A Connectionist Model of the Development of Seriation

Denis Mareschal
Department of Experimental Psychology
University of Oxford
South Parks Road
Oxford, England OX1 3UD
maresch@ox.ac.uk

Thomas R. Shultz
Department of Psychology
McGill University
1205 Penfield Avenue
Montréal, Québec, Canada H3A 1B1
shultz@psych.mcgill.ca

Abstract

Seriation is the ability to order a set of objects on some dimension such as size. Psychological research on the child's development of seriation has uncovered both cognitive stages and perceptual constraints. A generative connectionist algorithm, cascade-correlation, is used to successfully model these psychological regularities. Previous rule-based models of seriation have been unable to capture either stage progressions or perceptual effects. The present simulations provide a number of insights about possible processing mechanisms for seriation, the nature of seriation stage transitions, and the opportunities provided by the environment for learning about seriation.

Introduction

Recent research has indicated that basic phenomena in cognitive development can be successfully modeled with connectionist networks (cf. reviews by Bates & Elman, 1992; Plunkett & Sinha, 1992; Shultz, 1991). Generative learning algorithms, such as cascade-correlation (Fahlman & Lebiere, 1990), are particularly appealing for such simulations since they implement qualitative increases in network capacity during learning in addition to quantitative adjustments of network weights. It has long been held that much of the child's cognitive development is due to qualitative shifts in representational capacity (Piaget, *passim*), but until recently it has been unclear how such qualitative changes could be modeled.

A variety of aspects of cognitive development have been successfully modeled with cascade-correlation nets, including balance scale phenomena (Shultz & Schmidt, 1991), causal predictions (Shultz, Zelazo, & Strigler, 1991), and the acquisition of personal pronouns (Shultz, Buckingham, & Oshima-Takane, 1993). Here, we report on the application of cascade-correlation to another well known cognitive developmental phenomenon, seriation.

Seriation is the problem of sorting elements according to their respective values on a dimension. Sorting is a well studied problem in computer science with more than a dozen established symbolic algorithms (Knuth, 1973). Although there have been a few hand designed neural networks that sort (Atkins, 1989; Chen & Hsieh, 1990), we are presenting the first connectionist system that *learns* to sort.

Psychology of Seriation

Inhelder and Piaget (1969) developed the seriation task in order to demonstrate the presence of stages in the development of children's transitive reasoning. Children were presented with a disordered set of size graded sticks. After showing that they could identify the smallest and the largest of these sticks, the children were asked to build a staircase, i.e., to order the sticks from smallest to largest. Finally, children were presented with a few intermediate sticks and were told to insert these at their correct place in the series.

Four stages were identified in the child's ability to construct a series. Examples of these stages are shown in Figure 1. Each stage was interpreted as resulting from the presence or absence of radically different cognitive processes. During a first stage (*circa* 4 years old), children made no real effort at ordering the sticks and either lined them up in the order they appeared or moved them about randomly.

Children in the second stage (*circa* 5 years old) succeeded in combining sticks in terms of local absolute quantities such as *big* or *small*, but were unable to extend this order over the entire set of sticks. This led to series of uncoordinated pairs (pairs of large and small elements), uncoordinated triplets (one large, one medium, and one small element), seriation based on the correct alignment of only the tops of the sticks, roof-top seriation (in which the tops rise and then descend, or vice-versa), and correct seriation of only the first few elements.

This research was supported in part by a Fellowship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec to Mareschal and an operating grant from the Natural Sciences and Engineering Research Council of Canada to Shultz.

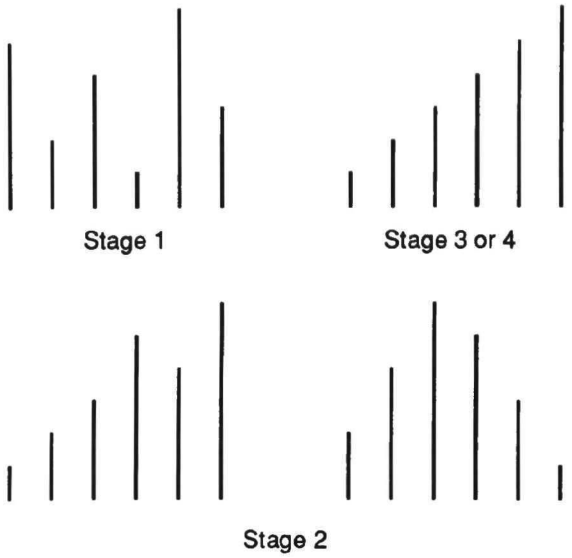

Figure 1. Examples of seriation stages.

