In Aleksander, I. & Taylor, J. (eds.) *Artificial Neural Networks II: Proceedings of the International Conference on Artificial Neural Networks,* Brighton, UK, September 4–7, 1992. Elsevier Publishers.

The Speed and Slips of Mental Arithmetic

Richard Dallaway (richardd@cogs.susx.ac.uk)

School of Cognitive and Computing Sciences, University of Sussex,
Falmer, Brighton BN1 9QH, United Kingdom

Abstract

The literature on arithmetic memory offers a number of theories for how we succeed and fail in recalling basic facts (such as $3 \times 4 = 12$). McCloskey, Harley, & Sokol [1991] have noted that these theories lack detailed implementations, and as such are difficult to appraise. This paper presents a connectionist model of memory for multiplication facts. The system's behaviour is compared to the error patterns and reaction times (RTs) of adults reported by Harley [1991] and Miller, Permutter, & Keating [1984].

Introduction

Adults exhibit well documented patterns of behaviour when recalling multiplication facts. RTs generally increase across the multiplication tables, but this "problem-size effect" has plenty of exceptions (see fig. 1). Tie problems (2×2, 3×3 etc.) are recalled relatively quickly. In addition to this Campbell & Graham [1985] found that adults make errors at the rate of 7.65 per cent, and of those errors 92.6 per cent fall into the following categories: *Operand errors*, for which the erroneous product for the problem $a \times b$ is a member of the a or b multiplication table (e.g., $6 \times 5 = 45$); *Close operand errors*, where the erroneous product is an operand error and also close in magnitude to the correct product. That is, for the problem $a \times b$, the error will often be correct for the problem $a \pm 2 \times b$ or $a \times b \pm 2$ (e.g., $6 \times 5 = 40$); *Frequent products* are the five products 12, 16, 18, 24 and 36; *Table errors*, where the erroneous product is the correct answer to some problem, but does not share any digits with the presented problem (e.g., $6 \times 5{=}16$); *Operation errors*, where the error to $a \times b$ is correct for $a + b$. The remaining 7.4 per cent of *non-table errors* were not correct products for any of the problems 0×0 to 9×9.

Five or six years of classroom drill does ensure that multiplication facts are ready to hand, but it does not ensure perfect recall. The problem is to produce a model which has correctly learnt the tables, yet can make slips when recalling answers. This aspect is missing from the other PDP model of mental arithmetic [Anderson, Spoehr, & Bennett 1991], which is based around the BSB algorithm. As presented the BSB model lacks the ability to correctly answer some problems, or fails to respond at all.

Given the observations on the types of erroneous responses, and the RT for correct responses, what assumptions must be made? The model presented here suggests that the initial skew in the frequency and order of presentation of multiplication fact [Campbell 1987, p. 118] is one of the important factors.

Architecture

A multilayered network is trained on all the problems 0×0 through 9×9 in a random order using backpropagation. The two digits that comprise a problem are coarse encoded on the two sets of 10 input units (representing the digits 0–9). For tie problems, an additional tie bit is set to 1. Activation flows through a hidden layer of 12 units, to the output layer. There is one output unit per product, plus a "don't know" unit.

During training, the presentation frequency of each pattern is skewed according to the problem frequencies reported by Siegler [1988]. Although problems with small products do occur more frequently in textbooks, there is no reason to believe this skew continues into adulthood [McCloskey et al. 1991, p. 328]. Hence, after 35 000 epochs on skewed data (to a mean total sum squared error, tss, of 0.07) the network is trained for a further 35 000 epochs with equal frequencies (reaching a mean tss of 0.002). After training both the "skewed" and "equalized" networks can correctly recall all the patterns in the training set.

