Memory & Cognition
2002, 30 (1), 97-106

A multidimensional scaling approach to mental multiplication

THOMAS L. GRIFFITHS and MICHAEL L. KALISH
University of Western Australia, Nedlands, Australia

Adults consistently make errors in solving simple multiplication problems. These errors have been explained with reference to the interference between similar problems. In this paper, we apply multidimensional scaling (MDS) to the domain of multiplication problems, to uncover their underlying similarity structure. A tree-sorting task was used to obtain perceived dissimilarity ratings. The derived representation shows greater similarity between problems containing larger operands and suggests that *tie* problems (e.g., 7 × 7) hold special status. A version of the generalized context model (Nosofsky, 1986) was used to explore the derived MDS solution. The similarity of multiplication problems made an important contribution to producing a model consistent with human performance, as did the frequency with which such problems arise in textbooks, suggesting that both factors may be involved in the explanation of errors.

Solving simple multiplication problems is an important part of the mathematics education of schoolchildren. By adulthood, individuals have developed sufficient expertise at solving multiplication problems to be capable of fast, accurate responses. Nevertheless, even adults produce consistent patterns of errors on simple problems (Campbell, 1994; Campbell & Graham, 1985). Explanations for these errors have focused on the contributions of the similarity of multiplication problems (Campbell, 1995; Campbell & Graham, 1985) and the frequency with which problems are encountered (Ashcraft & Christy, 1995). In this paper, we apply techniques from multidimensional scaling (MDS) and computational modeling to investigate the contributions of these factors to performance on mental multiplication tasks.

Patterns of Errors in Mental Multiplication

Performance on simple multiplication problems has been widely studied, mainly using the *production* task (e.g., Campbell & Tarling, 1996). This task involves visually presenting a problem, commonly between 2 × 2 and 9 × 9, and eliciting a verbal response. Dependent variables include error rates and response times, which are positively correlated (Campbell & Graham, 1985). The errors produced on simple multiplication problems show several trends, involving both the absolute error rate and the type of errors observed.

Absolute error rates. One common finding is that error rate increases as a function of the numerical magnitude of arithmetic problems (Geary, Widaman, & Little, 1986; Parkman & Groen, 1971; Stazyk, Ashcraft, & Hamann, 1982). Thus, large problems, such as 9 × 7, tend to have a higher error rate than do smaller problems, such as 3 × 4, a phenomenon termed the *problem size effect.* Problem size does not affect performance on all problems equally. The *tie problems*, with equal operands (e.g., 7 × 7), are less affected by problem size (Miller, Perlmutter, & Keating, 1984; Parkman, 1972), as are problems with five as an operand (Campbell, 1994).

The classification of errors. The errors made on simple multiplication problems differ in type as well as in frequency. Campbell and Graham (1985) identified three common types of error: *operand errors*, *table errors*, and *nontable errors.* An operand error is a response appropriate to one of the operands of the problem, but not to the other, such as responding *28* to 6 × 4, where 28 is the correct answer to 7 × 4. Table errors occur when the response is the correct solution to another simple multiplication problem with no shared operands, such as responding *27* to 6 × 4, where 27 is the correct answer to 9 × 3. Nontable errors are responses that are not the product of any of the operands in the problem set, such as *34* and *22.* Campbell and Graham had 60 adult participants complete 144 multiplication problems each and found an overall error rate of 7.6%. Of the errors reported, 79% were operand errors, 14% table errors, and 7% nontable errors. Other error types have also been identified, such as operand intrusions, where one of the operands of a problem appears in a response, and interoperation confusions, where the wrong operation is applied to a pair of operands (Campbell, 1994, 1997). How-

This research was supported by a Jean Rogerson Honors Studentship to the first author and by ARC Large Grant A79941108 to the second author. Jamie Campbell provided both insightful comments and the original computer program for the network interference model, for which the authors are grateful. For helpful comments on previous versions of this article, we thank Mark Ashcraft, Simon Farrell, Murray Maybery, and an anonymous reviewer. Correspondence concerning this article should be addressed to M. L. Kalish, Department of Psychology, University of Western Australia, Nedlands 6009, Australia (e-mail: kalish@psy.uwa.edu.au).

—Accepted by previous editorial team

ever, the problem size effect and the relative frequency of operand, table, and nontable errors serve as the basic phenomena that theories of mental multiplication have attempted to explain.

Theories of Mental Multiplication

Research concerning the performance of adults on simple arithmetic problems has found that the task is heavily reliant on retrieval from memory (Ashcraft, 1985; Koshmider & Ashcraft, 1991). Explanations of the errors made by adults on simple arithmetic problems have focused on the properties of this retrieval process (for reviews, see Ashcraft, 1992; Dehaene, 1992; McCloskey, Harley, & Sokol, 1991). Two of the most prominent theories have referred to interference between problems (Campbell, 1995; Campbell & Graham, 1985), and the strength of problem–solution pairs in memory (Ashcraft & Christy, 1995).

Interference. Campbell and Graham (1985) suggested that the errors observed in mental multiplication are due to interference between similar problems in memory. This theory has been formalized as a computational model, called the network interference model (Campbell, 1995; Campbell & Oliphant, 1992). The network interference model predicts the problem size effect because it assumes that similarity between problems increases as a function of the size of the answer. The model accounts for the high frequency of operand errors through the further assumption that problems sharing operands are similar.

