Interacting neighbors: A connectionist model of retrieval in single-digit multiplication

TOM VERGUTS and WIM FIAS
Ghent University, Ghent, Belgium

For most adults, retrieval is the most common way to solve a single-digit multiplication problem (Campbell & Xue, 2001). Many theories have been proposed to describe the underlying mechanism of arithmetical fact retrieval. Testing their validity hinges on evaluating how well they account for the basic findings in mental arithmetic. The most important findings are the problem size effect (small multiplication problems are easier than larger ones; cf. 3×2 and 7×8), the five effect (problems with 5 are easier than can be accounted for by their size), and the tie effect (problems with identical operands are easier than other problems; cf. 8×8 and 8×7). We show that all existing theories have difficulties in accounting for one or more of these phenomena. A new theory is presented that avoids these difficulties. The basic assumption is that candidate answers to a particular problem are in cooperative/competitive interactions and these interactions favor small, five, and tie problems. The theory is implemented as a connectionist model, and simulation data are described that are in good accord with empirical data.

How are arithmetic problems solved? Direct retrieval is an important way by which to come to a solution, and many theories have been proposed to describe the underlying mechanism of arithmetical fact retrieval (e.g., Ashcraft, 1987; Campbell, 1995; Groen & Parkman, 1972; Siegler, 1988). However, there is no consensus on this issue. The matter is complicated by the fact that retrieval is not the only way to solve an arithmetic problem. For instance, participants can also rely on elaboration strategies (e.g., $8 + 5 = 8 + 2 + 3$). Self-report measures show that nonretrieval strategies are relatively frequently used in addition, division, and subtraction, but less so in multiplication (e.g., only 3% of the trials for the non-Asian participants in Campbell & Xue, 2001; but see Hecht, 1999; Lefevre, Bisanz, et al., 1996). Because we intended to specifically address the issue of arithmetical retrieval, we focused on multiplication arithmetic. For the same reason, we focused on single-digit arithmetic, since multiplication tables of higher numbers (e.g., the 16-table) are not generally taught in Western curricula. Therefore, they are not expected to be stored for retrieval. Furthermore, since it is generally assumed that multiplication problems with 0 or 1 are not solved by retrieval but by rule (e.g., McCloskey, Aliminosa, & Sokol, 1991; McCloskey, Harley, & Sokol, 1991), we focused on the so-called *standard set* with operands ranging from 2 to 9.

How, then, are multiplication solutions stored and retrieved? An influential view is that problems are stored

in an associative network (e.g., Ashcraft, 1987; Campbell, 1995; McCloskey & Lindemann, 1992; Stazyk, Ashcraft, & Hamann, 1982; Winkelman & Schmidt, 1974). However, the internal organization of this network is heavily debated (McCloskey, Harley, & Sokol, 1991; see also our discussion of these models below), partly because none of the models proposed thus far has been able to account for the major findings regarding multiplication on the basis of well-motivated assumptions. In this article, we will show that a great deal of the behavioral effects can be understood from an associative network point of view, if one takes into account the regularities in the internal structure of multiplication tables. What we propose is that answers to multiplication problems are retrieved by the dual processes of cooperation and competition. If a particular problem is presented, a number of neighboring nodes in a semantic associative network are activated, each corresponding to a particular problem. For example, upon presentation of 7×4 , the representation of the problems 7×3 , 7×5 , 6×4 , and 8×4 will also be activated to some degree. Direct evidence for this assumption comes from Galfano, Rusconi, and Umiltà (2003), who showed that presentation of two numbers (e.g., 7 and 4) automatically activates not only the correct problem and its response (as demonstrated by Thibodeau, Lefevre, & Bisanz, 1996), but also the immediate neighbors of that multiplication problem (e.g., 21 and 35). Specifically, participants were given a cue pair of numbers (e.g., 8 and 3), followed by a different (probe) number (e.g., 16), and had to indicate whether the probe number matched any of the two cue numbers. They found that if the probe was operand and related to the product formed by the two cues [in the example, $16 = 8 \times (3 - 1)$], rejecting the probe as matching any of the two cues was more difficult than with a

This work was supported by Grant 5/04 from Interuniversity Attraction Poles and a GOA grant from the Ghent University Research Council to W.F. We thank Jamie Campbell for the use of his data. Correspondence concerning this article should be sent to T. Verguts, H. Dunantlaan 2, 9000 Ghent, Belgium (e-mail: tom.verguts@ugent.be).

neutral (not operand-related) cue pair. For the operand-related probes, only probes with a distance of one to the product formed by the two cues were used [e.g., 16 and $8 \times (3 - 1)$]. This result shows that activation of close operand-related answers occurs quickly and automatically.

Our key proposal is that, if the different candidate solutions converge on the same unit or decade part of the answer, cooperation will occur between these answers, and retrieval will be relatively easy. For example, the answer "21" points in the correct direction for the decade part of the problem 7×4 , but "35" does not. On the other hand, if many candidates point in a different direction (as "35" does for both decades and units), competition will ensue, and the retrieval process will be slowed down. If this simple principle is applied systematically to the standard set multiplication tables, the major effects reported in the literature (e.g., the size, tie, and five effects) are obtained. This cooperation/competition idea is formulated in a mathematical model so that specific predictions can be derived.

The remainder of this article will be organized as follows. First, we will review the major effects in response time (RT) and error data of single-digit multiplication arithmetic. Second, the most prominent models of multiplication arithmetic will be described. Third, we will propose a new model that is based on the cooperation/competition theory. Next, three simulation studies will show that the model accounts for the major effects in multiplication arithmetic. We will close with a general discussion in which our findings are integrated in a broader context.

MAJOR EFFECTS IN MULTIPLICATION ARITHMETIC

RT Effects

A first robust effect on RTs is the problem size effect (or the *size effect*, for short), meaning that problems with larger operands are solved more slowly than problems with smaller operands (e.g., Stazyk et al., 1982). A second effect is the *tie effect*, which means that tie problems (e.g., 4×4) are solved more quickly than nonties. Furthermore, an interaction between problem size and tie has been documented (Campbell & Gunter, 2002; Lefevre, Bisanz, et al., 1996), in the sense that the size effect is weaker for ties than for nonties. So a third relevant finding is the *size $\times$ tie* interaction. A fourth finding is the fact that problems with an operand of 5 are solved more quickly than would be predicted on the basis of their size (e.g., Siegler, 1988). We call this the *five effect*. Note that

it is logically possible that any of these four effects (size, tie, size $\times$ tie, and five) occurs in the absence of any of the other three, so they are four distinct facts.

Error Effects

Error rates. The four RT effects have a counterpart in error data, in the sense that there is a size, a tie, a size $\times$ tie, and a five effect in error rates. In particular, error rates are higher for large problems relative to small problems, for nontie problems relative to tie problems, and for five problems relative to other problems. Also, the size effect is weaker for ties than for nonties (size $\times$ tie interaction).

Error types. The types of errors also show interesting regularities (Campbell, 1994, 1997). Most important, the majority of errors are operand related. An error is operand related if it is the answer to a problem in which one of the operands is different (e.g., the answer 40 ($= 5 \times 8$) to the problem 5×7). Furthermore, operand-related errors are characterized by a distance effect: Responses are more likely to be correct answers for problems nearby in the multiplication tables than for problems far away. For example, in the problem 5×7 , the answer is more likely to be 40 [$= 5 \times (7 + 1)$] than 45 [$= 5 \times (7 + 2)$]. Besides operand-related errors, a second frequent type of errors is the table-related errors, which are errors that belong to the table but are not a multiple of one of the operands (e.g., the problem 6×8 , the answer 49). Error frequencies per type for the data in Campbell (1994) are shown in the first row of Table 1. Since there is some ambiguity in the classification of errors, in what follows we will tabulate only errors that can be unambiguously classified.

