

Connectionist Learning and Education: Applications and Challenges

Thomas R. Shultz (thomas.shultz@mcgill.ca)

Department of Psychology and School of Computer Science, McGill University, 1205 Penfield Avenue
Montreal, QC H3A 1B1 Canada

Abstract

Successful applications of connectionist learning to educational issues include word reading, single-digit multiplication, and prime-number detection. While closely modeling human learning, these applications underscore the importance of practice, feedback, prior knowledge, and well-structured lessons. Among the remaining simulation challenges in educational domains are other reading and arithmetic skills, learner-generated goals, social aspects of learning, and learning by direct instruction.

Keywords: Connectionism; education; reading; multiplication; prime-number detection.

Introduction

Because connectionism has dominated the theoretical study of learning for the last 20 years, it would be surprising if it yielded no useful applications to the field of education, where learning is obviously a key issue. However, most connectionist research is resolutely theoretical with very little effort expended on solving educational problems. The few systematic treatments of applying connectionism to education were written years ago and seem to have barely scratched the surface of possibilities. This paper reviews some of the more promising lines of connectionist research into learning in the key educational domains of reading and mathematics before attempting to abstract some recommendations that might be of interest to educational researchers. First though, modern connectionism is distinguished from the older, behaviorist connectionism, which had been applied to education nearly a century ago. Finally, some of the considerable challenges that education poses for connectionist modeling are discussed.

Connectionism Old and New

The central idea of modern connectionism is that mental processes can be modeled by interconnected networks of simple units. This is quite different from the older, behavioristic connectionism. Behaviorism emphasized the learning of associations between stimuli and responses and the idea that responses become habitual by being rewarded.

Despite some rather minor historical influence on, and superficial similarities with, modern connectionism, the differences are more profound than some critics believe (e.g., Fodor & Pylyshyn, 1988). Whereas behaviorists discussed a single association between a stimulus and a response, modern connectionism deals with large, multilevel, massively parallel networks. Moreover, many contemporary networks are designed with recurrent connectivity which allows for sequential processing and complex network dynamics. Following from these

differences, the old knowledge representation schemes were entirely local, whereas modern networks often employ distributed schemes in which each unit represents many different ideas and each idea is represented by many different units. These distributed representations are more efficient and robust, more biologically realistic, and account for a variety of interesting psychological phenomena. Behaviorism was uninterested in mental states, while modern connectionists invest considerable energy into determining what their networks know at various points in learning. Such knowledge-representation analyses often become an essential part of explaining psychological phenomena.

The *law of effect* emphasized that habit formation was controlled by rewards. Contemporary connectionist models have demonstrated the difficulty of learning from evaluative reward signals indicating that an organism is doing well or badly (Hertz, Krogh, & Palmer, 1991). In contrast, many neural networks learn from fully specified target vectors that indicate the correct response to particular inputs, making learning faster and more accurate. Taken together, these differences provide modern networks with vastly more learning power than simple habits possess. A habit implements only a simple linear relation between a stimulus and a response, whereas there are now proofs that a network with a single layer of hidden units can learn any continuous function to any degree of accuracy if this layer has enough hidden units (Hertz et al., 1991). There are also proofs that any function can be learned by a network with two hidden layers, if there are enough hidden units in each layer. Finally, although most of the theoretical work in behaviorism was vague and speculative, contemporary connectionism is characterized by working computational models that enable evaluation of the quality of data coverage and generate testable predictions.

Connectionism has been more concerned with establishing a theoretical understanding of learning than with developing applications to education or other practical fields. However, because connectionists have had so much success modeling the learning of reading and elementary mathematics, there is the possibility for applications to educational practice.

Models of Reading

A debate about whether it is better to teach reading with the rules of letter-to-sound correspondence or by learning to visually recognize whole words began in the 1960s (Foorman, 1994). There are hundreds of such phonic rules, but because letter-to-sound correspondence is only quasi-regular, they are not all that useful to learn as universally

quantified rules (Seidenberg, 2005). An example rule is *When there are two vowels side by side, the long sound of the first one is heard and the second is usually silent*. Another example is *When there are two vowels, one of which is final e, the first vowel is long and the e is silent*. It turns out that the first rule is correct only about 45% of the time and the second about 63% of the time (Adams, 1990).