There is some support for an effect of interference in mental multiplication. Campbell (1987) found that the time taken to formulate a correct response to a multiplication problem was related to how recently similar problems had been presented. Consistent with an interference effect, participants showed longer response times after the presentation of a number of similar problems. Graham and Campbell (1992) carried out a set of experiments using an "alphaplication" task, in which participants had to perform arithmetic with letters instead of digits, and found similar error patterns to multiplication. Graham and Campbell claimed that this was because the tasks share a common similarity structure and are thus equally affected by interference.

Strength. Ashcraft and Christy (1995) argued that the problem size effect might be due to the strength of association between problems and solutions. Assuming that repeated presentation of a stimulus produces a stronger memory trace, Ashcraft and Christy examined the frequency with which different multiplication problems appear in school math texts. They found a bias toward the presentation of small problems and suggested that the higher error rate shown with large multiplication problems may be a consequence of less exposure.

The low frequency of large multiplication problems may extend beyond school math texts. In any naturally occurring set of integers, large numbers are far less likely than small numbers, a relationship termed Benford's Law (Benford, 1938). This suggests that multiplication problems with large operands are less likely to be encountered in everyday experience. Fendrich, Healy, and Bourne (1993) found that giving participants equal practice on all problems reduced the problem size effect, again suggesting that problem frequency has an effect on performance.

Integrated theories. Empirical evidence suggests that both interference and strength affect the production of responses to simple multiplication problems. Accounts integrating these two factors have also been offered. One example of an integrated approach is that taken by Zbrodoff (1995), who explored the relationship between interference and strength in the explanation of the problem size effect in addition. In a series of experiments, Zbrodoff investigated the idea that addition problems interfere with one another and that problems with a stronger memory trace are more resistant to this interference. Participants were trained to do alphabet addition, similar to the alphaplication task used by Graham and Campbell (1992). Differential problem frequencies produced the problem size effect only at low levels of practice, and not after training was complete. Interference between problems produced the problem size effect in final performance, but not under conditions in which the problems had equal frequency. Zbrodoff concluded that interference and strength interact to produce the problem size effect in addition.

Exploring the Similarity Structure of Multiplication Problems

The network interference model predicts the errors in mental multiplication from the similarity between multiplication problems. Implementing this model requires specifying a similarity structure for multiplication problems. Campbell (1995) offered one such structure, in which the key parameters describing problem similarity "were estimated by examining the normative frequencies of specific errors" (p. 131). This approach involves making untested assumptions about the representation of multiplication problems, a fact that has been taken as a weakness of the network interference model (LeFevre et al., 1996). The aim of this paper is to address this issue by providing an empirical investigation of the similarity structure of multiplication problems.

Campbell's (1995; Campbell & Oliphant, 1992) network interference model postulates two aspects of similarity between multiplication problems: physical similarity and magnitude similarity. Physical similarity is the weighted sum of matches between the operator (i.e., whether the two problems both contain a multiplication sign), the operands, and the decades and units of the response. Magnitude similarity is a decreasing function of the difference in the solutions of the two problems, relative to the size of the larger solution. The total similarity is then the sum of the physical similarity and the magnitude similarity. If one of the problems involved is a tie or has five as an operand, the total similarity is multiplied by a further parameter to reduce the similarity between the problems.

Campbell's (1995) representational assumptions provide a clear specification of the similarity structure to which multiplication problems should adhere. Problems with matching operands should be more similar, the similarity between two problems should decrease as the difference in their solutions increases, and problems involving ties and fives should be especially similar to one another. These assumptions can be examined with techniques designed to extract representational information from similarity ratings.

Multidimensional Scaling and the Tree-Sorting Procedure

MDS is often used to provide insight into mental representations, especially in situations in which the stimuli are discrete, such as single-digit multiplication (e.g., Nosofsky, 1986; Shepard, 1980). Participants rate the perceived similarity of different stimuli, and this information is converted into a low-dimensional spatial representation through the assumption that similarity is inversely related to distance (Kruskal & Wish, 1978; Shepard, 1987). The properties of the derived representation can then help in understanding the psychological similarity structure of the stimuli.

The use of MDS techniques in psychological research is limited by the difficulty of collecting similarity ratings for large sets of stimuli. The similarity ratings provided to the MDS algorithm should ideally include one value for every possible pair of stimuli. Since these values are often assumed to be independent of the order of stimuli in the pair, an MDS solution featuring n stimuli requires $[n \times (n-1)]/2$ similarity ratings. This number rapidly becomes too large to be obtained conveniently: To scale the 64 multiplication problems between 2 × 2 and 9 × 9, 2,016 similarity ratings are needed.

Fortunately, techniques exist that allow participants to make only a small number of judgments and still produce a complete set of similarity ratings. The tree-sorting task (Fillenbaum & Rapoport, 1971) is one such technique, in which the close correspondence between multidimensional scaling and spanning trees is used (Shepard, 1980). Each participant follows a set of simple instructions (presented below) to form the stimuli into a *tree*, a structure that places connections between stimuli to form a unique route from each stimulus to every other. The placement of connections corresponds to the perceived similarity between stimuli, so that very similar items can be reached by traversing only a few connections. The number of connections that needs to be traversed to travel between stimuli is the *distance* between them, expressing their degree of dissimilarity. By averaging across participants, it is possible to produce distances that reflect the common trends in a set of trees: Stimuli consistently placed far apart will have large average distances, whereas those commonly placed together will have small average distances.