Developmental Trends

The developmental transition from procedure- and strategy-based calculation to retrieval strategies in multiplication has been widely documented (e.g., Cooney, Swanson, & Ladd, 1988; Lemaire & Siegler, 1995). This transition is gradual, but from around the fourth grade, most children appear to solve most single-digit multiplication problems by means of retrieval (Cooney et al., 1988). But this is not the final developmental stage: Within retrieval, further developmental changes occur. Although a systematic description of retrieval-based changes is lacking in the literature, the following four trends can be extracted. First, obviously, RTs are shorter and accuracy is higher for older children (e.g., Campbell & Graham, 1985). Second, the RT size effect is smaller for older children (Campbell & Graham, 1985; Koshmider & Ashcraft, 1991). Third, the proportion of operand-related errors in-

Table 1
Error Percentages per Error Type

	Total Error%	Operand ± 1	Operand $\geq \pm 2$	Table	Other
Campbell (1994)	7.16	50.64	22.88	14.91	11.57
Basic model	4.72	86.72	12.50	0.78	0
With frequencies	9.44	87.2	7.85	4.96	0
Developmental	22.75	58.18	23.98	17.84	0

creases with age (Butterworth, Marchesini, & Girelli, 2003; Campbell & Graham, 1985). Finally, the proportion of operand-related errors that differ from the correct answer by only one operand unit also increases with age—for example, to the problem 7×6 , the answer 48 [= $(7 + 1) \times 6$]; (Butterworth et al., 2003; Campbell & Graham, 1985). In other words, operand-related errors show a stronger distance effect with increases in age.

REVIEW OF MODELS OF MULTIPLICATION ARITHMETIC

Earlier models of mental arithmetic have dealt with a number of the issues described above. The most influential models are the class of *table search models* (e.g., Ashcraft & Battaglia, 1978; Geary, Widaman, & Little, 1986), the *distribution-of-associations model* of Siegler (1988), the *network retrieval model* of Ashcraft (1987), the neural network called *Mathnet*, developed by McCloskey and Lindemann (1992), and the *network interference model* of Campbell (1995). We will briefly describe each of these models and indicate how each deals with the four effects reported in the previous section (size, tie, size $\times$ tie, and five).

Table Search Models

In a table search model (e.g., Geary et al., 1986), each of the two problem operands corresponds to either the rows or the columns of a rectangular array. An answer is retrieved by stepping to the appropriate row and column and reading off the correct answer in the array.

How do such models fare in accounting for the behavioral effects? The size effect originates from the fact that it takes more time to “walk” to the appropriate row and column if the corresponding numbers are larger. However, tie and five effects cannot be explained within this model. Due to these problems, table search models are no longer considered viable candidates as models of mental arithmetic. Nevertheless, they have been important in shaping later models, such as the network retrieval model discussed below.

Siegler (1988):

Distribution-of-Associations Model

Siegler's (1988) model is motivated largely by procedural specificities involved in the learning of associations between problems and their answers during development. He assumes that children initially solve multiplication problems by nonretrieval (*backup*) procedures, such as repeated addition. Each time an answer is found with such an alternative strategy, the problem (e.g., 5×6) is associated more strongly with the answer (in the example, with 30 if the answer was correct, or with some other number if not). Later, if the different problems have accrued much associative strength to one or more answers, the answer is retrieved directly. In this account, the problem size effect derives primarily from the fact that more errors are made to problems that are difficult to solve by

means of such backup procedures. Problems with large operands are difficult to solve with a backup procedure because, for example, the number of required operations is generally higher (e.g., with the backup strategy of repeated addition, compare solving 9×2 and 9×6 as $9 + 9$ and $9 + 9 + 9 + 9 + 9 + 9$, respectively). Therefore, problems with large operands will accrue a relatively large amount of association toward incorrect solutions. It will, therefore, take longer to retrieve the correct answer for large problems, and the size effect follows. To account for the five effect, Siegler relies on the fact that the backup strategy of repeated addition is easier for fives than for other numbers (compare $4 \times 5 = 5 + 5 + 5 + 5$ with $4 \times 7 = 7 + 7 + 7 + 7$), so fewer incorrect associations are formed for five problems and, therefore, retrieval RTs will be faster.

To account for the tie effect, Siegler (1988) resorts to a frequency explanation. The author reports significantly fewer nontie than tie problems on the basis of frequencies from a third-grade workbook (his Table 4). However, Siegler's analysis was performed on problems involving all the digits from 0 to 9. Since the tie effect in arithmetic performance is especially prominent for larger problems (in line with the tie $\times$ size interaction discussed earlier), at the very least one should predict that the frequency difference between ties and nonties is also present when one restricts the analysis to the set of problems involving larger numbers (6–9) only. Hence, we compare the frequencies of tie versus nontie problems for operands 6–9 on the data reported in Table 4 of Siegler. Tie problems turn out not to be more frequent [$t(3) = 0.874, p = .446$]. Besides Siegler, Ashcraft and Christy (1995) reported similar frequency data for multiplication problems. When their data are analyzed in this respect, the effect even seems to be in the reverse direction [$t(3) = -5.246, p = .013$]. This calls into question the basic fact on which the frequency explanation for the tie effect is based. A similar argument has already been made by Baroody (1999).

Ashcraft (1987): Network Retrieval Model

Ashcraft (1987) described a network model with two input fields, one for each operand, and one output field. Each operand node in the input field is connected to its answer nodes in the output field (e.g., the input node for 7 in each of the two input fields is connected with 7, 14, 21, and so on, at the output side). Furthermore, each of the answer nodes spreads activation via connections to its neighbors [e.g., activation in the 35 (= 5×7) answer node spreads activation to all (other) $N \times 7$ and $5 \times N$ nodes for different values of N]. After activation has spread, the number associated with the most active node is given as a response. RT is determined as a monotonically decreasing function of activation.

In this model, the size effect originates from the lower presentation frequency of numbers with larger operands. Lower frequencies lead to weaker connections from input nodes to answer nodes for problems with larger operands, which in turn results in lower activation values

of the correct answer for problems with larger operands and, hence, to larger RTs. As a motivation for this differential frequency assumption, Ashcraft and Christy (1995) reported larger frequencies of small-number problems in arithmetic books for elementary school in Grades 2–6. Explanation of the tie and five effects rely on a frequency account as well. Specifically, ties and five problems are assumed to be relatively more frequent than would be expected on the basis of their size.

Larger problems indeed occur less frequently than small problems (e.g., Ashcraft & Christy, 1995), so the explanation for the size effect is plausible. However, the frequency account has trouble dealing with the tie effect, as was discussed in the context of Siegler's model. Concerning the frequencies of five problems, the evidence is mixed: In Grades 5 and 6, they are more frequent than four and six problems (the "neighbors" of five problems), but this is not the case in Grades 2–4 (see Ashcraft & Christy, 1995, Appendix, Tables G–K). Hence, whereas the size effect could plausibly be attributed to lower frequencies of larger problems, this is not the case for the tie effect (and hence, also the tie $\times$ size interaction) or the five effect.

McCloskey and Lindemann (1992): Mathnet

Mathnet is a connectionist model containing three layers of nodes. The input layer codes the problem operands, and the output layer codes the answer. The hidden layer is initially undedicated, in the sense that it has random connections with the input and output layers, but these connections are trained in an initial learning phase. After training, upon presentation of a problem, activation is passed through the network until all activation values in the network are stable. The number of cycles needed to reach this stable state is used to model RT.

If Mathnet was presented multiplication problems with equal frequencies during the learning phase, no size effect was observed. However, a size effect was obtained if the authors presented problems with different frequencies. Hence, McCloskey and Lindemann (1992) account for the size effect with a frequency argument. Unfortunately, the tie and five effects were not present in Mathnet's performance, so it remains to be seen whether and how the model can be extended to account for these findings. Moreover, we showed above that frequency does not provide a consistent ground on which to account for the tie and the five effects.

Campbell (1995): Network Interference Model

Campbell's (1995) model has undergone extensive revision throughout the years (Campbell, 1995; Campbell & Graham, 1985; Campbell & Oliphant, 1992), and only the last version will be discussed here. In this model, there are different *problem nodes* that correspond to different operand–operation–answer combinations (e.g., $\{4, 8, \times, 32\}$). When a multiplication problem is given to a participant, each of these nodes receives activation depending on how strongly the current problem matches the constituents of a problem node. For example, the

problem $4 \times 8 = ?$ would strongly activate the problem node $\{4, 8, \times, 32\}$. However, the problem $3 \times 8 = ?$ also partially activates this node. Campbell further assumes that a magnitude code is calculated, which approximates the size of the answer. If a large problem such as 7×8 is presented, this magnitude code activates all large problems ($7 \times 8, 8 \times 9$, etc.). Hence, each problem receives input both from the problem characteristics (operands, operation) and from the magnitude code. As in McCloskey and Lindemann's (1992) model, RT is modeled by the number of cycles the network needs to reach threshold activation at one of the problem nodes.