During a third stage (*circa* 6 years old), children were able to construct a series but only through extensive trial and error; as though they were completing a difficult jigsaw puzzle piece by piece.

Finally, during the fourth stage (*circa* 7 years old), children could construct the series rapidly and with few errors by applying a systematic strategy that Piaget called the operational method. Operational seriation consisted in selecting the smallest unordered stick and placing it immediately in its appropriate place in the series under construction.

More stringent studies initially replicated Piaget's observations (Elkind, 1964). However, in depth protocol analyses suggested that seriation was non-algorithmic and more flexible than Piaget had suggested (Young, 1976). Retschitzki (1978) claims that no stage can truly be described by a single procedure. Random selection strategies are observed in children of all ages including those well into the fourth, operational stage (Kingma, 1982). Other partial seriator characteristics are also present at all ages (Kingma, 1983).

Moreover, perceptual constraints were found to influence performance on seriation tasks. If differences between adjacent sticks are too large, stage 3 seriators perform at stage 4 since the perceptual information that stage 3 children rely on is so salient as to allow them to construct a series reliably and efficiently (Piaget, 1965). If discriminations between sticks are too small, seriation performance deteriorates (Elkind, 1964; Kingma, 1984).

Koslowski (1980) found that 4-year-old children could seriate at a stage 4 level with four elements but failed to generalize this to the traditional 10 element task where they were diagnosed as being in stage 1. She argues that, even though the children failed the traditional task, their sorting in the abbreviated task implies that they were using the required operations but with less refined size discriminations. Koslowski suggests that one component along which development occurs is the precision with which seriation operations are carried out.

Previous Computational Models

Due to Piaget's early systematic description of seriation, several symbolic computational models of seriation performance have been created. As early as 1964, Frey outlined a set of automata that would incorporate the cognitive structures thought by Piaget to be present at each stage and that would produce the observed behavior.

Two rule-based approaches were also published. One of these (Baylor et al., 1973) was based on the protocols of three children, each in a different developmental stage. No transition mechanisms between stages were specified, although development was thought to be driven by a progressive sophistication in the child's structuring of the experimental environment, a developmental trend towards stimulus independence, and a tendency towards the development of more efficient procedures.

Young's (1976) rule-based model was designed to match the idiosyncrasies of individual children at different ages. He suggested that development occurred along three different dimensions: selection of an item, evaluation of that choice, and correction of incorrect choices. Development was seen as a continuous process consisting of the acquisition of new rules. Contrary to Baylor et al.'s model, children do not acquire more efficient algorithmic procedures but add more discriminating rules to an existing kernel.

None of these models are truly developmental since they do not provide a mechanism for passing from one stage to the next.[1] Therefore, they fail to capture stage progressions. Nor do they address the issues surrounding perceptual saliency. Moreover, these models are overly rigid in that hand tailored sets of rules function consistently, thereby failing to capture the flexibility inherent in children's protocols. Connectionist approaches might be better suited to the modeling of children's performance on the seriation task due to their ability to capture rule-like behavior and perceptual effects without sacrificing flexibility.

---

[1] Anzai (1987) presented a production rule model that constructed a few new seriation rules from an initial set of 31 productions. It used two different learning strategies: backtracking to avoid bad moves and looking ahead to select good moves. The model did learn to sort more efficiently, but it was not evaluated for psychological realism.

Cascade-correlation

Cascade-correlation is a generative algorithm that builds its own feed-forward network topology when it needs to (Fahlman & Lebiere, 1990). The network begins with the minimal topology of an input layer fully connected to an output layer. Weights are trained using the quickprop algorithm (Fahlman, 1988) to minimize the difference between observed and desired activation across the output units. Quickprop is a minimization algorithm loosely based on Newton's second order method. Information about the current and previous derivatives are used to construct a local approximation of the curvature of the potential to be minimized.