A theoretical framework for an influential series of connectionist models of reading that are not based on rules is pictured in Figure 1 (Seidenberg, 2005). The solid rectangles represent groups of network units encoding information on orthography, phonology, or semantics, while the dashed rectangles represent groups of hidden units that encode nonlinear transformations between the spelling, sound, and meaning encodings. The bidirectional arrows indicate connection weights traveling in both directions. By adjusting these connection weights, the system can learn to transform written words into pronunciations or meanings, meanings into written words or pronunciations, and pronunciations into written words or meanings. Most of the research to date has concentrated on the mapping from written words to pronunciations, i.e., reading aloud. Such network-based models learned to pronounce the words they were trained on, such as *gate* and *save*, and generalized successfully to novel words such as *rave* (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996).

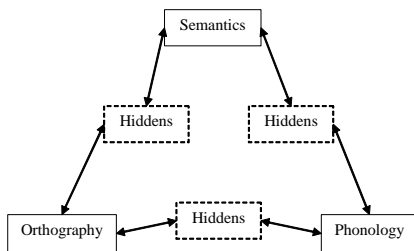


Figure 1: Theoretical framework for Seidenberg's connectionist models of reading.

These connectionist models also covered a number of well-documented psychological regularities such as frequency, similarity, and regularity effects. The frequency effect holds that common words are read more quickly than rare words. Network models also correctly predicted an interaction of frequency with similarity, namely that frequency effects would be smaller for words with many similar neighbors (e.g., *save*) than for more isolated words such as *sieve*. The regularity effect is that words with regular neighbors are read more quickly than words with irregular neighbors. For example, the word *gave* has a regular pronunciation but has irregular neighbors like *have*. Consequently, words like *gave* take longer to read aloud than words such as *must*, which have no irregular neighbors. Such regularity effects are larger for lower-frequency words and for less-skilled readers.

In neural networks, such effects can be understood in terms of weight sharing. Because all words share the same set of network weights and neural learning attempts to

reduce as much error as possible, frequent words, words highly similar to other words, and words with regular neighbors are read more quickly and accurately. Similar words or regular words tend to support each other in terms of pronunciation. However, high frequency words can achieve speed and accuracy without similarity and regularity because of their sheer frequency. These three factors of frequency, similarity, and regularity can compensate for each other in that words at a disadvantage in one respect might benefit from another factor.

A curiosity of Seidenberg's models is that a network has no explicit representation of lexical entries, i.e., words. Language researchers typically not only assume a lexicon, but the frequency effect in reading is customarily explained by storing frequencies for each lexical entry. Nonetheless, Seidenberg's networks cover the frequency effect without possessing an explicit lexicon, suggesting that it may be unnecessary in people as well.

Recent models that also include the mapping from orthography to semantics enabled connectionist models to address the issue of phonics vs. visual recognition in teaching reading (Harm & Seidenberg, 1999). In contrast to previous models assuming a conflict between visual and phonetic routes, their networks revealed collaboration between the two routes. Early in training, networks relied somewhat more on the orthography-phonology route, but with additional training, the orthography-semantics-phonology route increased in importance, simulating a psychological progression observed in children. Just as with children, skilled reading of words involved convergent contributions of both of these routes from orthography to phonology.

Dyslexia can be simulated in these network models by impairing either the network or its training (Harm & Seidenberg, 1999). Reducing the number of hidden units could be analogous to a child with limited cognitive resources (Seidenberg & McClelland, 1989). Making each letter string activate more orthographic units could mimic a visual impairment (Seidenberg, 1992). Limiting the amount of training could correspond to inadequate educational opportunity (Seidenberg, 2005). Ignoring training in the orthography-phonology route could simulate teaching without phonics (Harm & Seidenberg, 1999). In all of these impaired cases, network learning focuses on the largest current source of error, namely that contributed by regularly-pronounced words, thus sacrificing the reading of words with exceptional pronunciations.

Other, rule-based computational models have relatively more difficulty accounting for this diverse set of psychological phenomena, underscoring the conclusion that these networks make a compelling descriptive model of how children learn to read.

The main prescriptive implication of these models for reading instruction is that students would learn to read well if provided with plenty of examples of printed words, and their meanings and pronunciations. Pointedly, children should not be trained in the application of specific phonic

rules (because those have too many exceptions) and should not be trained without attention to phonological codes.