In earlier versions of Campbell's model, the size effect is explained by frequency and order of acquisition. However, in the 1995 version, the explanatory burden is transferred to the magnitude code. Specifically, the magnitude code obeys a compressive effect (e.g., Dehaene, 1992), so that a large magnitude code activates more problem nodes than a small magnitude code does. For this reason, larger problem nodes suffer more inhibition from other activated problem nodes, leading to higher RTs and higher error rates. The tie effect is explained by assuming that tie problems are stored separately from all other problems: Specifically, when a tie problem is presented, nontie problems are only weakly activated (and vice versa for nonties), and since there are more nonties than ties, the latter will receive less inhibition and result in better performance. The five problems are also assumed to be stored separately. The latter two assumptions, however, are empirically unmotivated, except by the fact that it helps to better account for the data, and they are, therefore, post hoc assumptions.

Moreover, the assumption of a separate store for ties can also be questioned on empirical grounds. Whereas the model predicts that errors to tie problems should primarily be other tie answers, an analysis of the data reported in Campbell (1994) shows that of all the errors to tie problems, only 16% were answers to other tie problems. Hence, the hypothesis that ties are stored separately is not supported.

Conclusion

The models described above were designed primarily to explain the problem size effect in mental arithmetic, and all are successful in doing so. None of the models, however, is also able to account for both the tie and the five effects with empirically well-founded assumptions. In the following, we will present a model that shows that it is possible to account for the size, tie, and five effects on the basis of a unitary mechanism, without the need to invoke different principles for each of the effects.

NEW MODEL

Our basic hypothesis is that many of the relevant findings can be understood from an associative network approach (e.g., Campbell, 1995; Stazyk et al., 1982) and, in particular, from the principles of cooperation and competition. Specifically, presentation of a multiplication prob-

lem diffusely activates a number of nodes in the network, and these nodes can cooperate (if they lead to the same response) or compete (if they lead to different responses). This is in line with many other cognitive models—for example, in language processing, where it is assumed that the number of *friends* and *enemies* of a target word determines naming performance (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996).

Basic Principles

The core component of the model is a *semantic field* representing the multiplication tables. There are two crucial assumptions pertaining to this semantic field. The first is that the network is internally organized on the basis of the size of the operands. This implies that problems with similar operands (e.g., 6×7 and 7×7) are stored closely together, an assumption that is shared by the table search models. Recall that Galfano et al. (2003) found that presentation of a problem activates not only the answer, but also the neighbors of this problem, which provides direct evidence for this assumption. Niedeggen and Rösler (1999) have provided event-related potential (ERP) evidence for this same assumption, which we will discuss below in more detail. The second assumption is that not all problems are represented; rather, for each commutative pair of problems (e.g., 7×5 and 5×7), there is only one representational unit. This assumption is based on a number of grounds. First, Rickard and Bourne (1996; see also Rickard, Healy, & Bourne, 1994) observed in a learning experiment that there is almost as much beneficial transfer from a problem to its reverse (e.g., 2×7 and 7×2) as from a problem to itself. For example, solving the problem 2×7 in a training phase leads to shorter RTs in solving 2×7 in a test phase, and almost the same benefit is found for solving 7×2 in the test phase (but there is no transfer to other problems). This suggests that there is only one representational unit for both problems (see Campbell, 1999, for additional evidence on this assumption). Second, dyscalculic patients exhibit a highly correlated pattern of errors between commutative pairs of problems (e.g., $r = .93$ in Patient P.S., reported in Sokol, McCloskey, Cohen, & Aliminosa, 1991), suggesting that only one representation determines their performance. These results do not say anything about any systematicity regarding which problem in a commutative pair is represented. The recent developmental study of Butterworth et al. (2003) is informative in this respect. They found that, in multiplication, children progressively come to systematically respond more quickly to problems in which the maximum operand is presented as the first digit (e.g., 7×5 is solved more quickly than 5×7). Interestingly, this trend appears only from the fourth grade, which is the grade at which retrieval starts to become the dominant strategy (Cooney et al., 1988).

In line with Butterworth et al.'s (2003) data, we assume that the mental representation of multiplication tables in memory is reduced, in the sense that, on the basis of the commutativity principle, only half of all problems are ex-

plicitly represented. Specifically, we assume that only problems with a $\max \times \min$ organization are explicitly represented. It is worth noting that the latter assumption is not critical; a systematic $\min \times \max$ organization yields exactly the same results. Hence, even if different participants have different storage patterns (either $\max \times \min$ or $\min \times \max$), the results hold. For concreteness, however, we will assume $\max \times \min$ organization. Given such an organization, it follows that, when a problem is presented in the inverse order ($\min \times \max$), the problem is first transformed to a $\max \times \min$ order. This operation takes some time (e.g., LeFevre & Liu, 1997), but the time it takes cannot contribute substantially to the *differences* between problems with different operands (unlike in addition; see Butterworth, Zorzi, Girelli, & Jonkheere, 2001).

The semantics of a table search model can be represented as a rectangle, where rows represent one operand and columns the other. If only $\max \times \min$ problems are stored, it follows that the upper right part of the triangle (i.e., above the diagonal, containing $\min \times \max$ problems, such as 2×3) can be removed. Therefore, the semantic field can be conceived of as a triangular structure, as in the semantic field depicted in Figure 1A. Thus, this structure is reminiscent of the table search models, but unlike these models, the shape of the table is not rectangular, due to the commutativity principle. It is important to note that we do not assume that the network is actually stored in a true triangular form; by no means do we impose any geometrical properties on the model. What matters is the internal pattern of connections in the model, and the triangular structure is introduced only to conveniently represent these connections.

From this semantic field, activation is projected gradually to a field that represents the decade of numbers, and also to a field that represents the unit of numbers. The fact that such a decomposed unit/decade representation exists is motivated by recent data in Nürk, Weger, and Willmes (2001). They found an effect of unit/decade compatibility in two-digit number comparisons, in the sense that RTs were shorter if the response based on the units was the same as that based on the decades (e.g., the pair 63–75) than otherwise (e.g., 63–71). Also, Ratnckx, Brysbaert, and Fias (2005) observed that a prime digit in the correct position facilitated naming a target two-digit number (e.g., the target 12, the prime 18), whereas a prime digit in the other position inhibited naming the target (e.g., the target 12, the prime 23). These findings suggest that units and decades are represented separately.

From the two decomposition fields, activation is projected to a response field, which contains one unit for each possible response (numbers 1–99). If one of the nodes in the response field reaches a fixed threshold value, a response is emitted, and the corresponding response can be read off from the network.

Implementation

The architecture of the model is shown schematically in Figure 1A. Each of the fields in this model contains units, and each unit represents either a number (e.g., 6)

Figure 1. (A) Architecture of the model. For an explanation of the symbols in each of the fields, see the Appendix. (B) Activation curve in the semantic field some time after presentation of the problem 6×4 . (C) Graphical illustration of weights between the max input field and the semantic field. The influence of parameters α (peakedness) and C (height) are illustrated. Solid lines denote weights with $\alpha = .25$. Dotted line denotes weights with $\alpha = .75$. Lines with asterisks denote weights with $C = .5$. Line without asterisks denotes weights with $C = .2$.

or a pair of numbers (e.g., 6×4). For example, one unit in the max-operand field is activated if the number 6 is the maximal number in the problem. Furthermore, one node in the semantics field represents the problem pair 6×4 and can, therefore, be called a 6×4 node. The equations describing the flow of activation within and between these units are presented in the Appendix.

Input fields. The two operands of a multiplication problem are represented in two separate input fields (*max operand* and *min operand* in Figure 1A), the first of

which represents the maximal number, and the second the minimal number. The maximal number activates exactly one node in the max input field (e.g., the fifth node, if 5 is the maximal operand), and the minimum number also activates exactly one node in the min input field.

Semantic field. The two-dimensional semantic field is where the different solutions are represented. This is the triangular structure depicted in Figure 1A (*semantics*). Note that we make no assumptions on the discussion in the literature as to whether this semantic field is

phonological (e.g., Dehaene & Cohen, 1997) or of an “abstract meaning-based form” (Whalen, McCloskey, Lindemann, & Bouton, 2002, p. 519).