The tree-sorting task was originally used to collect similarity judgments for the development of semantic representations (Fillenbaum & Rapoport, 1971). It has subsequently been applied to face perception (Rhodes, 1985) and is recognized as an efficient means of dealing with large sets of stimuli (Coxon, 1982). However, few data exist on the validity or the reliability of the results derived through this technique. In Experiment 1, we examined the reliability and validity of the tree-sorting task. In Experiment 2, we then used the tree-sorting task to assess the similarity structure of multiplication problems.

EXPERIMENT 1
The Reliability and Validity of the Tree-Sorting Task

One of the classic sets of stimuli to which MDS techniques have been applied is the set of integers between 0 and 9. Shepard, Kilpatric, and Cunningham (1975) found that the MDS solution for this stimulus set could be divided along axes corresponding to parity and magnitude. This result has been confirmed through a variety of experimental procedures (Lewandowsky & Newman, 1993; Miller, 1992; Miller & Gelman, 1983). Here, we examine the reliability and validity of the tree-sorting task using these stimuli.

Method

Participants. Two groups of 20 undergraduate psychology students from the University of Western Australia participated for partial course credit.

Materials. The 10 integers between 0 and 9 were printed on cards, 9 cm in width and 5.5 cm in height. Each number was printed in 72-point Times in black, on a white background.

Procedure. The participants were tested individually. Each participant performed a number of mathematical tasks, including multiplication, factorization, computing squares, and making magnitude comparisons, and then went on to perform the similarity rating task. The participants were given the following instructions for the tree-sorting task:

> From the set of 10 numbers, pick the two numbers which you think are *most similar* to each other. Take these two cards and place them next to each other on the table. Now you have two options:
>
> *Option 1*: You may go carefully over the remaining numbers (of which there are now 8) and pick the number which you think is most similar to *either* of the *two* numbers you have already selected. Move this card and put it next to the one to which you believe it is similar.
>
> *Option 2*: You may look over the remaining numbers and decide that two of them are more similar to each other than any one of them is to either of the two numbers already selected. If so, you may select these numbers and place their cards next to one another, just as you did with the first pair.
>
> After taking Option 1 or 2, proceed in exactly the same way. Search over the remaining numbers and choose Option 1 or Option 2. When you take Option 1 you add a number to an already linked group of numbers (which is called a tree). When you take Option 2 you start a new tree.
>
> As the experiment proceeds, a new option becomes available:
>
> *Option 3*: If you find that you have made several trees, you may want to connect any two of them together. If you find two numbers, on two separate trees, that are more similar to each other than any other two numbers are to each other, you should move these cards next to each other, linking the trees.

The participants were directed to use these options repeatedly, until all 10 cards were used and a single tree remained. The 10 cards were then shuffled and placed in front of the participant, who formed them into a tree so that each integer could be reached from every other integer. The experimenter noted the order of the connections.

Results and Discussion

Inspection of the data showed that no participants formed linear trees determined solely by magnitude. A triangular 10 × 10 distance matrix was formed for each participant by summing the ordered connections between all the items. The matrices thus derived were then averaged to form an aggregate distance matrix for each group. In forming the aggregate matrices, it was observed that the means and variances of the distances were related. The relationship between the logarithms of the means and the standard deviations was examined and found to be linear. Rhodes (1985) obtained a similar result, using the tree-sorting task, and transformed the data so that $y = x^{(1-b)}$, where b is the slope of the regression line. The regression of log (SD) on log (M) yielded $r(43) = .10$, where $b = 0.10$ for the first group, and $r(43) = .75$, $b = 0.60$ for the second. Following the transformation, $r^2(43) = .003$ and $r^2(43) = .0001$, respectively.

The average distances for the two groups were correlated, yielding $r(43) = .80$. Applying the Spearman–Brown correction for split-half reliability, the corrected reliability for the whole set is $r = .89$. The high correlation between the derived distances supports the reliability of the tree-sorting task. The aggregate matrix for the first group was supplied to the ALSCAL algorithm. Stimulus configurations above three dimensions incorporated more free parameters than the data had degrees of freedom and were not examined. The two-dimensional solution provided clearly interpretable dimensions corresponding to magnitude and parity, with a stress of .072. The resulting representation thus had the same properties as those found in previous research (Lewandowsky & Newman, 1993; Miller, 1992; Miller & Gelman, 1983; Shepard et al., 1975), supporting the validity of the tree-sorting task.

EXPERIMENT 2
Deriving a Similarity Structure for Multiplication Problems

Method

Participants. The participants were 20 undergraduate psychology students from the University of Western Australia, participating for partial course credit.

Materials. The 64 problems between 2 × 2 and 9 × 9 were printed on cards, 9 cm in width and 5.5 cm in height, with the same properties as the cards in Experiment 1.