If cascade-correlation fails to reduce the error within an acceptable criterion, it then proceeds to construct and insert a hidden unit with inputs from all units in the network other than the outputs. These input weights are trained so as to maximize the correlation between each of several candidate hidden units' activations and the residual error at the outputs. Once the correlations reach asymptote, the hidden unit whose activations correlate best with the error is installed in the network with trainable connections to the outputs. Once installed, the weights leading into the new unit are frozen and can never be changed. The network then reverts back to the error minimization phase but with the added power of a new hidden unit tuned to the residual error. If necessary, the cycle is repeated until success is achieved.

Simulations

The simulations presented here focus on the ability to construct a six element series. Although this is fewer than the 10 elements used by Piaget, it is more than the four used in abbreviated tasks and corresponds closely to the number used in the perceptual saliency studies.

Our models consist of a component devoted to the processing of seriation information. This component receives information about the present state of the series, processes this information, and outputs a move. A move is defined as the identification of a stick and of the position to which that stick should be moved. The move is not actually carried out by the network, but by auxiliary software that maintains the array. Psychologically, moves are hypothesized to be carried out in the environment by an auxiliary motor component. As in Young's (1976) model, the entire seriation performance is constituted by the serial juxtaposition of independent moves.

The seriation component is composed of two distinct modules, each processing the same input array but responding independently (Figure 2). Processing occurs in parallel; information is only integrated once it has reached the output units. One module is trained to identify which stick to move, whereas the other is trained to identify where this stick should be placed. Since the weights in the two modules are updated independently, behavioral development of the whole system can be attributed to the interaction of development in the individual skill modules. The modules are trained to respond as dictated by Piaget's operational procedure, i.e., to move the smallest stick that is out of order to its correct place.

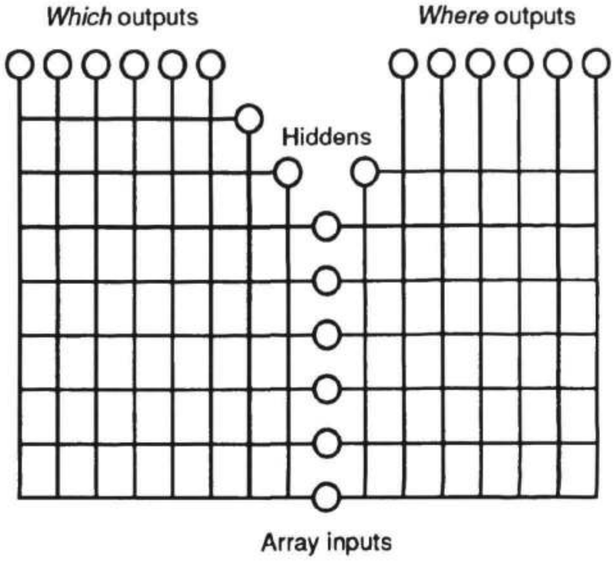

Figure 2. Schematic representation of the composite seriation network. The two modules are independent except for sharing the same input. Here, the *which* net has two hidden units and the *where* net has one hidden unit.

The *which* and *where* modules are each individual cascade-correlation networks. In the standard cascade-correlation package, it is not possible to test performance during the hidden unit recruitment phase. This does not imply that the network has entered some kind of hibernation, but that development of the behavior inferred from the outputs has reached an asymptote. Thus, during the hidden unit recruitment phase, the output response is taken to be that given by the last epoch immediately preceding the phase transition.

Each seriation component is trained on 120 examples of seriation moves. The input consists of an array configuration; the output consists of either which stick should be moved or where a stick should be moved to, depending on the module. For example, given the array {1 3 5 6 2 4} the composite networks learn to identify that the stick in the fifth position should be moved to the second position. The input is coded on a bank of six linear units with integer activation indicating the relative heights of the sticks. The output is coded on a bank of six non-linear sigmoid units on which the unit coding the correct position is turned maximally on and all others are maximally inhibited. All hidden units have a sigmoid activation function ranging from -0.5 to +0.5.