Each multiplication problem diffusely activates this field. For example, if 6×4 is presented, activation in the semantic field will be most strongly concentrated at row 6, column 4, but there will also be some activation at neighboring positions (e.g., row 6, column 5). The evidence for this assumption, obtained by Galfano et al. (2003), has already been discussed. Figure 1B depicts the activation distribution in the semantic field some time after presentation of 6×4 . The unit at the location (row 6, column 4) is maximally activated, and activation drops gradually as one walks away from the maximum. It is clear that activation remains relatively high at operand-related problems (e.g., 7×4 , 8×4), but there is also some activation at other (table-related) problems (e.g., 5×3). ERP evidence for this type of activation distribution was recently described by Niedeggen and Rösler (1999), who found that activation spreads only to table-related problems, and only to table-related problems with a small distance (up to two) to the original problem. This type of distribution also arises in the model. The reason is that activation in the semantic field is a monotonic function of the *sum* of the activation values in the input fields (see the Appendix for details). This activation distribution is very similar to a typical activation distribution assumed on the mental number line (e.g., Dehaene, 1992), except that the present situation refers to a two-dimensional, rather than a one-dimensional, representation.

Such an activation distribution is characterized by its location in the semantic field, its height, and its peakedness. The location of the distribution is determined by which units in the input field are active. To describe the parameters determining the height and peakedness of the distribution, consider the connection weight between a node in the max operand field and a node in the semantic field. We denote this connection weight by w_{ij}^{in} , for a node i in the max operand field and any node in row j of the semantic field. Hence, all connections from a particular node in the max operand field to each of the units in a particular row of the semantic field have the same value. However, the values are different for different rows. The equation describing the connection weight between two such units equals

$$w_{ij}^{\text{in}} = \exp(-\alpha|i - j|) - C. \quad (1)$$

Figure 1C illustrates Equation 1 and its parameters graphically. It shows connection weight values (ordinate) coming from node 6 in the max operand field, going to units in rows 4–8 in the semantic field (abscissa). The solid line represents a small α ($\alpha = .25$, a weakly peaked curve), and the dotted line represents a large α ($\alpha = .75$, a strongly peaked curve). Larger values of α lead to a more strongly peaked pattern of weights across rows, and this leads to more strongly peaked activation distributions (as in Figure 1B). The parameter C determines

the height of the activation distribution; large values of C imply that the weights are small, which leads to low activation values overall (see Figure 1C for a graphical illustration of the influence of both α and C on the connection weights w_{ij}^{in}). The parameters used in the activation distribution of Figure 1B were $\alpha = .75$ and $C = .50$.

Just as each node in the maximum operand field is connected with equal strength to all the units in a particular *row* of the semantic field, likewise, each unit in the minimum operand input field is connected with equal strength to all the units in a particular *column* of the semantic field. Correspondingly, Equation 1 applies to a particular unit i in the min field and to all the units in column j in the semantic field. As a result of this connection pattern, the unit at, for example, row 6 and column 4 in the semantic field will be the most active unit when the problem 6×4 is presented as an input. For this reason, this unit can be called the 6×4 unit.

Decomposition fields. There are two decomposition fields: a decade field and a unit field (*decades* and *units* in Figure 1A). They are called decomposition fields because they decompose the answer in its base-10 lexical representations. For example, if 6×4 is presented, the 6×4 unit and the surrounding units in the semantic field become activated. From there, activation is propagated to the decade and unit decomposition fields. The 6×4 unit in the semantic field passes activation to the *twenty* unit in the decade field and to the *four* unit in the unit field. As a result of the diffuse pattern of activation in the semantic field, multiple units in the decade and unit fields will be activated. Competitive interactions occur in each of the decomposition fields in the form of lateral inhibition, as is depicted schematically by the within-field arrows in Figure 1A.

Response field. The information in the decomposition fields must be transformed into a single representation for output. This is the purpose of the fourth and final field (*response* in Figure 1A), which contains a node for each possible nonzero integer number below 100. If one of the response units reaches threshold activation value, a response is executed. In this way, both choice and RT data can be recorded. In most experiments, only the standard set of problems is presented (operands 2–9). Furthermore, errors are predominantly answers to other problems belonging to this standard set (e.g., 96% of all the errors in the data in Campbell, 1994). For this reason, we assumed that only answers of the standard set can be activated in the response field. One way to conceptualize this is that a (context) node specific for a multiplication context is connected to all the standard set answers, so that they have a prior advantage for finishing first, relative to nonstandard set answers. Without this assumption, errors made by the model were predominantly answers outside the standard set. This context node does not in any way provide the answer to the network: It simply activates the whole set of plausible (i.e., standard) responses. Technically, we assumed that the rate of activation increase of nonstandard set answers was zero. By

increasing this rate, one could, in principle, model the generation of the 4% nonstandard set answers.

SIMULATION 1 Basic Model Without Frequencies

One factor that has been hypothesized to determine RTs in mental multiplication is problem frequency. We assume this also and will incorporate this factor into the model. However, in a first simulation, we will *not* use frequency, in order to show that for many of the effects, frequency is not needed. In fact, as was noted above, the tie and five effects *cannot* be explained by a frequency account, because tie and five problems are not more frequent than other problems. Hence, by first presenting the results without frequency, it is clear where exactly frequency has an explanatory role in the model and where not.

The model's performance will be compared with the data in Campbell (1994). He presented addition and mul-

tiplication problems in both Arabic and verbal formats, but we focus on the multiplication data in Arabic format. The results were very similar when other relevant data were analyzed in this respect (e.g., Campbell, 1997; Campbell & Xue, 2001).

Parameter Settings

In each simulation round, all 36 problems are presented once to the network (recall that the order of the problems is not modeled explicitly, so for example, 4×6 and 6×4 are the same problem for the model). One hundred simulation rounds were performed. Most parameters of the model are scaling parameters that do not influence the qualitative pattern of data predicted by the model. Their values are given in the Appendix. The parameter α (which determines the peakedness of the activation curves) was set at .75. To obtain variability in model performance, the parameter C (which determines the height of the activation curves) was sampled in each simulation round from a uniform distribution over (.4,

Figure 2. Observed response time (RT) data (panels A, B, F, and G) and simulated RT data (panels C–E and H–J). Simulated RTs are expressed in arbitrary model time units. (A) Mean RT per operand for adult data (Campbell, 1994). (B) Mean RT per operand for children's data (Campbell & Graham, 1985). (C) Simulation of mean RT per operand for model without frequencies. (D) Simulation of mean RT per operand for model with frequencies. (E) Simulation for developmental model. (F–J) Interaction of problem size (small [Sm] or large [La]; see the text for definition) with the tie status of the problem, for adults (panel F), children (panel G), model without frequencies (panel H), model with frequencies (panel I), and developmental model (panel J). Note that the scale of the developmental simulation (panels E and J) is different from the scale of the other two (adult) simulations, because the scale is different for the adults' and the children's data also.

.9). Hence, on some presentations of a problem, many neighbors were activated (low C), and on other presentations of that same problem, no or only a few neighbors were activated (high C). Variability in C can be thought of as either within-subjects or between-subjects variability. The value of α and the range of values for C were chosen to obtain approximately the same number of errors in the model data as in the human data.

Results

RT effects. Figure 2A shows the mean RTs per operand for Campbell's (1994) data. Figure 2F shows the mean RT for small and large problems, for ties and nonties separately. In line with Campbell and Gunter (2002), small problems were all the problems with both operands smaller than or equal to five, and larger problems were those with both operands larger than five. Figures 2C and 2H show the corresponding model-based results. The size and the five effects are apparent both in the human data (Figure 2A) and in the model data (Figure 2C). The correlation between observed and simulated RTs equals .97 ($p < .01$; calculated over data points in Figures 2A and 2C). The tie effect and the size $\times$ tie interaction are also apparent in the data (Figure 2F) and, although to a lesser extent, in the simulation (Figure 2H) as well.

Error rate effects. The error rates per operand are shown in Figure 3A. Figure 3F shows error rates for small and large problems, for ties and nonties separately. The corresponding model data are shown in Figures 3C and 3H. The size and five effects appear in the human data (Figure 3A); the five effect is simulated by the model (Figure 3C), but the size effect is not. The correlation between model and data error rates (data points in Figures 3A and 3C) equals .23 ($p = .32$). The tie effect and the size $\times$ tie interaction appear in the empirical error rates (Figure 3F) and also in the model error rates (Figure 3H).

Error types. The total error rate and error rates per type are shown in Table 1. Although the model is not quantitatively accurate, it captures the three basic effects in error types. First, operand errors are the most common kind of error. Second, there is a distance effect in the sense that the operand ± 1 error is the most common of all operand errors. Third, there is a small number of table errors.

Discussion

We now will describe in detail why the cooperation and competition principle that drives the network accounts for the RT effects. Error rates and error types will be discussed afterward.