Procedure. The participants were tested individually. Each participant was asked to go through the stack of randomly shuffled cards and say aloud the answer to each problem. They then received instructions for the tree-sorting task, emphasizing the importance of rating the abstract similarity of the problems, rather than their physical resemblance. The cards were shuffled again, and 16 cards were selected. The participants performed the tree-sorting task on this reduced stimulus set, as a means of illustrating the demands of the task. The cards were collected, shuffled back into the deck, and the full set of cards was spread out on a table in an 8 × 8 grid. Following the rules of the task, the participants arranged the stimuli according to the similarity between them: Pairs of similar items were selected until all the stimuli were linked, and the order of choice was recorded.

Results and Discussion

A triangular 64 × 64 distance matrix was formed for each subject by summing the ordered connections between all the items. The matrices thus derived were then averaged to form an aggregate distance matrix. The regression of log (SD) on log (M) yielded $b = 0.53$, with $r(2{,}014) = .56$. The exponential transformation used in Experiment 1 produced $r^2(2{,}014) = .03$. In order to check the reliability of the data, the 20 matrices were randomly split into two

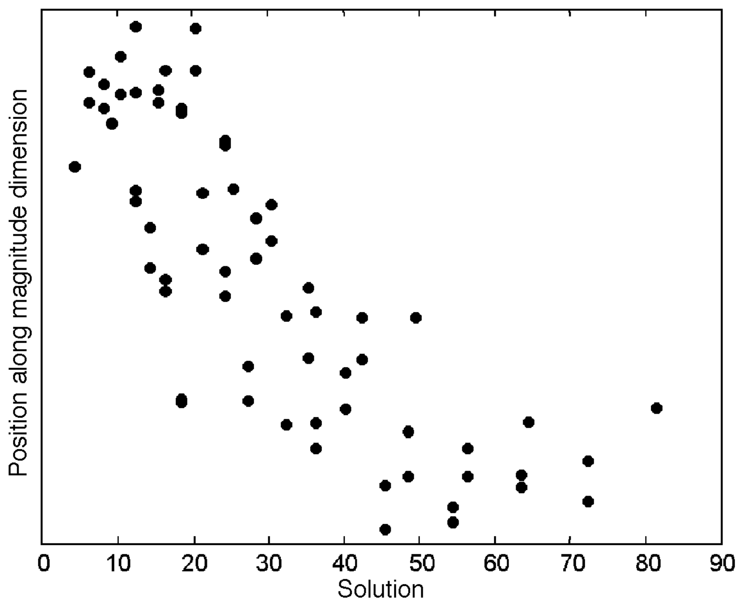

Figure 1. Problem locations on the first dimension of the five-dimensional multidimensional scaling solution found in Experiment 2, as a function of problem solution. This dimension appears to correspond to solution magnitude.

groups, and group aggregates were derived. The aggregate matrices were correlated against each other, yielding $r(2{,}014) = .55$. Applying the Spearman–Brown correction gave a final $r(2{,}014) = .67$. Although low, this value is comparable with similar indices reported in other research using this method (e.g., Rhodes, 1985).[1]

The aggregate matrix was supplied to the ALSCAL algorithm to derive an MDS representation of the stimuli. Stimulus configurations were derived for between two and six dimensions. The stress values were 0.24, 0.17, 0.13, 0.10, and 0.09, for two to six dimensions, respectively, showing no points of inflection. Since our primary concern was with the similarity of problems that share an operand, we chose the representation that best preserved the mean distance between the members of each operand family. These mean distances were computed for the raw data and for each MDS solution. The mean distances in the five-dimensional solution best matched the raw distances, yielding $r(6) = .92$. The first dimension of the derived solution was related to the magnitude of the solution to the problem [$r(62) = .81, p < .0001$], as can be seen in Figure 1. The other four dimensions were involved in