Pilot studies revealed that the disorder of the array presented was important in determining the network's success. Therefore, the training set consisted of 50 randomly selected arrays whose sum-squared distance from the target array {1 2 3 4 5 6} was less than or equal to 20, and 50 randomly selected arrays whose sum-squared distance was greater than 20.[2]

This corresponds to the assumption that many of the events from which the young child learns about seriation involve nearly seriated arrays. When items are completely disordered, the child is less likely to conceive of them in terms of a series and therefore may not attempt to seriate them. On the other hand, if items are largely ordered, the child is more likely to conceive of them as a nearly completed series. Such a nearly completed series might then serve as a cue for completion of the ordered series.

In order to reflect the finding that even young children can seriate small sets of sticks (Koslowski, 1980), 20 randomly selected arrays containing only three elements (e.g., {0 0 3 1 2 0}) were added to the training set.

Macroscopic seriation behavior is evaluated by presenting the network with a standard array {5 2 4 1 6 3}, not present in the training patterns.[3] The move output by the network is carried out by moving the selected stick and adjusting the others in order to fill the empty slots. The resulting array is then cycled back as input to the network. Since this system is deterministic, the cycling process is terminated once an array has appeared twice. The resulting collection of arrays constitutes a trace of the network's seriation performance.

Stage diagnosis requires information on both the final state of the array and the method used to arrive at that final array. Diagnosis is controversial, even in children, as a variety of criteria are used (Kingma, 1982). Here, stages 1 and 2 are diagnosed as described by Piaget. To distinguish between stage 3 (empirical) and stage 4 (operational) seriators, both the procedure used and the number of self-corrections criterion are simultaneously applied. A network is classified as stage 4 if it correctly constructs a series according to the operational method with at most one error from which it continues using the same operational method, or if it seriates in the same or fewer moves than required by the operational method. It is classified as stage 3 if it constructs a completed series in any other way. To be classified as a proper stage, a behavior must last at least four consecutive epochs.

Results

Under these conditions, networks typically exhibited behavior in all four stages. Figure 3 shows a representative network progressing through stages 1, 2, 3, and then 4. Seven of the 20 networks progressed through this four stage sequence, often with a large overlap between adjacent stages suggesting that, for some of the developmental period, no single stage is truly characteristic of the network's behavior. The transitions are soft, with adjacent stages gradually merging into one another.

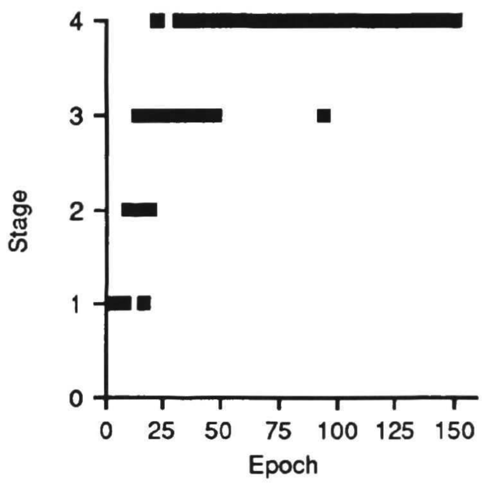

Figure 3. Stage diagnosis in one seriation network across epochs of learning.

Three of the 20 networks progressed through stages 1, 2, and 4. These nets showed some stage 3 behavior, but it was too weak to be considered as consistent. Three nets progressed through stages 1, 2, and 3. These networks often showed an early but brief stint of stage 4 behavior followed by regressions either to stage 3 or to even lower level stages. This suggests that over training may have occurred. Five networks progressed through stages 1, 3, and 4. Again there was evidence of stage 2 behavior in some of these nets, but it was not sufficiently marked to constitute a stage. The remaining two nets only showed strong behavior in two of the four possible stages. That is not to say that other stages were not represented, but that they were briefly present and only in co-occurrence with another form of behavior.

A key question that remains to be addressed is whether these networks respond to perceptual variations in the same way as children do. To test this, we ran three additional simulations in which the

---

[2] The unbiased population of arrays consists of 151 (21%) low disorder patterns and 569 (79%) high disorder patterns.

[3] This array is maximally disordered from both decreasing and increasing series (Retschitzki, 1978).

training set consisted only of the 100 six element arrays described earlier. The three simulations differed only in the size of the difference between adjacent ordered sticks: 1.0, 0.5, or 0.25. The proportions of 20 networks able to complete a full sort by the end of training (i.e., be diagnosed as stage 3 or 4) were 0.85, 0.55, 0.15, respectively. Thus, as with children, the more easily the stick sizes can be distinguished, the better the network's seriation performance.