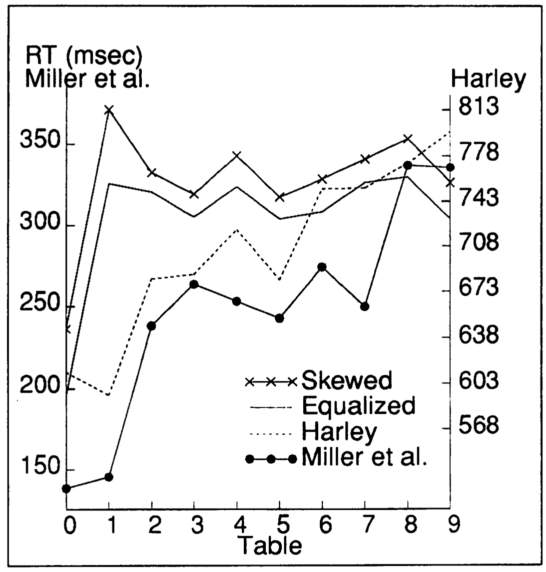

Figure 1. Plot of mean correct RT per multiplication table collapsed over operand order for: median RT of 42 adults [Harley 1991, app. D]; median RT (adjusted for naming time) of 6 adults [Miller et al. 1984, table A1]; mean RT for 20 networks trained on skewed frequencies; and, the same 20 networks after continued training on uniform frequencies (both networks equally scaled).

The "cascade" equations [McClelland & Rumelhart 1988, p. 153] are used to simulate the spread of activation in the network. Each unit's activity is allowed to build up over time:

$$\mathrm{net}_i(t) = k \sum_j w_{ij} a_j(t) + (1-k)\ \mathrm{net}_i(t-1),$$

where: k determines the rate with which activation builds up (0.05); w is the weight matrix; and $a_j(t)$ is the activation of unit j at time t. The net_i is passed through the logistic squashing function to produce the activation value, a_j, and the response values are taken to be the normalized activation values (outputs sum to 1.0). Fig. 2 is a time plot of output responses using the cascade equations.

Simulations

At the start of cascade processing the initial state of the network is that which results from processing an all-zeros input pattern (to give all patterns a common starting point). The network is trained to activate the "don't know" unit for all-zeros input. For each problem a random threshold is selected between 0.4 and 0.9. Processing continues until a product unit exceeds this threshold. The RT (number of cascade steps) is recorded for a correct response, and erroneous responses are classified into the five categories itemized above. The network is presented with each of the 100 problems 50 times, and the mean correct RT is recorded. This is repeated for 20 networks with different initial weights.

With a large threshold the correct product is reliably recalled, but occasionally the threshold will be low enough to accept an incorrect product. For example, "12" would be a close operand error for 4×4 in fig. 2. Presumably the mild time pressure of the experimental situation results in subjects lowering their thresholds.

Results

The mean RTs plotted in fig. 1 show some of the basic features of the problem-size effect. For the skewed networks the RT correlates $r = 0.22$ ($p = 0.013$) with adult RT reported

Table 1. Percentage breakdown of errors. Figures are mean values from twenty different networks, and mean values from 42 adult subjects [Harley 1991, app. B]. Adult scores other than error frequency were recomputed from Harley's data.

	Networks		Adults
	Skewed	Equalized	
Operand errors	93.56	93.15	86.20
Close operand errors	78.14	74.12	76.74
Frequent product errors†	22.56	20.14	26.99
Table errors	6.43	6.85	13.80
Error frequency	10.64	15.58	6.3

†Percentage of operand errors.

by Miller et al. [1984]. This increases to $r = 0.37$ ($p = 0.000067$) for the equalized networks. Slightly lower correlations were found to the Harley [1991] RTs. Note that the RTs have reduced and flattened out for the equalized networks, which is just what is expected after continued practice [Campbell & Graham 1985, p. 349]. One feature of the RT plot is the drop in RT for the nine times table. Children in grades 3 to 5 respond faster to $\times 9$ than $\times 8$ problems [Campbell & Graham 1985], but this levels out for adults. The tie bit is required to ensure that tie problems are among the fastest. All tie problems were below the mean for their table, except for 6×6. A comparison of error types is presented in table 1.

A correlation was found between problem error rate and correct RT. Campbell [1987, p. 110] reports a correlation of 0.93 for adults. For the skewed and equalized networks $r = 0.74$ and $r = 0.76$ respectively. It is not obvious that any model would necessarily predict that slower problems produce more errors.