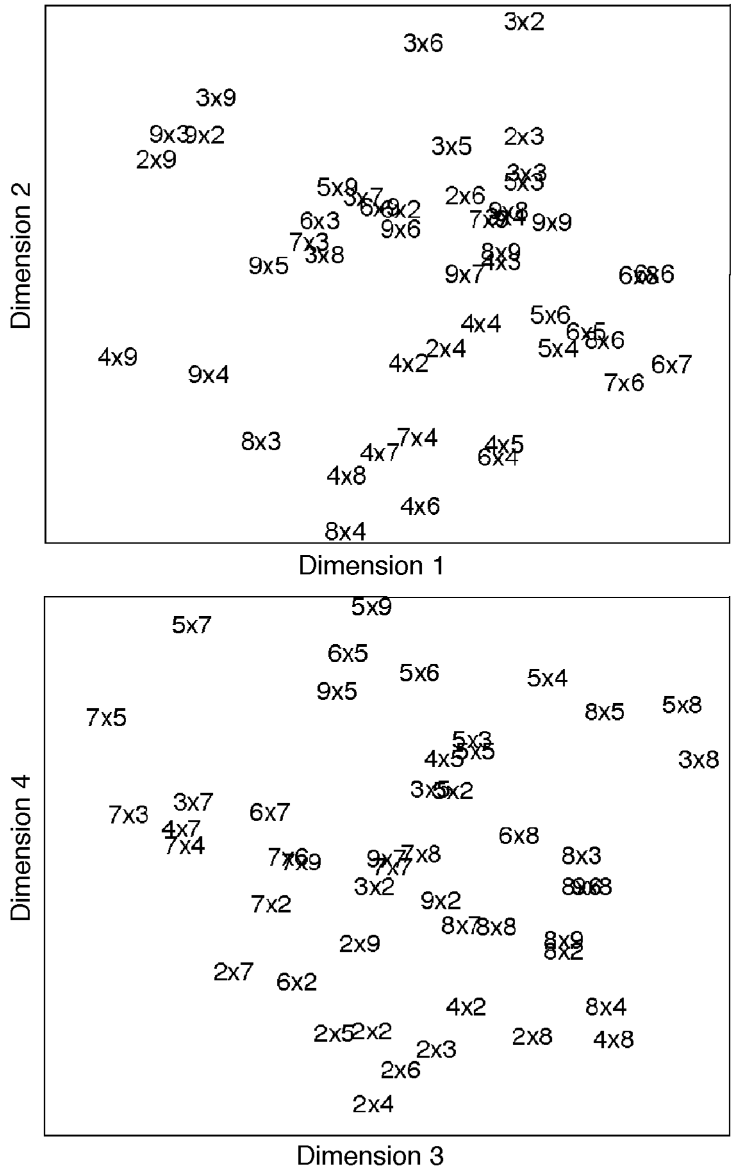

Figure 2. The remaining four dimensions of the multidimensional scaling solution derived in Experiment 2. To illustrate the way in which they code for problem operands, the dimensions were rotated to maximize the mean of each operand family along one dimension and minimize it along the other three while retaining orthogonality. Only those problems separated by the axes of each graph are displayed.

coarsely coding the operands of the problems. Figure 2 shows that these dimensions separate problems containing 9 as an operand from those containing 6 as an operand, 4 from 3, 7 from 8, and 2 from 5, respectively.

To aid in the interpretation of the five-dimensional solution, a linear model of the set of derived distances was generated, with factors coding for particular characteristics of the relationship between problems. The factors were *commute*, coding for commutative pairs (e.g., 3 × 4 and 4 × 3); *tie*, coding for whether the two problems were both ties; *magnitude*, the difference in magnitudes between the two problems; and *two* to *nine*, coding for the presence of a particular operand in both problems. A dummy coding scheme was used for all binary factors, in which each predictor was assigned a value of 1 if a pair of problems possessed the target attribute and 0 otherwise. The resulting parameter values are reported in Table 1. The regression model gave $R(11, 2004) = .62, p < .0001$.[2]

The parameters of the linear model increase slightly for factors *two* to *four*, then decrease sharply. This trend can be reflected by correlating each parameter against the magnitude of the operand family to which it refers [$r(6) = .77$ for all eight operands and $r(4) = .97$ for the last six].[3] The negative parameters observed for all the operands support increased similarity between problems sharing operands. However, Campbell's (1995) claim that problems with 5 as an operand should show the greatest similarity to one another is not supported by the data. In fact, similarity gradually decreases for problems featuring 2, 3, or 4 as an operand and then increases sharply as a function of operand magnitude.[4]

The parameters of the linear model also suggest that commutative pairs are closer together than would be expected on the basis of a match between their operands alone. This is consistent with Campbell's (1995) notion of magnitude similarity, which is maximized for commutative pairs. The magnitude of the difference between solutions contributes to the distance between problems, large distances being associated with problems being further apart, a result that is also consistent with the basic idea of magnitude similarity in the network interference model. Finally, the derived representation supports the assertion that tie problems should be categorically distinct from nonties, showing greater similarity to other tie problems than to nontie problems. The negative *ties* parameter in the linear model is consistent with tie problems being separated by a lower than average distance.

Table 1
Parameter Values for the Linear Model of the Distances Between Problems

Factor	Slope Parameter	Standard Error
(Intercept)	2.935	0.034
Commute	−0.638	0.150
Tie	−1.417	0.140
Two	−0.536	0.078
Three	−0.445	0.077
Four	−0.201	0.077
Five	−0.262	0.076
Six	−0.430	0.075
Seven	−0.951	0.075
Eight	−0.982	0.075
Nine	−1.188	0.075
Magnitude	0.019	0.001

Note—The mean distance between problems was 3.05. All factors were statistically significant at $\alpha = .01$.

MODELING THE EFFECTS OF SIMILARITY AND FREQUENCY

The MDS solution found in Experiment 2 shows a qualitative correspondence to the assumptions of Campbell (1995). The MDS approach pursued in this paper also provides the opportunity for the exploration of the quantitative properties of this representation. The question of how spatial representations can be mapped onto behavior has been thoroughly investigated in the literature of cognitive psychology (e.g., Shepard, 1987), resulting in models such as Nosofsky's (1986, 1991a) generalized context model (GCM). The GCM has been applied to a number of cognitive domains, including stimulus identification (Nosofsky, 1986), categorization (Nosofsky, 1987), recognition memory (Nosofsky, 1991b), and category learning (Kruschke, 1992). Use of the GCM to predict errors in mental multiplication provides the opportunity to gain further insight into the properties of the derived representation.