Figure 3. As in Figure 2, but with error rate as the dependent variable.

Size effect. There is *cooperation* between multiplication problems if the semantic node for a particular problem has consistent neighbors in its unit/decade decomposition. For example, if 6×4 is presented, the 28 semantic node will also be (weakly) activated. Since this node also projects to the 20s node in the decade field, activation of that node will increase, and in this sense, 28 will cooperate with 24 when 6×4 is solved. For this reason, we say that neighbors 6×4 and 7×4 ($= 28$) are *consistent* in the decade part of their answer. In the decade decomposition field, such cooperations occur more often for small than for large problems. This can be seen in Table 2A, which shows the decade part of each problem. For example, the problem 5×3 has three consistent neighbors and one inconsistent neighbor in the decade part¹: The neighbors 5×2 , 6×3 , and 4×3 are consistent; the neighbor 5×4 is inconsistent. On the other hand, a problem such as 8×7 has four inconsistent neighbors and no consistent neighbors. For this reason, activation development in the decade field will be slower for 8×7 than for 5×4 , which is part of the reason why a size effect appears in the model. Note that Table 2A reflects the triangular structure of the semantic field.

There is also a *competition* aspect at work in the decade and unit fields. Specifically, all elements within each decomposition field inhibit one another. Therefore, a problem such as 8×7 will have a double handicap relative to 5×4 : It has fewer consistent neighbors (cooperation) and more inconsistent neighbors (competition). In general, large problems have more competition and less cooperation than small problems do, and the size effect follows. To sum up then, the problem size effect arises in the model because there is less cooperation and more competition from neighboring nodes for large problems than for small problems. Hence, given the principles of cooperation and competition, the internal structure of the multiplication table is sufficient to account for the size effect.

Tie effect. Due to the assumption that all commutative pairs are represented by one node in the semantic field, this field can be represented as a triangular structure. All tie problems are on the main diagonal of this triangle and, therefore, have fewer neighbors. Since in general there are more competitive neighbors than cooperative neighbors (see Tables 2A and 2B), there will be a tie advantage relative to the nontie problems.

Size $\times$ tie interaction. An inspection of Table 2A shows that there is a positive correlation between size and number of competitors (see the discussion of size effect above), but this correlation is smaller for the tie problems (on the diagonal) than for the nontie problems (off-diagonal). Hence, the size effect is smaller for the tie problems than for the nontie problems. For example, the number of interfering decades of neighbors in the 2×2 and 7×7 problems is 0 and 1, respectively (see Table 2A). On the other hand, for the problems 3×2 and 8×7 , the number of alternative decades is 0 and 4, respectively. A size $\times$ tie interaction is thus obtained.

Five effect. The five effect is generated by the model due to the fact that, for neighbors two steps away from the other operand (e.g., neighbors 3×5 and 7×5 for the problem 5×5), the unit value is always consistent (and hence, cooperative; see Table 2B). The same is not true for any other operand. These *far* neighbors have a small influence, but it is enough to account for the relative speed for five problems.

Error rate effects. The RT effects are determined primarily by the tendency for a problem and its neighbors to point to the same decade and unit digit (cooperation/competition). The error rate is also determined by this tendency, but relative to the tendency of neighbors to point *consistently* to a *different* decade or unit digit. In particular, if there is a consistent trend toward a particular unit or decade among these neighbors (different from the correct unit or decade), the error rate will be high. This observation accounts for the tie effect: Ties have fewer neighbors than nonties do, and hence, less cooperation is possible among tie neighbors, thus leading to fewer errors on ties. Five problems are less error prone because there is a strong tendency for neighbors of a particular problem to point to the *correct* unit digit (e.g., 3×5 and 7×5 for the problem 5×5). This tendency is stronger than the opposing tendency for immediate neighbors to point to a different unit (4×5 and 6×5). Hence, a five effect follows. However, there is no strong tendency for larger problems to have more neighbors that consistently point to a decade or unit different from the correct decade or unit, so the size effect does not occur in the model simulations.

Error types. Not only is the correct node in the semantic field activated, but the neighbors of the correct

Table 2A
Decade Parts of Answers to Multiplication Problems

Maximum Operand	Minimum Operand							
	2	3	4	5	6	7	8	9
2	0							
3	0	0						
4	0	1	1					
5	1	1	2	2				
6	1	1	2	3	3			
7	1	2	2	3	4	4		
8	1	2	3	4	4	5	6	
9	1	2	3	4	5	6	7	8

Table 2B
Unit Parts of Answers to Multiplication Problems

Maximum Operand	Minimum Operand							
	2	3	4	5	6	7	8	9
2	4							
3	6	9						
4	8	2	6					
5	0	5	0	5				
6	2	8	4	0	6			
7	4	1	8	5	2	9		
8	6	4	2	0	8	6	4	
9	8	7	6	5	4	3	2	1

node are activated as well; in the response field, one of the neighboring answer nodes will sometimes reach threshold earlier than the correct node if this neighboring node gets sufficient cooperation from other problems activated in the semantic field. The type of activation spreading in this model (see Figure 1B) leads to relatively high levels of activation for operand-related problems, so operand-related errors are the most frequent error type. Within operand-related errors, problems closer to the correct answer have higher levels of activation, and this generates an operand distance effect. However, table errors are incidentally made as well, because non-operand-related problems are also weakly activated.

To sum up, general processing properties of spreading of activation and cooperation/competition mechanisms were implemented in a model that stores multiplication tables in a semantic field, but reduced due to commutativity. From this semantic field, there were separate projections to unit and decade representations. With this setup, many of the effects in mental multiplication could be traced back to the internal structure of the multiplication tables that were stored in the semantic field. The next question is whether the fit of the model can still be improved by taking into account the different frequencies of problems (e.g., Ashcraft & Christy, 1995). Given the ubiquitous frequency effects in other domains (e.g., reading aloud; Seidenberg & McClelland, 1989), it is plausible that frequency also is a factor in multiplication performance. This will be the subject of the next section.

SIMULATION 2 Model With Frequencies

Frequency may play a role into the speed of incorporating a problem into the retrieval network. Once the semantic field is constructed, however, frequency may continue to exert an influence, and one very plausible source of influence for frequency is the peakedness of the activation distribution (Figure 1B). Indeed, more frequently trained problems are solved more accurately (Pauli, Bourne, & Birbaumer, 1998), and the most straightforward way to increase accuracy on a problem in the model is to increase the peakedness of the activation distribution that is generated in the semantic field upon presentation of that problem.

Since the peakedness of the activation distributions is determined by the parameter α (see Equation 2), we scaled this parameter for each problem according to that problem's frequency. In particular, rather than setting each parameter α at a fixed value of .75, we set a minimum value of $\alpha = .50$ for the operand pair with the lowest frequency and a maximum α value of 1.0 for the operand pair with the highest frequency. Values of α for other operand pairs were linearly interpolated on the basis of their frequency. This procedure was based on the frequency data of Ashcraft and Christy (1995), but the frequency data of Siegler (1988) gave similar results. Except for the adjustment of this parameter, exactly the same procedure was followed as that in Simulation 1.

Results and Discussion

RT effects. The trends in the RT data are similar to those from the previous simulation, as can be seen in Figures 2D and 2I. The correlation between observed and simulated mean RTs (Figures 2A and 2D) equals .98 ($p < .01$).

Error rate effects. The error rates are shown in Figures 3D and 3I. It is clear that there is again a five effect and a tie effect, as in the previous simulation, but the size effect is now also present. The correlation between observed and simulated error rates (Figures 3A and 3D) equals .70 ($p < .05$). There are only two small deviations now, for operands 2 (or 3) and 9. However, for some problems, the reported frequencies may yield distorted estimates. For example, in solving the problem $9 \times N$, it can be safely assumed that participants often resort to a strategy of the type $(10 \times N) - N$, which is much easier than the direct retrieval. As a consequence, the reported frequencies for the $9 \times N$ problems are overestimations of the frequency with which $9 \times N$ problems are retrieved (rather than solved by strategy). More generally, the frequencies that are derived from schoolbooks may be only rough estimates of real-world frequencies. Given this difficulty in obtaining reliable frequency data, adapting the model to the empirical frequencies does a good job of accounting for the size effect in the error rate data, in addition to the other data that were already well accounted for in Simulation 1.

Error type effects. Incorporating frequencies did not substantially change the type of errors made by the model (see Table 1).