Further, 85% of the nets trained with 1.0 size differences were diagnosed at stage 4, as compared to only 25% of the nets trained with 0.5 size differences. This supports Piaget's (1965) view that stage 3 performers would be classed at stage 4 as size differences between sticks increase.

In order to obtain a better idea of how the information from the input array is processed by the network to select and move a stick, Hinton diagrams were generated at selected epochs that represented consistent stage behavior. These diagrams provide a visual display of the strength and sign of the weights in the network. Hence, it is possible to follow the weight evolution which leads to the particular solution settled on by the learning algorithm. Inspection of the Hinton diagrams revealed no drastic differences in weights between adjacent stages. Instead, stage differences were marked by rather small modifications in the sizes of weights.

The Hinton diagrams also revealed that the building of a representational structure in the network began by adjusting weights leading to those units dealing with the short end of the series and was progressively extended along the length of the series until finally appropriate weights were found for those units coding the larger end of the series. This is consistent with Trabasso's (1977) suggestion that children build a linear order mental representation of the seriation task by proceeding from the ends of the series inwards.

Discussion

This work contributes to the view that connectionist methods can be successfully applied to the study of cognitive development. The present models capture stage transitions and perceptual effects in seriation, both of which had eluded previous rule-based models.

This approach to seriation highlights features of the network and environment which have repercussions for understanding children's ability to seriate. A closer examination of the assumptions of such models might identify principles which could constitute the beginnings of a novel theory of cognitive development, including seriation development.

Two key facets of the learning environment can be stressed. Our most realistic models are biased towards lower disorder stimuli in the learning environment. There is also a bias towards learning from small sets of elements. A precocious ability to seriate small arrays could produce the empirical tries characteristic of stage 3 on larger arrays if children were applying the operational method to adjacent subsets of the series, rather than to the entire series.

With regard to the processing involved in seriation, our model points to a modularization of the task into two independent sub-tasks. One module is devoted to identifying which element in a set needs to be moved, whereas the other is devoted to identifying where an element should be moved. The observed macroscopic behavior is therefore a result of the interaction of the developmental states of these two modules.

In contrast, pilot seriation simulations with a non-modular net failed to both learn seriation and capture psychological regularities. Ten to 20 hidden units were often recruited during learning, yet these nets failed to develop past stage 2 performance. The success of a modular approach can be attributed to functional decomposition (i.e., the elementary components of a complex function tend to be easier to learn than the complex function itself) and to greater generalizability due to the constraints built into the architecture (Jacobs, Jordan, & Barto, 1991).

The present seriation simulations were the first of our cascade-correlation simulations of cognitive development to require modularity. The other simulations generated psychologically realistic data with a single, non-modular network. We expect to have to resort to more modularity as we simulate larger and more difficult developmental problems.

Our present network model also shows that the processing involved in seriation can be achieved using only brain style computational methods. The underlying principles involved are those of activation and inhibition rather than the explicit application of symbolic rules or application of the logical principles assumed by Piaget.

Analysis of Hinton diagrams of network weights reveals that seriation development is essentially due to an increase in the precision of processing rather than to any fundamental reorganization of processing. This finding is radically opposed to Piaget's initial conclusions, but conforms to Koslowski's (1980) suggestion concerning developmental increases in the precision with which systematic sorting is applied.

This gradualism stands in contrast to some other cascade-correlation models in which macroscopic changes in stage are due to underlying qualitative changes in network structure (Shultz & Schmidt, 1991). Some developmental changes appear to require qualitative changes in representational power, whereas others do not. This is one of several issues in cognitive development which can be investigated more easily in neural networks than in children. One advantage of employing generative, as opposed to static, networks in such simulations is that the necessity of qualitative change can be assessed.

Finally, the zeitgeist of the connectionist approach focuses our attention on the need for a tighter integration of perceptual and cognitive factors when studying cognitive development. The perceptual effects simulated in our seriation model are reminiscent of other perceptual effects found in cognitive developmental simulations (Shultz & Schmidt, 1991). Such perceptual phenomena seem particularly difficult for symbolic rule-based models to explain. Connectionist simulations are showing that many of these perceptual effects can be handled within the same processing system that generates rule-like and stage-like behaviors.

