1 and 2 in Experiment 1 (relative to Experiment 2). Hence, an interaction should be obtained between experiment (1 vs. 2) and number (1–2 vs. 8) in *same* trials. The between-subjects experiment factor is a factor of physical similarity: It indicates whether the number 8 is (physically) similar to one of the other numbers used in the experiment. This interaction can indeed be observed [$F(1,21) = 5.99, MS_e = 1,001.767, p = .023$]. *Same* trials containing the numbers 1 or 2 are 17 msec longer in Experiment 1 than in Experiment 2, but *same* trials containing the number 8 are 63 msec longer in Experiment 1 than in Experiment 2. The extra term of 46 msec ($= 63 - 17$) can be interpreted as the *physical similarity cost*. Interestingly, this extra term is almost the same as the size effect in Experiment 1, which was 44 msec. This finding corroborates our interpretation of the size effect obtained in Experiment 1. The analogous analysis on the accuracy data yielded no significant experiment \times size interaction ($F < 1$).

GENERAL DISCUSSION

We have contrasted two general, competing views on the structure of the mental number line. One key difference between the two views is whether the size effect in single-digit processing originates from the mental number line. There was a distance effect in our data, which, according to both views, implies access to the mental number line. Given a distance effect, it follows that if the size effect originates from the mental number line, a size effect should be obtained as well. This prediction was falsified, suggesting that the size effect originates elsewhere. One possible objection could be that numbers up to 8 (as in Experiment 2) are simply too small to generate a size effect. However, this cannot be true, because in magnitude comparison, a size effect is regularly obtained even for smaller numbers (e.g., Schwarz & Stein, 1998).

Besides these two hypotheses, the data also allow us to rule out the possibility that a size effect is obtained whenever two numbers have to be compared. Indeed, in

both magnitude comparison and *same/different* judgment, there are two numbers to be compared, but a size effect is regularly obtained for magnitude comparison, but not for *same/different* judgment, as the present data show. In contrast, the present findings are consistent with our earlier proposal (Verguts et al., 2005) that the size effect in magnitude comparison originates from task-specific mappings to output components, rather than from the mental number line. We emphasize that a similar representation of small and large numbers is assumed to hold only for single-digit numbers. Indeed, if our neural network model is trained on multidigit numbers also, it represents them more vaguely than single-digit numbers, for the simple reason that multidigit numbers are much less frequent in daily life (and hence, in the training set). Exact processing of multidigit numbers can occur, then, only by decomposing the numbers so that the (precise) representations of single digits can be used (Verguts & De Moor, 2005). In this sense, not only the symbols, but also the representations of single digits, serve as the building blocks for exact number processing.

REFERENCES

- CAMPBELL, J. I. D., & CLARK, J. M. (1988). An encoding-complex view of cognitive processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General*, **117**, 204–214.
- CAREY, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis [Review of the book *The number sense: How the mind creates mathematics*]. *Mind & Language*, **16**, 37–55.
- DEHAENE, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1–42.
- DEHAENE, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- DEHAENE, S. (2001). Author's response: Is number sense a patchwork? *Mind & Language*, **16**, 89–100.
- DEHAENE, S., & AKHAVEIN, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 314–326.
- 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.
- FIAS, W., & VERGUTS, T. (2004). The mental number line: Exact and approximate. *Trends in Cognitive Sciences*, **8**, 447–448.
- GALLISTEL, C. R., & GELMAN, R. (1992). Preverbal and verbal counting and computation. *Cognition*, **44**, 43–74.
- HALE, S., & JANSEN, J. (1994). Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychological Science*, **5**, 384–389.
- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, **1**, 476–490.
- MOYER, R. S., & LANDAUER, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, **215**, 1519–1520.
- REYNVOET, B., & BRYNBAERT, M. (1999). Single-digit and two-digit numerals address the same semantic number line. *Cognition*, **72**, 191–201.
- REYNVOET, B., CAESSENS, B., & BRYNBAERT, 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.

- VERGUTS, T., & DE MOOR, W. (2005). Two-digit comparison: Decomposed, holistic, or hybrid? *Experimental Psychology*, **52**, 195-200.
- VERGUTS, T., & FIAS, W. (2004). Numerical representations in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, **16**, 1493-1504.
- VERGUTS, T., FIAS, W., & STEVENS, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, **12**, 66-80.

NOTE

1. With numbers in nonsymbolic notation (e.g., a number of dots or a time interval), our framework also predicts a size effect in tasks other than magnitude comparison. Details on the differential predictions for symbolic and nonsymbolic notation are provided in Verguts and Fias (2004) and are summarized in Fias and Verguts (2004).

(Manuscript received May 26, 2004;
revision accepted for publication December 27, 2004.)