Subjects sometimes respond with a number that is not a correct product for any of the problems 2×2 to 9×9 (e.g., $2 \times 3 = 5$). As the network can only produce 0×0 to 9×9 products as output, it cannot produce these non-table errors. Campbell & Graham [1985] report that only 7.4 per cent of errors are of this kind.

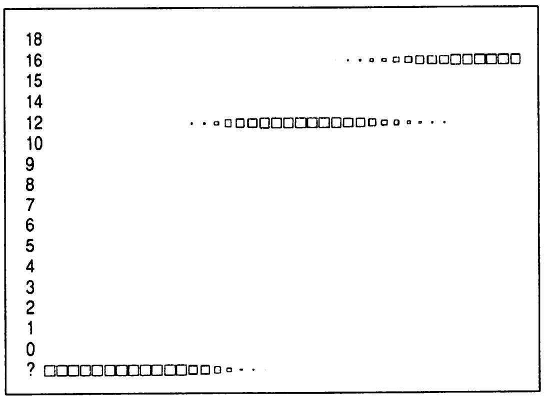

Figure 2. Response of the output units over 40 time steps for the problem 4×4. Output units for products over 18 are not shown on this graph.

Analysis and discussion

Presentation frequency, product frequency, initial weight values, and input encoding are all involved in determining the weights, which in turn determines the RT. Further assumptions suggested by other theories [Siegler 1988; Campbell & Graham 1985] were not required. In particular there was no need to explicitly train erroneous products, nor incorporate explicit global magnitude information, or between-product connections.

The presentation frequency peaks slightly for the five times table. This may be the reason behind the speed of the fives, but earlier simulations using a linearly skewed frequency curve also produced the characteristic dip. It should be noted that all the products in the five times table are unique in the sense that they do not occur outside the context of 5 (unlike 12 which occurs twice as often as any 5 product). The "12" output unit will need to respond to two different encodings, where as all 5 output units will only respond to one encoding (the networks learn the same encoding for $a \times b$ and $b \times a$).

It is often stated that $0 \times N$ problems are answered by rules. The evidence cited to support this notion is mostly from the relative RTs of zero problems [Miller et al. 1984; Stazyk, Ashcraft, & Hamann 1982]. From the model's RTs it appears that a rule is not required: zero problems have strong associations and are quick to be recalled. However, this means that zero also occurs as an error in unrealistic situations ($3 \times 8 = 0$). Harley's [1991] results show subjects producing errors of the form $0 \times N = N$, which is rarely observed in this model. Also, Sokol, McCloskey, Cohen, & Aliminosa [1991] produce evidence to support the notion of a rule: one of their brain damaged patients showed errors of $0 \times N = N$ on almost all $0 \times N$ problems, and then suddenly improved to almost perfect performance on these problems [p. 358]. These points indicate that it is the error performance, and not the RT, that suggests the special status (though not necessarily a rule) for $0 \times N$.

Acknowledgements

Funded by the SERC in conjunction with Integral Solutions Ltd. Thanks to Harry Barrow and David Young. Simulations were performed using a modified version of the McClelland & Rumelhart [1988] *bp* program, and POPLOG POP-11.

References

Anderson, J. A., Spoehr, K. T., & Bennett, D. J. [1991]. A study of numerical perversity: Teaching arithmetic to a neural network. Tech report 91–3, Department of Cognitive and Linguistic Sciences, Brown University.

Campbell, J. I. D. [1987]. The role of associative interference in learning and retrieving arithmetic facts. In Sloboda, J. A. & Rogers, D., eds., *Cognitive Processes in Mathematics*, pp. 107–122. Clarendon Press, Oxford.

Campbell, J. I. D. & Graham, D. J. [1985]. Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, *39*(2), 338–366.

Harley, W. S. [1991]. *Associative Memory in Mental Arithmetic.* Ph.D. thesis, Johns Hopkins University, Ann Armor, MI.

McClelland, J. L. & Rumelhart, D. E. [1988]. *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises.* MIT Press, Cambridge, MA.

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, and Cognition*, *17*(3), 377–397.

Miller, K., Permutter, M., & Keating, D. [1984]. Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology*, *10*(1), 46–60.

Siegler, R. S. [1988]. Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*(3), 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 subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 355–376.

Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. [1982]. A network approach to mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 320–335.