The simplest version of the GCM takes the similarity between stimuli i and j, η_{ij}, to be an exponentially decreasing function of the Euclidean distance between i and j, $\eta_{ij} = \exp(-c\, d_{ij})$, where c is the specificity of the similarity function. The distances, d_{ij}, are computed from a multidimensional representation in which the weight of each dimension is set by a parameter w_k, with the constraint that $\Sigma w_k = 1$. The response selection process uses these similarity values, although the exact procedure varies according to the task being modeled. When there is a single response for each stimulus, as in mental multiplication, the activation of the jth response on a trial in which stimulus i is presented, a_j, is simply the similarity between i and j, η_{ij}. These activations are converted into response probabilities via a choice rule. Kruschke (1992) endorses an exponentiated version of Luce's (1963) choice rule, taking the form

$$P(R_j \mid S_i) = \frac{\exp(\Phi b_j a_j)}{\sum_k \exp(\Phi b_k a_k)}, \tag{1}$$

where $P(R_j \mid S_i)$ is the probability of response j given stimulus i, Φ is a parameter setting the sensitivity of the response selection process to similarity values and b_j is the bias associated with response j.

The b_j parameters, which describe the bias toward a particular response, have a natural interpretation in the context of mental multiplication. Since these parameters are multipliers for the activation of problems, they determine the extent to which each problem is resistant to interference. This is essentially the role of strength, as described by

Ashcraft and Christy (1995) and by Zbrodoff's (1995) formulation of the interaction between strength and interference. The relationship between problem strength and frequency of presentation can be incorporated into the model by making the b_j parameters proportional to problem frequency.

The use of the GCM thus provides the opportunity to examine the relative contributions of problem similarity, expressed in terms of the MDS solution derived in Experiment 2, and problem frequency, reflected in the b_j parameters. Models that differ in their parameters also allow the assessment of whether each parameter makes a statistically significant contribution to the fit of the model to data, just as in multiple linear regression. Examining the contributions of the parameters associated with similarity and frequency can thus help us gain insight into the role of these factors in producing errors.

In order to test our statistical models, we require a set of data to which the models can be fit and an appropriate metric of model performance. Campbell (1994) reported frequencies of integer responses to multiplication problems between 2×2 and 9×9, presented either as digits (e.g., 2×2) or as words (e.g., two × two). These frequencies came from 64 participants giving four responses to nontie problems and eight responses to ties. We chose to use the data from problems presented as digits, since these were the stimuli for the MDS procedure. These data demonstrate the two major phenomena of mental multiplication performance: the problem size effect and the classification of errors.

The problem size effect and the classification of errors can be captured by collapsing the set of response frequencies into four categories for each problem, representing correct responses, operand errors, table errors, and nontable errors. This forms a 64×4 contingency table, the 64 rows corresponding to the 64 problems between 2×2 and 9×9 and the four columns representing the category of response. Given such a table, G^2 is a suitable lack-of-fit statistic (Wickens, 1989). However, the use of multiple responses gathered from participants in a single testing session violates the assumption of independence underlying χ^2 tests, given the observation of practice effects and response priming in solving multiplication problems (Meagher & Campbell, 1995; Rickard, Healy, & Bourne, 1994; Stazyk et al., 1982). Wickens (1989) recommends the use of G^2/k for this situation, where k represents the number of scores in each related group. Thus, the frequencies were divided by the number of responses given by each participant for each problem, four for nonties and eight for ties. The error columns also contain a large number of zero values, which can reduce the accuracy of the G^2 statistic. Since most small values occur in the nontable errors column, this column was removed from the contingency table, reducing it to 64×3. The contingency table still contained 41 cells with zero values. In order to avoid a disproportionate effect of these cells on the results, a value of 0.01 was added to all cells, chosen to be less than one tenth the size of the smallest observed frequency. All reported G^2 and χ^2 values have $N = 4{,}086.8$.

Four models were fit to this data. The first model provided a baseline for assessing the contributions of other factors, taking the similarity between problems to be equal and setting the b_j to be uniform. The second and third models assessed the effects of similarity and problem frequency independently, whereas the fourth model examined the consequences of including both these factors. The optimal parameter values and fits for these models are given in Table 2. In addition to the G^2 value, the table reports the result of a χ^2 test for homogeneity conducted on the model predictions, which reflects the strength of the problem size effect in these data, and the relative proportions of correct responses, operand errors, and table errors. The final column shows these same quantities computed directly from the modified contingency table representing the data.

As might be expected, the baseline model gave a poor fit to the data, failing to capture the appropriate frequencies of errors and showing extremely homogenous predictions across problems. Introducing the effect of problem frequency to this model resulted in little improvement, although the predictions did show a weak problem size effect, reflected in the higher χ^2 score. The effect of problem frequency was implemented by setting the b_j parameters, using frequencies from Ashcraft and Christy's (1995) assessment of elementary school textbooks, summing over all grades. The b_j values were generated by normalizing $b_j = f_j^{\gamma}$, where f_j is the raw frequency of the jth problem and γ is a free parameter that was optimized. The introduction of problem frequency in this fashion resulted in a significant improvement in the fit of the model over baseline [$G^2(1) = 42.38, p < .01$].