To sum up, frequency improved the fit of the model; in particular, predictions on the size effect in error rates were more accurate with frequencies. Note, however, that frequency is not the most critical variable in the model: As was shown in Simulation 1, most of the effects are accounted for without frequency, and as discussed in our review of earlier models, frequency cannot even account for many of the effects. Also, it is more appropriate to think of frequency as a factor that modulates the more fundamental cooperation/competition principle (i.e., frequency modulates the peakedness of the activation curves), rather than being an additional principle.

SIMULATION 3 Developmental Trends

As was noted earlier, from the fourth grade on, the dominant strategy for solving multiplication problems is retrieval. However, this does not mean that development stops there, because there are clear developmental changes long after this stage. To recapitulate, four effects can be extracted from the literature for children in the fourth grade: First, RTs become shorter, and error rates become smaller; second, the size effect in RTs becomes smaller; third, the proportion of operand-related errors increases; and fourth, within operand-related errors, the proportion of errors of the type operand ± 1 increases.

In the previous simulation, the peakedness of the activation distributions was adjusted on the basis of the fre-

quencies of the multiplication problems. To simulate an early developmental stage, it is appropriate to assume lower frequencies for all the problems, relative to the *adult* model. In line with the logic of Simulation 2, this can be accomplished by decreasing the peakedness parameter α for all the problems. To implement this, we subtracted .25 from the minimum and maximum α , so the minimum and maximum values became .25 and .75, respectively. Other α values were linearly interpolated. Except for the adjustment of this parameter, exactly the same procedure was followed as that in Simulations 1 and 2. As an empirical standard on which to base the model–data correlations, we took the RT and error data reported in Campbell and Graham (1985) for Grade 4 children (results were similar for the Grade 5 data reported in Campbell & Graham, 1985). These RT data are shown in Figures 2B and 2G (RT per operand and size $\times$ tie data, respectively). The corresponding error data are shown in Figures 3B and 3G.

Results and Discussion

RT effects. Simulated mean RT data per operand and RT size $\times$ tie data are shown in Figures 2E and 2J, respectively. The qualitative trend is similar to that observed in Simulations 1 and 2. The correlation between observed and simulated mean RTs (Figures 2B and 2E) equals .86 ($p < .01$).

With respect to the developmental trends, RTs are higher than those in Simulation 1 or 2. Also, the size effect is larger in Simulation 3 than in Simulation 1 or 2. The size effect (expressed as the range of the operand-specific RTs divided by overall mean RT) is 5.6%, 5.8%, and 8.7% in Simulations 1, 2, and 3, respectively.

Error rate effects. Simulated error rates are shown in Figures 3E and 3J. Due to the more diffuse activation distributions (lower α values), the overall error rate (22.75%) is higher than that in the previous two simulations. The correlation between observed and simulated error rates (Figures 3B and 3E) equals .97 ($p < .01$).

Error type effects. Error rates per type are shown in the last row of Table 1. In line with the developmental data, the percentage of operand-related errors is lower in the developmental model than in the other two (*adult*) models. Also, within the operand errors, there are fewer errors with a distance of one (± 1 errors in Table 1) in the developmental model.

The picture that emerges from this modeling effort is that children with an intact retrieval system initially have very broad activation distributions around or close to the solution in the semantic field. With experience, these distributions gradually become more peaked, so that the RT size effect becomes smaller. Also, errors become increasingly of the operand-related type with a distance of one.

GENERAL DISCUSSION

A model has been described that captures a wide range of findings in mental multiplication. The basic assumption is that problems, or answers to problems, un-

dergo cooperation and competition and that all effects arise from the internal structure of the multiplication tables. In this sense, we agree with earlier theorists (e.g., Campbell, 1995; Stazyk et al., 1982), who claimed that many major findings in single-digit multiplication arithmetic (RT, error rate, and error types) can be explained in terms of interference between problems and answers to problems. In fact, the idea that problems can interfere with each other is the basic assumption in almost all the models discussed above, with the exception of the table search models. Other assumptions in the present model that are not shared with the earlier models are the reduction of the semantic field on the basis of commutativity and the separate unit/decade representations. We have argued that these assumptions are empirically valid. Importantly, to account for the five and tie effects, the present model does not need the assumption that these problems are more frequent (as in the distribution-of-associations model or the network retrieval model) or that these problems are stored separately (as in the network interference model), since we have argued that these assumptions are not empirically supported.

An advantage of the present work is that we have explicitly shown by model simulation that the concept of interference can, indeed, account for many of these findings with independently and empirically motivated assumptions. These effects are size, tie, and five effects and a size $\times$ tie interaction, in both RT and error rate. Furthermore, we successfully predicted the most common error type (operand-related errors) and a number of developmental findings within retrieval (more operand-related errors with increases in age and closer operand-related errors with increases in age).

We have tacitly assumed that all multiplication problems are solved by direct retrieval rather than by, for example, repeated addition or some other strategy. The issue of retrieval versus nonretrieval will now be taken up first. After this, we will integrate our findings into a broader context by interpreting a number of related phenomena from the point of view of the model and its underlying assumptions.

Retrieval and Nonretrieval

To what extent multiplication problems are solved by retrieval in adults is a matter of considerable debate (e.g., Kirk & Ashcraft, 2001; Lefevre, Bisanz, et al., 1996). If multiplication problems are not solved primarily by retrieval but by such strategies as repeated addition or multiplication by 10 followed by subtraction [e.g., $9 \times 8 = (10 \times 8) - 8$], it might be that some or even all of the major effects could be attributed to nonretrieval strategies. LeFevre, Sadesky, and Bisanz (1996) have convincingly demonstrated that, in the case of addition, the size effect emerges primarily from strategies other than retrieval. To a lesser extent, a similar observation was made by Lefevre, Bisanz, et al. for the case of multiplication. On the other hand, all the models we reviewed, including our own, but with the exception of Siegler's (1988) distribution-of-associations model, focus exclu-

sively on retrieval. It is, therefore, a genuine concern whether retrieval trials dominate single-digit multiplication data. To investigate this issue, we analyzed the data in Campbell and Xue (2001). These authors were interested in multiplication differences across different cultures: For present purposes, to ensure comparability with the data in Campbell (1994) that were discussed above, we will look at the performance of the non-Asian Canadian participants in Campbell and Xue's data set. The authors asked participants, on a trial-to-trial basis, to report the strategy they had used on each problem. Although retrieval was clearly the dominant way of solving multiplication problems (reported on 97% of all the trials), we separately analyzed the problems on which retrieval was reported as the strategy. In these trials, all major effects appeared: There were size, tie, and five effects and a size $\times$ tie interaction in both RT and error data.² This corroborates our assumption that single-digit multiplication data are indeed dominated by retrieval, in line with the findings of Kirk and Ashcraft. The same conclusion may or may not be valid, however, for other operations, such as addition (Lefevre, Sadesky, & Bisanz, 1996), subtraction, or division (Campbell & Xue, 2001). This is not to say, however, that participants of all ages and all cultures always use retrieval in multiplication (Hecht, 1999; Lefevre, Bisanz, et al., 1996). The point here is simply that, in data sets in which there is evidence for strong retrieval-based performance, the standard effects (size, tie, and five) are found. Hence, these effects cannot be solely or primarily the result of strategy use or the result of mixing retrieval and strategy use trials, as has been argued for the size effect in addition (Lefevre, Sadesky, & Bisanz, 1996).

Extensions

There are different ways in which the model can be extended to account for different task dimensions. Some straightforward extensions are the following. First, we have discussed only the Arabic input format. However, it is easy to model other input formats (e.g., verbal) by simply adding extra fields at the input side that code for numbers presented in other input formats. These fields would then project to the semantic field in a way similar to that for the Arabic input field. A second extension would be to assume that after a problem is solved, some activation remains in the network. This would allow us to study performance of the network with priming paradigms (e.g., Campbell, 1987, 1995; Galfano et al., 2003). Third, we have discussed only production tasks until now, but a second frequently used paradigm is *verification*, in which participants have to evaluate the correctness of a proposed result (e.g., Campbell & Graham, 1985). According to Zbrodoff and Logan (1990), the two tasks share the same representations and microprocesses (the network structure and network activation) but operate with different macroprocesses (how the network is accessed). Hence, the verification task may be based on the same underlying representations. Fourth, additional connective pathways could be added to the model. For

example, the input fields may directly activate the response field, so that, for example, the *five* node in the first input field activates response nodes that contain a 5. In this way, naming errors can be modeled also (e.g., Noël, Fias, & Brysbaert, 1997). Finally, we identified larger values of α (steeper activation distributions, more precise access of multiplication facts) with later phases in development. However, also among adults, there can be differences in the precision with which multiplication facts are accessed. Such individual differences can be straightforwardly modeled in the same way, by varying the peakedness of distribution curves (α) across participants. This short review of possible extensions indicates that the model is sufficiently rich to allow future study of other aspects of mental multiplication. This leads us to the broader question of whether other operations, such as division, can also be understood in the same framework.