Models of Mathematics

Mathematics is another area of unnatural skills taught over several years of formal instruction which has also been modeled with neural networks. The two examples considered here are learning of the single-digit multiplication table and prime-number detection.

Multiplication

Learning the single-digit multiplication table requires about 5 to 6 years of schooling and even adults continue to make some errors (Campbell & Graham, 1985). The most commonly studied multiplication problem is the so-called *production* task in which two single-digit multiplicands are presented and the participant is asked to provide their product. Several regularities are evident in the psychological literature on single-digit multiplication:

1. Computational methods such as repeated addition ($m \times n = \text{adding } m, n \text{ times}$) are gradually replaced by recall of products (Sieglar, 1988).
2. Reaction time increases with the size of the multiplicands, except that the 5s table and tie problems (e.g., 3×3 , 8×8) are quicker than would be expected (Campbell & Graham, 1985).
3. Adults who are under mild time pressure make errors on about 8% of the problems (Campbell & Graham, 1985).
4. Errors are typically close to the correct product, and often substitute a close multiplicand for the correct answer (McCloskey, Harley, & Sokol, 1991).
5. There is a sizeable correlation ($r = .93$) across problems between reaction time and error (Campbell, 1987).

Building on the successes and overcoming some of the limitations of earlier connectionist models (Anderson, Spoehr, & Bennett, 1991; McCloskey & Lindermann, 1992; Stazyk, Ashcraft, & Hamann, 1982), Dallaway (1994) designed a feedforward network that captured phenomena 2-5. The topology of Dallaway's model for multiplying the digits 2-9 is shown in Figure 2. Target output vectors were designed by turning on one product unit and leaving the others off, implementing so-called 1-of- n coding. Percentage of error types plotted in Figure 3 indicate a good fit of the model to adult errors, although the overall error rate was higher for the networks at 14.1%. As shown in Table 1, *operand* errors are characterized by changing one of the operands, *close-operand* errors by changing to an operand close to a multiplicand, and *frequent-product* errors by giving a frequently occurring product. *Table* errors involve answering with a less frequent product that is in the multiplication table but does not share multiplicands with the problem being tested. *Operation* errors are produced by adding instead of multiplying the given multiplicands. As measured by settling time, networks reacted about as quickly to multiplication by 6 as to multiplication by 5. This

was unexpected and it complicates any easy explanation of the speedup on 5s problems.

Variants models did not fit the human data nearly as well as the foregoing model did. For example, model fit deteriorated when the 0 and 1 multiplication tables were trained along with the 2-9 tables. Some researchers believe that multiplication by 0 and 1 is rule governed, rather than being based on connectionist pattern matching (McCloskey, Aliminos, & Sokol, 1991), but it seems possible that the greater regularity of 0 and 1 multiplications only makes them seem rule governed. Also, fit to human reaction time data was worse when the training sample was no longer biased in favor of smaller multiplicands. Even with these limitations, just as with computational models of learning to read, connectionist models here have few rivals for fitting human performance.

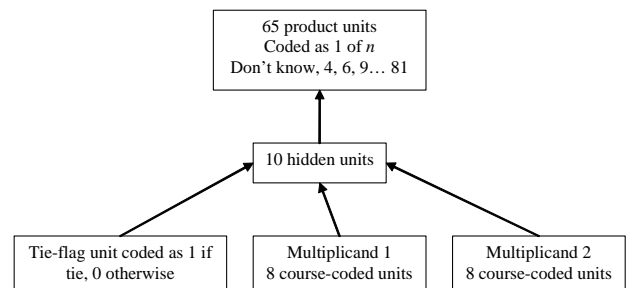


Figure 2: Dallaway's (1994) multiplication network.

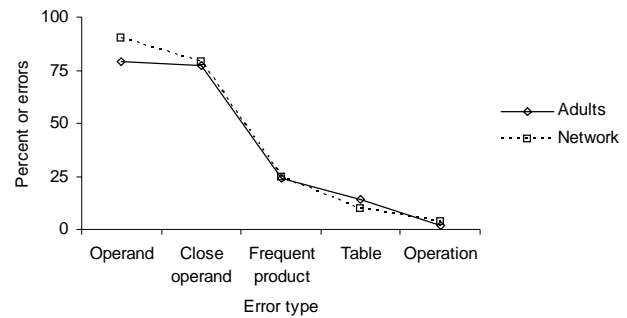


Figure 3: Percent of error types (Dallaway, 1994). Percents do not sum to 100 because types are not mutually exclusive.