Insights generated by these simulations will, of course, need to be tested in research with children.

Acknowledgements

We are grateful to Scott Fahlman for providing the cascade-correlation code and to Sheldon Tetewsky and Dave Buckingham for helpful comments on an earlier draft of this paper.

References

Anzai, Y. 1987. Doing, understanding, and learning in problem solving. In D. Klahr, P. Langley, & R. Neches. *Production system models of learning and development*, 55-97. Cambridge, MA: MIT Press.

Atkins, M. 1989. Sorting by Hopfield net. *Proceedings of the International Joint Conference on Neural Networks* 2:65-68.

Bates, E. A., & Elman, J. L. 1992. Connectionism and the study of change. Technical Report 9202, Center for Research in Language, University of California, San Diego, CA.

Baylor, G. W., Gascon J., Lemoyne G., & Pothier, N. 1973. An information-processing model of some seriation tasks. *Canadian Psychologist* 14:167-196.

Chen, W., & Hsieh, K. 1990. A neural sorting network with O(1) time complexity. *Proceedings of the International Joint Conference on Neural Networks* 1: 87-95.

Elkind, D. 1964. Discrimination, seriation, and numeration of size and dimension differences in young children: Piaget replication study VI. *Journal of Genetic Psychology* 104:256-296.

Fahlman, S. E. 1988. An empirical study of learning speed in back-propagation networks. Technical Report, CMU-CS-88-162, School of Computer Science, Carnegie Mellon University, Pittsburg, PA.

Fahlman, S. E. & Lebiere, C. 1990. The Cascade-correlation learning architecture. Technical Report, CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, Pittsburg, PA.

Frey, L. 1964. Sériation et transitivité. *Cahiers de psychologie* 7:143-157.

Inhelder, B., & Piaget, J. 1969. *The early growth of logic in the child.* New York: Norton Library.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. 1991. Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science* 15:219-250.

Kingma, J. 1982. A criterion problem: The use of different operationalizations in seriation research. *Perceptual and Motor Skills* 55:1303-1316.

Kingma, J. 1983. Some behavioral characteristics of the partial seriators reconsidered. *Journal of General Psychology* 108:231-247.

Kingma, J. 1984. The influence of task variation on seriation research: Adding irrelevent cues to the stimulus materials. *Journal of Genetic Psychology* 144: 241-253.

Knuth, D. E. 1973. *Sorting and searching*. Reading, MA: Addison-Wesley.

Koslowski, B. 1980. Quantitative and qualitative changes in the development of seriation. *Merrill-Palmer Quarterly* 26:391-405.

Piaget, J. 1965. *The child's conception of number.* New York: Norton.

Plunkett, K., & Sinha, C. 1992. Connectionism and developmental theory. *British Journal of Developmental Psychology* 10:209-254.

Retschitzki, J. 1978. L'évolution des procédures de sériation: Etude génétique et simulation. *Archives de Psychologie* 46, Monographie 5.

Shultz, T. R. 1991. Simulating stages of human cognitive development with connectionist models. In L. Birnbaum, & G. Collins Eds., *Machine learning: Proceedings of the Eighth International Workshop*, 105-109. San Mateo, CA: Morgan Kaufman.

Shultz, T. R., Buckingham, D., & Oshima-Takane, Y. 1993, in press. A connectionist model of the learning of personal pronouns in English. In S. J. Hanson, M. Kearns, T. Petsche, & R. L. Rivest (Eds.), *Computational learning theory and natural learning systems*. Cambridge, MA: MIT Press.

Shultz, T. R., & Schmidt, W. C. 1991. A cascade-correlation model of balance scale phenomena. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, 635-640. Hillsdale, NJ: Erlbaum.

Shultz, T. R., Zelazo, P. R., & Strigler, D. 1991. *Connectionist modeling of the development of the concepts of potency and resistance in causal prediction.* Paper presented at the meeting of the Society for Research in Child Development, Seattle.

Trabasso, T. 1977. The role of memory as a system in making transitive inferences. In R. U. Kail, & J. W. Hagen Eds., *Perspectives on the development of memory and cognition*, 333-366. Hillsdale, NJ: Erlbaum.

Young R. 1976. *Seriation by children: An artificial intelligence analysis of a Piagetian task.* Basel: Birkhauser.