Division, Addition, and Subtraction

Recent evidence has converged on the fact that division is often solved indirectly via multiplication (e.g., Campbell, 1999; Mauro, LeFevre, & Morris, 2003). It is easy to adapt the model for solving division problems. For example, to solve the problem 72/8, the 8 node in the max input field and the 72 node in the response field can be clamped (i.e., set to a fixed value). Then activation can flow from these two fields, via the semantic field, to the min input field. If no node in the min input field reaches sufficiently high activation, the 8 node in the min input field can be clamped, together with the 72 node in the response field, and activation can flow to the max operand field until a unit reaches activation. In this way, the problem can be solved. This procedure would amount to reformulating the division problem 72/8 as a problem of the type $8 \times ? = 72$. Interestingly, this is exactly what Mauro et al. (2003) found that their participants did for large division problems. On the other hand, Mauro et al. found that small division problems were not mediated by multiplication. Hence, division could be solved by a separate memory structure devoted to small division problems, in combination with an alternative process operating on the multiplication structure (Figure 1). This is again an example of different macroprocesses (in the sense of Zbrodoff & Logan, 1990) for multiplication and division, operating on the same representations (the multiplication fields) and with the same microprocesses. The theory of division outlined here accounts for the fact that all findings in multiplication (size, tie, five, and size $\times$ tie) are also found in division (e.g., Campbell & Xue, 2001). However, given that not enough is known yet about exactly which (small) division problems are stored separately, it is premature to extend the model in this direction, but it is clear that such an extension would be relatively straightforward.

Application of the same memory structure for addition and subtraction, however, is less obvious. Unlike for division, there is no simple one-step procedure that transforms an addition or subtraction problem into a multiplication problem. Hence, it is unlikely that the multiplication

memory structure is used for addition or subtraction. Also, Dehaene, Piazza, Pinel, and Cohen (2003) have reviewed both behavioral and neuroscientific evidence that multiplication and small addition problems are solved by retrieval but that larger addition and subtraction problems are solved by different (e.g., decomposition) strategies. Finally, addition and subtraction problems are usually not taught by rote learning of tables. For these reasons, we think that insights obtained in multiplication/division should be applied to addition or subtraction with caution.

Neuropsychological Data

Any plausible model of multiplication should, at least in principle, be able to account for neuropsychological findings regarding multiplication. The basic finding is that multiplication can be (doubly) dissociated from other numerical operations. For example, it can be dissociated from numerical calculation procedures (McCloskey, Aliminosa, & Sokol, 1991), number comparison (Cipolotti & Butterworth, 1995), and subtraction (Dehaene & Cohen, 1997). Given that the semantic field is specific for multiplication, these findings are easy to incorporate into the framework presented here. The hypothesis that division is mediated by the multiplication memory structure predicts that only one type of dissociation is possible. In particular, division can be impaired when multiplication is not, but not vice versa, and this conclusion is consistent with current neuropsychological findings (Cipolotti & de Lacy Costello, 1995).

A more fine-grained effect is that, in some patients, many errors occur on specific problems (idiosyncratic for each patient), but not on others (e.g., Patients P.S. and G.E. in Sokol et al., 1991). This could be implemented in different ways in the model. For example, the weights to the semantic field could be (randomly) damaged. Alternatively, the nodes in the semantic field itself could be damaged. Either of these manipulations leads (obviously) to increased error rates, but also to errors on specific problems, but not on others, depending on where exactly the damage occurs.

One unresolved issue is whether patients regularly have a larger size effect than healthy participants do. Patients typically make more errors on large than on small problems, but so do healthy participants. Hence, if the size effect with patients is similar to that with healthy adults (although the absolute level of errors is, of course, higher in patients with calculation impairments), no additional explanation is required to account for the patient data. Unfortunately, the findings in the literature are inconsistent: For example, Patient P.S. in Sokol et al. (1991) appears to have a larger size effect than is usual, but Patient G.E., reported in that same article, does not.

A last phenomenon that we mention here is that some patients have proportionally many more nontable (other) errors than control participants do (e.g., answer 17 to the problem 6×2 ; Patients F.W. and T.M. in McCloskey, Aliminosa, & Sokol, 1991). In our simulation, we assumed that only table responses took part in the race and, therefore, nontable responses could not be executed. Equiva-

lently, we may assume that the race times for table responses are much shorter (e.g., because they are used repeatedly during multiplication and, hence, have a higher baseline activation), so that effectively, the table responses are the only participants in the race. However, if we do let nontable responses participate, the number of nontable errors increases strongly. In terms of the model, patients with many nontable errors may not be able to activate table responses to a sufficient degree or may not be able to maintain high baseline activation, so that nontable responses are made from time to time. Given the small amount of data on this issue, it would be premature to make strong assumptions as to exactly what goes wrong in these types of dyscalculia, but it is clear that the model could provide a useful framework for interpreting the performance profiles reported in the literature.

This review has shown that although the model is restricted in scope, it is nevertheless sufficiently rich that it can be useful for understanding other phenomena in mental multiplication, or mental arithmetic more generally. Moreover, many phenomena in memory retrieval at least superficially resemble those observed in mental multiplication. For example, when retrieving the capital of Uruguay, people may (erroneously) say "Caracas" but never "Paris." Mental multiplication may turn out to be an especially clear testing field for investigating such phenomena.

REFERENCES

- ASHCRAFT, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In C. J. Brainerd, R. Kail, & J. Bisanz (Eds.), *Formal methods in developmental research* (pp. 302-338). New York: Springer-Verlag.
- ASHCRAFT, M. H., & BATTAGLIA, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 527-538.
- ASHCRAFT, M. H., & CHRISTY, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1-6. *Journal for Research in Mathematics Education*, **5**, 396-421.
- BAROODY, A. J. (1999). The roles of estimation and the commutativity principle in third graders' mental multiplication. *Journal of Experimental Child Psychology*, **74**, 157-193.
- BUTTERWORTH, B., MARCHESINI, N., & GIRELLI, L. (2003). Multiplication facts: Passive storage or dynamic reorganization? In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetical concepts and skills* (pp. 189-202). Mahwah, NJ: Erlbaum.
- BUTTERWORTH, B., ZORZI, M., GIRELLI, L., & JONKHEERE, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *Quarterly Journal of Experimental Psychology*, **54A**, 1005-1029.
- CAMPBELL, J. I. D. (1987). Network interference and mental arithmetic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 109-123.
- CAMPBELL, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, **53**, 1-44.
- CAMPBELL, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, **1**, 121-164.
- CAMPBELL, J. I. D. (1997). On the relation between skilled performance of simple division and multiplication. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1140-1159.
- CAMPBELL, J. I. D. (1999). Division by multiplication. *Memory & Cognition*, **27**, 791-802.
- CAMPBELL, J. I. D., & GRAHAM, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, **39**, 338-366.