The effect of similarity was examined in a model with the b_j uniform, but with distances determined by the MDS solution derived in Experiment 2. This required the introduction of six new parameters to the baseline model: the c parameter and the weights for the five dimensions of the space, w_k. The resulting model gave better predictions of the relative frequencies of errors, as well as showing evidence of a weak problem size effect, reflected in the higher χ^2 value. The introduction of these additional parameters resulted in a significant improvement in fit over the baseline model [$G^2(5) = 68.84, p < .01$].

Table 2
Parameter Values and Fit Statistics for Variants of the Generalized Context Model

Parameters and Fit Statistics	Baseline	Factors in Model: Frequency	Similarity	Both	(Data)
Φ	751.51	767.35	569.97	551.78	–
γ	–	0.117	–	0.153	–
c	–	–	4.39	5.07	–
w_1	–	–	0.32	0.33	–
w_2	–	–	0.23	0.28	–
w_3	–	–	0.06	0.09	–
w_4	–	–	0.18	0.17	–
G^2	286.79	244.41	207.95	186.24	–
df	127	126	122	121	
$\chi^2(126)$	2.73	54.28	75.71	117.65	201.29
P(Correct)	.965	.965	.964	.964	.965
P(Operand)	.014	.013	.023	.022	.030
P(Table)	.021	.022	.013	.014	.005

Finally, the model incorporating both similarity determined by the MDS solution and problem frequency in the response biases gave the best overall account of the data. The predictions of this model accounted for error frequencies in a fashion comparable with similarity alone but showed a problem size effect much stronger than either of the models incorporating just one of these factors. The introduction of the parameters implementing the effects of similarity resulted in a significant improvement in fit over the model with frequency alone [$G^2(5) = 58.17, p < .01$], and the introduction of the effect of frequency resulted in a significant improvement in fit over the model with similarity alone [$G^2(1) = 21.71, p < .01$].

The results of the model fitting suggest that similarity and problem frequency make complementary contributions to errors in mental multiplication. Using problem frequency to set the response biases of a model in which problems were equally similar improved the prediction of the problem size effect but gave a poor account of the relative frequency of different types of errors. Using the representation derived in Experiment 2 to determine similarity was sufficient to produce appropriate error frequencies, but only a weak problem size effect. Incorporating problem frequency into the model with the appropriate similarity structure resulted in the best overall account of the data, including the prediction of a problem size effect much stronger than that given by the models incorporating each factor alone.

GENERAL DISCUSSION

The aim of this paper was to investigate the properties of the representation used by adults in solving simple multiplication problems. One important aspect of this investigation was to address the acceptability of the representational assumptions of Campbell's (1995) network interference model. In Experiment 1, we examined the validity and reliability of the tree-sorting task, a particular method for collecting ratings for the purposes of MDS. In Experiment 2, we used the tree-sorting task to derive a spatial representation of the similarity between multiplication problems. This representation showed a reasonably strong correspondence to the kind of structure that might be expected if Campbell's (1995) assumptions are valid. In order to obtain a better understanding of the derived similarity structure, a simple model was used to evaluate the effects of similarity and problem frequency on errors. The model fitting suggested that both factors can contribute to an account of multiplication performance, with complementary effects.

The present results provide some insight into the representational structure underlying multiplication problems. It appears that the kind of representation assumed by Campbell (1995) is consistent with the results of applying MDS to similarity ratings of multiplication problems. Most strikingly, the derived solution includes dimensions that separate operand families into individual clusters in the derived space. The linear model of the derived distances supports Campbell's (1995) assumptions, indicating less distance between problems sharing operands, commutative pairs, and tie problems. The effect of response magnitude in the MDS solution is also consistent with the inclusion of magnitude similarity in the network interference model.

The model fitting suggests that the derived representation is sufficient to produce a weak problem size effect, with some influence of problem magnitude on errors. However, the extent to which error frequency deviates from homogeneity is not as great as that in the human data. This result needs to be reconciled with Campbell's (1995) finding that incorporating magnitude similarity in the network interference model was sufficient to produce the problem size effect. There are three factors that could contribute to this difference in results: the properties of the derived representation, the choice of fit metric, and the limitations of the modeling assumptions.

Two differences between the derived representation and that predicted by Campbell's (1995) account are the lack of special similarity for problems in the five operand family and the nonmonotonicity in the average distance between problems in a given operand family. Although the average distance between problems generally decreases with operand magnitude, the reversal of this trend with small operands can complicate the production of the problem size effect. Essentially, models are forced into a compromise between making too many errors with small operands and making too few with large operands, resulting in an artificial homogeneity on responses.

The predicted response frequencies resulting from the recovered similarity structure are, in part, a result of the use of a global fit metric like G^2, which forces every cell in the contingency table to make a contribution to the final fit. Our model-fitting procedure evaluates how well the model accounts for the data represented by the entire contingency table and then checks for the presence of the appropriate error frequencies and the problem size effect. In contrast, Campbell (1995) used the network interference model to reproduce selected empirical phenomena. Optimizing the current models to capture such phenomena might produce a stronger problem size effect, at the cost of accuracy in reproducing other specific error frequencies.

Finally, although the models considered provided the opportunity to implement a simple account of the interaction between similarity and problem frequency, they did not take into account several important aspects of adult performance on multiplication problems. Specifically, no attempt was made to account for response time data, and several known phenomena of cognitive arithmetic were neglected, including operand intrusions and interoperation confusions (cf. Campbell, 1994).