Table 1: Multiplication table with three types of error highlighted

x	2	3	4	5	6	7	8	9
2	4	6	8	10	12	14	16	18
3	6	9	12	15	18	21	24	27
4	8	12	16	20	24	28	32	36
5	10	15	20	25	30	35	40	45
6	12	18	24	30	36	42	48	54
7	14	21	28	35	42	49	56	63
8	16	24	32	40	48	56	64	72
9	18	27	36	45	54	63	72	81

Close operand errors for 4x5: □
 Operand errors for 9x8: ■
 Frequent products: *italics*

Applications of this and related models to educational practice remain tentative, but likely would be similar to those made for reading: include many examples of correct multiplication that go just beyond the student's current ability. The role of addition in learning and understanding multiplication should probably be explored in future computational modeling and psychological experimentation because of its apparent role in children's learning and its possible role in several multiplication errors (Lemaire & Siegler, 1995).

Prime-number detection

Detection of prime integers is a relatively advanced mathematical skill that has also been modeled with neural networks. An integer greater than 1 is a prime number if it has two divisors, 1 and itself. An integer greater than 1 having more than two divisors is a composite number. The integer 1 is neither prime nor composite.

It might seem that the primality of an integer n could be determined by checking whether n is divisible by any integer between 2 and $n - 1$. Prime-number detection can be done in that fashion, but it can also be done more efficiently. The only divisors really needed are prime numbers from 2 to the integer part of \sqrt{n} . Still more efficiency can be gained by starting with the smallest prime number and increasing divisor size until locating a divisor that divides evenly into n . Starting with small divisors and increasing divisor size in this way is efficient because the smaller the prime divisor, the more composite numbers it can detect in any fixed range of integers.

A connectionist system called knowledge-based cascade-correlation (KBCC) discovered this efficient algorithm from learning examples and recruiting previously-learned knowledge of divisibility (Egri & Shultz, 2006). KBCC is based on a somewhat simpler connectionist algorithm called cascade-correlation (CC). CC learns from examples by recruiting single hidden units as needed to reduce network error. CC was used to simulate a large number of developmental phenomena (Shultz, 2003). Compared to CC, KBCC has the added advantage that it can recruit previous knowledge stored in networks as well as recruiting single hidden units (Shultz & Rivest, 2001). Both CC and KBCC are constructive neural learners that build their new learning on top of existing knowledge.

In the prime number simulation (Egri & Shultz, 2006), the pool of source knowledge contained networks that had previously learned whether an integer could be divided by each of a range of divisors. There was a divide-by-2 network, a divide-by-3 network, a divide-by-4 network, etc., up to a divisor of 20. The source networks had been trained on integers in the range of 2-360. Then 20 KBCC target networks were trained on 306 randomly-selected integers in the range of 21-360. As these target networks learned, they opted to recruit only source networks that involved prime divisors below the square root of the largest number they were trained on, in this case 360. These sources were recruited in order from small to large, and installed on a

single layer. Moreover, the target networks avoided recruiting single hidden units, source networks with composite divisors, any divisors greater than square root of 360 even if they were prime numbers, and divisor networks with randomized connection weights.

KBCC target networks never recruited a divide-by-2 source network, but this was because they instead used the least significant digit of n , which was coded in binary form, to directly determine if n was odd or even. Like people who use the least significant digit of base-10 numbers to check for divisibility by 5 or 10, this is a handy shortcut to having to divide by 2.

The KBCC target networks learned to classify their training integers about three times faster than did knowledge-free CC networks, and they generalized nearly perfectly to test integers that were untrained. Without divisibility knowledge, networks did not generalize better than chance. As predicted by this simulation, adults testing the primality of integers also used mainly prime divisors below \sqrt{n} and ordered divisors from small to large, showing that the networks provided an accurate model of human performance. The main recommendation for education is not only to use examples, but also to structure curricula so that learning can build on existing knowledge.