- CAMPBELL, J. I. D., & GUNTER, R. (2002). Calculation, culture, and the repeated operand effect. *Cognition*, **86**, 71-96.
- CAMPBELL, J. I. D., & OLIPHANT, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 331-364). Amsterdam: Elsevier.
- CAMPBELL, J. I. D., & XUE, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, **130**, 299-315.
- CIPOLOTTI, L., & BUTTERWORTH, B. (1995). Toward a multi-route model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, **124**, 375-390.
- CIPOLOTTI, L., & DE LACY COSTELLO, A. (1995). Selective impairment for simple division. *Cortex*, **31**, 433-449.
- COLTHEART, M., RASTLE, K., PERRY, C., LANGDON, R., & ZIEGLER, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.
- COONEY, J. B., SWANSON, H. L., & LADD, S. F. (1988). Acquisition of mental multiplication skill: Evidence for the transition between counting and retrieval strategies. *Cognition & Instruction*, **5**, 323-345.
- DEHAENE, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1-42.
- DEHAENE, S., & COHEN, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, **33**, 219-250.
- DEHAENE, S., PIAZZA, M., PINEL, P., & COHEN, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, **20**, 487-506.
- GALFANO, G., RUSCONI, E., & UMILTÀ, C. (2003). Automatic activation of multiplication facts: Evidence from the nodes adjacent to the product. *Quarterly Journal of Experimental Psychology*, **56A**, 31-61.
- GEARY, D. C., WIDAMAN, K. F., & LITTLE, T. D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition*, **14**, 478-487.
- GROEN, G. J., & PARKMAN, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329-343.
- GROSSBERG, S. (1980). How does the brain build a cognitive code? *Psychological Review*, **87**, 1-51.
- HECHT, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition*, **27**, 1097-1107.
- KIRK, E. P., & ASHCRAFT, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 157-175.
- KOSHMIDER, J. W., & ASHCRAFT, M. H. (1991). The development of children's mental multiplication skills. *Journal of Experimental Child Psychology*, **51**, 53-89.
- LEFEVRE, J., BISANZ, J., DALEY, K. E., BUFFONE, L., GREENHAM, S. L., & SADESKY, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, **125**, 284-306.
- LEFEVRE, J., & LIU, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from China and Canada. *Mathematical Cognition*, **3**, 31-62.
- LEFEVRE, J., SADESKY, G. S., & BISANZ, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 216-230.
- LEMAIRE, P., & SIEGLER, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, **124**, 83-97.
- MAURO, D. G., LEFEVRE, J., & MORRIS, J. (2003). Effects of problem format on division and multiplication performance: Division facts are mediated by multiplication-based representations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 163-170.
- MCCLOSKEY, M., ALIMINOSA, D., & SOKOL, S. M. (1991). Facts, rules, and procedures in normal calculation: Evidence from multiple single-patient studies of impaired arithmetic fact retrieval. *Brain & Cognition*, **17**, 154-203.
- MCCLOSKEY, M., HARLEY, W., & SOKOL, S. M. (1991). Models of arithmetic fact retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 377-397.
- MCCLOSKEY, M., & LINDEMANN, A. M. (1992). Mathnet: Preliminary results from a distributed model of arithmetic fact retrieval. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 365-409). Amsterdam: Elsevier.
- NIEDEGGEN, M., & RÖSLER, F. (1999). N400 effects reflect activation spread during retrieval of arithmetic facts. *Psychological Science*, **10**, 271-276.
- NOËL, M.-P., FIAS, W., & BRYBAERT, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition*, **63**, 335-374.
- 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.
- PAULI, P., BOURNE, L. E., JR., & BIRBAUMER, N. (1998). Extensive practice in mental arithmetic and practice transfer over a ten-month retention interval. *Mathematical Cognition*, **4**, 21-46.
- PLAUT, D. C., MCCLELLAND, J. L., SEIDENBERG, M. S., & PATTERSON, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.
- RATINCKX, E., BRYBAERT, M., & FIAS, W. (2005). *Naming two-digit Arabic numerals: Evidence from masked priming studies*. Manuscript submitted for publication.
- RICKARD, T. C., & BOURNE, L. E., JR. (1996). Some tests of an identical elements model of basic arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1281-1295.
- RICKARD, T. C., HEALY, A. F., & BOURNE, L. E., JR. (1994). On the representation of arithmetic facts: Operand order, symbol, and operation transfer effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1139-1153.
- SEIDENBERG, M. S., & MCCLELLAND, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.
- SIEGLER, R. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, **117**, 258-275.
- SOKOL, S. M., MCCLOSKEY, M., COHEN, N. J., & ALIMINOSA, D. (1991). Cognitive representations and processes in arithmetic: Inferences from the performance of brain-damaged patients. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 355-376.
- STAZYK, E. H., ASHCRAFT, M. H., & HAMANN, M. S. (1982). A network approach to mental multiplication. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 320-335.
- THELEN, E., SCHÖNER, G., SCHEIER, C., & SMITH, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral & Brain Sciences*, **24**, 1-86.
- THIBODEAU, M. H., LEFEVRE, J.-A., & BISANZ, J. (1996). The extension of the interference effect to multiplication. *Canadian Journal of Experimental Psychology*, **50**, 393-396.
- WHALEN, J., MCCLOSKEY, M., LINDEMANN, M., & BOUTON, G. (2002). Representing arithmetic table facts in memory: Evidence from acquired impairments. *Cognitive Neuropsychology*, **19**, 505-522.
- WINKELMAN, J. H., & SCHMIDT, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology*, **102**, 734-736.
- ZBRODOFF, N. J., & LOGAN, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 83-97.

NOTES

1. We discuss only the neighbors that have a one-unit difference from one of the operands, since these are the most strongly activated, besides the correct problem node itself.
2. We did not use Campbell and Xue's (2001) data for explicit comparison with our model, for a number of reasons: The sample size was smaller than that in Campbell (1994), the standard set of multiplication problems was not used, and detailed error type data were not available to us.

APPENDIX

Here, we will describe the exact equations that govern the flow of activation in the network. All equations have the form

$$\frac{d}{dt}q(t) = \text{decay} + \text{excitation} - \text{inhibition}$$

for the activation $q(t)$ of an arbitrary unit. Equations of this type were proposed in many earlier cognitive models (e.g., Grossberg, 1980; Thelen, Schöner, Scheier, & Smith, 2001). For each of the two input fields, for an input unit i at time t , activation x_i of that unit at time t is described by

$$\frac{d}{dt}x_i(t) = -x_i(t) + I_i(t)[B - x_i(t)]. \quad (\text{A1})$$

In Equation A1, $-x_i(t)$ functions as a decay term. The factor $I_i(t)$ is an indicator function which, at time t , equals one if number i was presented in its input field and zero otherwise. The parameter B is an upper bound to the activation values x_i . Hence, the combination of the decay term and the $I_i(t)[B - x_i(t)]$ terms restricts x_i between zero and B .

The activation y_i of a semantic unit i has the following equation:

$$\frac{d}{dt}y_i(t) = -y_i(t) + \sum_j w_{ji}^{\text{in}} x_j (B - y_i). \quad (\text{A2})$$

In Equation A2, the sum is taken over all the nodes in both input fields. Again, the factor $B - y_i$ bounds the equation below B . The equation is also bounded above zero: For example, in the absence of input, the equation becomes $d/dt y_i(t) = -y_i(t)$, and y decays exponentially toward zero.

The equation describing activation z_j^T of unit j in the decade field (with superscript T for tens, or decades) is

$$\frac{d}{dt}z_i^T(t) = -z_i^T(t) + M \sum_j w_{ji}^T y_j (B - z_j^T) - z_i^T \sum_{j \neq i} z_j^T, \quad (\text{A3})$$

and similarly, for the unit field, with superscript U for units,

$$\frac{d}{dt}z_i^U(t) = -z_i^U(t) + M \sum_j w_{ji}^U y_j (B - z_j^U) - z_i^U \sum_{j \neq i} z_j^U. \quad (\text{A4})$$

The terms

$$z_i^T \sum_{j \neq i} z_j^T \text{ and } z_i^U \sum_{j \neq i} z_j^U$$

reflect competition within the decade and the unit fields, respectively. These terms indicate that all other nodes ($j \neq i$) inhibit node i equally. This inhibition is multiplied by the activation of the node i itself (z_i^T or z_i^U), so the activation remains bounded above zero.

The weights w_{ji}^T and w_{ji}^U are zero or one, depending on whether node j should feed into decomposition field unit i or not. For example, the weight from the 24 node in the semantic field to the 20s node in the decade field is one, to the 30s node it is zero, and so on.

In the response field, for node i with activation $r_i(t)$,

$$\frac{d}{dt}r_i(t) = -r_i(t) + \sum_j w_{ji}^R z_j (B - r_i). \quad (\text{A5})$$

The weights w_{ji}^R are zero or S , again depending on whether the connection between the decomposition (decade or unit) node and the response node should be formed or not. The parameter S is a scaling parameter that influences the overall speed of responding in the network.

In addition to these parameters, there is a threshold that has to be reached before a response is executed, which is denoted τ . To sum up, the parameters used in the model are B , M , S , and τ , in addition to the parameters α and C described in the text. The first four of these are scaling parameters, in the sense that changing their values changes the simulated RTs (e.g., RTs are generally longer with a higher threshold τ) but does not change the qualitative pattern of predictions. The values of these scaling parameters used in the simulations were $B = 20$, $M = 10$, $S = 2$, and $\tau = 19.5$.