Beyond the direct implications for the network interference model, the present results provide information that is relevant to other theories of cognitive arithmetic. The results of the model fitting are consistent with the claims of Ashcraft and Christy (1995) about the importance of problem strength in explaining performance on mental arithmetic problems. Despite the deviations from monotonicity in the effect of operand matches and any constraints imposed by the fit metric, using frequency to modulate the effects of interference improved the fit of the model. This suggests that although similarity may contribute to the problem size effect, the magnitude of the effect may also be due to the influence of problem strength. This result parallels Zbrodoff's (1995) claim that the explanation of adult errors on addition problems involves an interaction between problem strength and interference.

Despite extensive practice, adults consistently produce errors on simple multiplication problems. One prominent explanation for these errors rests on the similarity structure of multiplication problems, formalized in Campbell's (1995) network interference model. In this paper, we investigated the representation underlying multiplication through the application of MDS techniques and computational modeling. The empirically derived similarity structure was largely consistent with the assumptions of the network interference model and produced appropriate frequencies of different kinds of errors. Simulations showed that this representation produced appropriate error frequencies and a weak problem size effect, although the strength of the problem size effect could be increased by allowing the effect of similarity to be modulated by problem frequency. These results provide further support for the development of integrated theories of cognitive arithmetic, focusing on the interaction between explanatory factors such as interference and problem strength.

REFERENCES

Ashcraft, M. H. (1985). Is it farfetched that some of us remember our arithmetic facts? *Journal for Research in Mathematics Education*, **16**, 99-105.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of theory and data. *Cognition*, **44**, 75-106.

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*, **26**, 396-421.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, **78**, 551-572.

Campbell, J. I. D. (1987). Network interference and mental multiplication. *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 model and simulation. *Mathematical Cognition*, **1**, 121-164.

Campbell, J. I. D. (1997). Reading-based interference in cognitive arithmetic. *Canadian Journal of Experimental Psychology*, **51**, 74-81.

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., & 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., & Tarling, D. P. M. (1996). Retrieval processes in arithmetic production and verification. *Memory & Cognition*, **24**, 156-172.

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling: With special reference to the MDS(X) library of computer programs.* Exeter, NH: Heinemann.

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

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1993). Mental arithmetic: Training and retention of multiplication skill. In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 111-133). Hillsdale, NJ: Erlbaum.

Fillenbaum, S., & Rapoport, A. (1971). *Structures in the subjective lexicon*. New York: Academic Press.

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.

Graham, D. J., & Campbell, J. I. D. (1992). Network interference and number-fact retrieval: Evidence from children's alphaplication. *Canadian Journal of Psychology*, **46**, 65-91.

Koshmider, J. W., & Ashcraft, M. H. (1991). The development of children's mental multiplication skills. *Journal of Experimental Child Psychology*, **51**, 53-89.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

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.

Lewandowsky, S., & Newman, D. A. (1993). Chronometric scaling of numbers. In R. Steyer, K. F. Wender, & K. F. Widaman (Eds.), *Psychometric methodology: Proceedings of the 7th European Meeting of the Psychometric Society in Trier* (pp. 272-277). Stuttgart: Gustav Fischer Verlag.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.

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.

Meagher, P. D., & Campbell, J. I. D. (1995). Effects of prime type and delay on multiplication priming: Evidence for a dual process model. *Quarterly Journal of Experimental Psychology*, **48A**, 801-821.

Miller, K. F. (1992). What a number is: Mathematical foundations and developing number concepts. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 3-38). Amsterdam: Elsevier.

Miller, K. [F.], & Gelman, R. (1983). The child's representation of number: A multidimensional scaling analysis. *Child Development*, **54**, 1470-1479.

Miller, K. [F.], Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 46-60.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87-109.

Nosofsky, R. M. (1991a). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-393). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1991b). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.

Parkman, J. M. (1972). Temporal aspects of simple multiplication and comparison. *Journal of Experimental Psychology*, **95**, 437-444.

Parkman, J. M., & Groen, G. J. (1971). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology*, **89**, 335-342.

Rhodes, G. (1985). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, **17**, 43-63.

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.

Shepard, R. N. (1980) Multidimensional scaling, tree-fitting, and clustering. *Science*, **210**, 390-398.

Shepard, R. N. (1987) Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, **7**, 82-138.

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.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Zbrodoff, N. J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations for the problem-size effect. *Memory & Cognition*, **23**, 689-700.

NOTES

1. As a further check of the reliability of the results, the simulations described below were run with MDS solutions derived from these split data sets. The conclusions drawn were unchanged by which solution was used.

2. The R value is a poor index of the importance of these factors because of the nature of the distance data. Points in a metric space obey the triangle inequality—the distance between two points is less than or equal to the sum of the distances of those points to a third point. This structural limitation implies that points can be close together despite having no shared characteristics, if they each share some characteristics with a third point. The regression model does not account for these relationships.

3. The authors are grateful to Jamie Campbell for making this observation.

4. This nonmonotonic change in intrafamily distances was also observable in the mean distances within the corresponding operand families. It was not a consequence of the derived representation that was selected, since it appeared in the distances for the raw data and all MDS solutions.

(Manuscript received March 29, 2001;
accepted for publication August 19, 2001.)