Educational Relevance

As neural network modeling of learning continues, further applications to education could become more apparent. Some implications of such models of reading and mathematics were already noted, concerning the use of examples and fully specified feedback on what to do with those examples. Teaching with examples is compatible with the classically important idea of learning by doing. Because students may vary considerably in their current skills, providing such examples can be challenging in a classroom setting. However, it could be accomplished with materials that vary enough in difficulty to continuously provide at least some examples just beyond the ability of each student.

Another important recommendation from connectionism is to accompany examples with complete feedback about appropriate responses to each problem. Full target feedback is more informative than the evaluative feedback provided by rewards which was emphasized in classical connectionism. Computational results show that converting information about being wrong into a useful target vector makes learning both slower and more difficult (Hertz et al., 1991).

Full target feedback is also more informative than the cues to disequilibrium that were characteristic of the educational recommendations of Piagetian theory (Piaget, 1970). In Piaget's view, disequilibrium occurred when there was an imbalance between the processes of assimilation and accommodation. This could occur when a child copied without understanding or distorted reality to fit internal conceptions. In either case, cognitive change might be stimulated, but not much useful information was provided about how to improve and thus restore equilibrium.

Another clear educational recommendation stemming from connectionist modeling is that repetition and patience are often required for successful learning. This idea derives partly from the fact that connectionist learning is often quite slow. Some insight into reasons for this slowness has been attained from systematic study of variation in learning-rate parameters. When learning rate is set too high, networks often oscillate across error minima. To settle closer to such minima, it is often necessary to lower learning rate in order to take small steps in connection-weight adjustments, thus slowing down learning. This could also be true of brain networks, and if it is, then educators should not expect success by rushing through difficult material.

As might be expected, methods for increasing both the speed and accuracy of network learning are under active investigation. For example, it is becoming clearer that networks learn faster and more accurately if they can bias their learning by recruiting relevant existing knowledge. In the case of prime-number detection, successful generalization to untrained integers actually required recruiting existing knowledge about divisibility. This suggests the use of curricula designed to ensure that lessons are presented in some optimal order. Network simulations might be useful in identifying lesson sequences likely to facilitate learning and generalization.

Yet another implication of connectionist simulations is that context is important and that it can limit the amount of generalization. Connectionist learning algorithms naturally exhibit context effects whenever it is the case that contextual cues aid learning. The tradeoff is that such contextual effects ensure that generalization is not universal. If more generalization is desired, teachers might want to decontextualize learning. Decontextualization could be accomplished by varying contextual cues while learning from examples, thus allowing a learner to generalize the basic target function across different contexts. Again, it might be the case that exploratory network simulations could help to determine how best to accomplish this.

Many of these educational recommendations coming from connectionist research (practice, feedback, prior knowledge, well-structured lessons) at first may appear more consistent with teacher-centered, rather than child-centered, approaches to education. This seems a bit paradoxical given that constructive connectionist approaches (such as CC and KBCC) are quite consistent with a Piagetian approach to knowledge acquisition that serves as the psychological basis for much child-centered education.

Whereas teacher-centered education focuses on structured lesson plans, extensive practice, and feedback, child-centered education emphasizes curiosity, problem solving, and learning by discovery (Chall, 2000). Although these approaches are often portrayed as being in opposition, constructivist connectionist modeling suggests a possible rapprochement, by providing computational demonstrations that effective learning incorporates both of these approaches. In connectionist learning, knowledge representations are constructed and abstracted by the

learner, rather than merely memorized. Moreover, this learning is particularly effective when lessons are well structured, building more complex ideas on top of simpler earlier ideas, and well practiced, with detailed information about correct responses.

Challenges for Future Research

There are a number of educational concerns that are currently well beyond the ability of current connectionist models. Some of these stem from the fact that current models do not yet cover many of the phenomena in learning to read or perform mathematics. For example, existing connectionist models of reading do not cover more than the reading of monosyllabic words. Reading of multisyllabic words, phrases, sentences, paragraphs, and large bodies of text all await further investigation. This is also true of other paths within Seidenberg's word-reading framework that relate print, sound, and meaning.

There are likewise many aspects of arithmetic that are still uncovered by connectionist models including counting and subitizing; addition and subtraction; multi-column addition, subtraction, and multiplication; division; ratios and proportions; algebra; and many other topics. At this point, even the conceptual origin of integers is still a mystery, particularly but not exclusively for connectionist approaches that represent unit activations in terms of real numbers. In both reading and arithmetic, connectionism is only just getting started.

Another unsolved problem is the neglect of the goals that learners might have. Much of connectionist learning is built on the principle of error reduction, where error is the discrepancy between actual and target output values. Error reduction could well be a goal of human learners, but it is likely that humans sometimes create other learning goals. Some examples of alternate goals could be completing an assignment, achieving happiness, accepting a challenge, or enjoying social interaction. It is unclear whether connectionism could capture the setting of goals and learning under such goals.

In general, the social aspects of learning are not addressed by current connectionist modeling. These would include the teacher-student relationship as well as peer interaction, competition, and collaboration.

A particularly glaring omission from the connectionist literature is that of direct, explicit instruction. Most formal education involves lectures or lessons presented in a verbal fashion by an instructor to rather passive students. Although examples and student activity can sometimes be skillfully included in direct instruction, it is not clear how neural networks would be able to learn from direct verbal instruction itself. It may be that typical connectionist weight adjustment procedures are far too slow to capture the rapid learning that sometimes follows direct verbal instruction. A particularly vivid example concerned training humans to detect the gender of day-old chicks by either example (which takes years) or direct instruction (which takes only minutes) (Biederman & Shiffar, 1987).

One promising approach is to represent verbal instruction as a pattern of activation in a constraint-satisfaction network (Noelle & Cottrell, 1995). Attractor basins in such a network could be trained as a kind of instruction language. Additional attractors might be realized by interactions among trained attractors. Direct instruction input could settle into an attractor basin and modulate a feed-forward task-learning network. Applied to multi-column addition of binary numbers, with instruction sequences such as *write a sum*, *announce a carry*, and *move to next column*, a version of this system learned instructions but failed to generalize to novel problems. Because of the clear importance of learning from instruction, and its likely interaction with learning by doing, further work on this problem is probably warranted.

Acknowledgments

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

- Adams, M. J. (1990). *Beginning to read*. Cambridge, MA: MIT Press.
- Anderson, J. A., Spoehr, K. T., & Bennett, D. J. (1991). *A study of numerical perversity: Teaching arithmetic to a neural network*. Department of Cognitive and Linguistic Sciences, Brown University.
- Biederman, I., & Shiffar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 640-645.
- Campbell, J. I. D. (1987). The role of associative interference in learning and retrieving arithmetic facts. In J. A. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics*. Oxford: Clarendon Press.
- Campbell, J. I. D., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, *39*, 338-366.
- Chall, J. S. (2000). *The academic achievement challenge: What really works in the classroom?* New York: Guilford.
- Cottrell, G. W., & Tsung, F.-S. (1993). Learning simple arithmetic procedures. *Connection Science*, *5*, 37-58.
- Dallaway, R. (1994). Dynamics of arithmetic: A connectionist view of arithmetic skills. *Cognitive Science Research Papers* 306. Brighton: University of Sussex.
- Egri, L., & Shultz, T. R. (2006). A compositional neural-network solution to prime-number testing. In *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.
- Foorman, B. R. (1994). The relevance of a connectionist model of reading for "The Great Debate". *Educational Psychology Review*, *6*, 25-47.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491-528.
- Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Reading, MA: Addison Wesley.
- 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.
- McCloskey, M., Aliminoso, D., & Sokol, S. M. (1991). Facts, rules, and procedures in normal calculation: Evidence from multiple single-patient studies of impaired arithmetic fact retrieval. *Brain and 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, and Cognition*, *17*, 377-397.
- McCloskey, M., & Lindermann, 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*. Amsterdam: Elsevier.
- Noelle, D. C., & Cottrell, G. W. (1995). A connectionist model of instruction following. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Pittsburgh: Erlbaum.
- Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.
- 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.
- Seidenberg, M. S. (1992). Dyslexia in a computational model of word recognition in reading. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition*. Hillsdale, NJ: Erlbaum.
- Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*, 238-242.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.
- Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.
- Shultz, T. R., & Rivest, F. (2001). Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science*, *13*, 1-30.
- Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258-275.
- 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*, 320-335.