Simulating Development by Modifying Architectures

Gary Jones (gaj@psychology.nottingham.ac.uk)
Frank E. Ritter (frank.ritter@nottingham.ac.uk)
ESRC Centre for Research in Development, Instruction and Training
Department of Psychology; University of Nottingham
Nottingham, NG7 2RD (UK)

Abstract

In order to ground our understanding of cognitive development we have started to create a model of how children and adults solve a well-studied three-dimensional puzzle. We started with a model that fits the adult behaviour on the puzzle. We then modified the model's cognitive architecture (ACT-R) and its perceptual/motor architecture (the Nottingham Interaction Architecture) in three ways to simulate a younger problem solver by: (a) reducing the accuracy of vision, (b) reducing working memory, and (c) doing both. The modifications, particularly reduced working memory (and its combination with reduced visual accuracy), allow the model to approximate, on some measures, the behaviour of seven year olds on the puzzle. The results suggest that cognitive models and their architectures can help answer the question of "What develops?".

Introduction

As children develop, they are more able to learn new strategies and tasks, and become more efficient at the strategies and tasks they already know (e.g. Siegler, 1986). What changes occur in order for this to happen? It has long been noted that it would be useful to be able to specify in information processing terms how the behaviour seen at each age is achieved, and therefore what the differences are between ages (e.g. Simon, 1962).

Computational modelling allows behaviour to be specified in information processing terms. The first step in computational modelling across ages is defining the behaviours that occur at each age. A model that matches the observed behaviour can suggest what knowledge and procedures children may be using. To the extent that the behaviour has not or cannot be measured, the model can make predictions about the missing elements. This provides a way to examine to what extent changes in task performance can be attributed to differences in knowledge and to what extent it can be attributed to differences in processing due to development.

A Task to Study Development

The Tower of Nottingham (Wood & Middleton, 1975) task is to build a pyramid from 21 wooden blocks (shown in Figure 1). There are six layers to the pyramid; the lower five consist of four blocks, with a single block as the top layer. The blocks in the lower layers all share the same characteristics, differing only in size. Each layer is normally formed via two sets of paired blocks. For example, placing the peg of block A into the hole of block B brings the two half holes together to form a pair having a hole (a hole-pair). Similarly, placing block C and block D together forms a pair with a peg (a peg-pair). A layer is then formed by fitting the peg of the peg-pair into the hole of the hole-pair. Other strategies for creating a layer also exist, such as fitting a half peg into a half hole to form a pair having two pegs (blocks A and C), and a pair having two holes (blocks B and D).

Why Use This Task?

The Tower of Nottingham is a physical problem solving puzzle. A detailed analysis of task behaviour is possible via videotape. Many strategies are readily visible, reducing the need for the experimenter to infer mental structures and strategies. This enables a more accurate computational model to be created.

The Tower is also a suitable task in which to study development, because performance improves with age. Older children accomplish more correct operations, produce less errors, and take less time than their younger counterparts (Murphy & Wood, 1981; Wood & Middleton, 1975). This allows the study of children's problem solving behaviours on the Tower across ages. We start to develop here a way to specify how these behaviours are generated.

Simulating Development

There are two natural ways to create a series of developmental models. One way is to model a lower performance level (that of children) and modify the model to fit higher performance levels. The other way is to begin at the highest performance level (that of adults), and then modify the model to fit lower performance levels. We have chosen to start with the simpler (adult) behaviour and work towards the more chaotic (child-like) behaviour.

We will attempt to model behaviour (including learning) at different ages by making modifications within a fixed architecture (i.e. the terms of reference for the

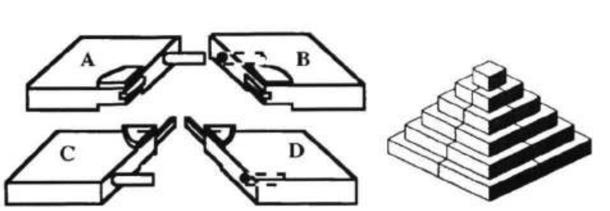

Figure 1: The blocks, left, that make up each layer, which are then stacked to create a tower, right.

changes are determined by the architecture). There are few previous models that suggest what changes to make to this architecture. In fact, the first definitions and implementations of cognitive architectures stressed that architectures do not change across tasks (Newell, 1990, p. 81). Most models of child development model development as just learning (e.g. McClelland & Jenkins, 1991; Siegler & Shipley, 1995) or differences in knowledge (Klahr & Wallace, 1976; Young, 1973).

The model of adult behaviour which we begin with has been improved since it was first reported (in Jones & Ritter, 1997). The model now fits the adult data better, and can be modified in more theoretically motivated ways. We will modify the model and architecture in three ways. For each of the changes we describe why they are suggested by developmental data, how they have been implemented (either in ACT-R or the Nottingham Interaction Architecture), and the effect the change has on the model's behaviour.

The Simulation and the Adult Model

Figure 2 shows the Tower simulation, which is written in Garnet (Myers, et al., 1990). The simulation contains a full graphical representation of the task (all blocks and features), which is 2 1/2 dimensional—blocks cannot be turned on their side or held in mid-air, but can be face-up or face-down.

The simulation also includes an eye and two hands. The eye and hands, as an instantiation of the Nottingham Interaction Architecture, are designed to meet a set of requirements identified for creating a psychologically plausible architecture for interacting with multiple external tasks (Baxter & Ritter, 1996). The architecture has been implemented several times in a variety of user interface management systems and graphics programming languages (Baxter et al., 1997). The eye is able to saccade and fixate, and passes to the model what blocks and constructions it sees (e.g. a peg-pair will be represented as a construction of two blocks flush on their outer edges with their quarter circles and half pegs aligned).

The visual information passed to the model depends upon where blocks are positioned in relation to the simulated eye. Three areas are defined: fovea, parafovea, and periphery. Full information is passed for blocks or features in the fovea and parafovea, though the parafovea adds noise to features and to block sizes. For items in the periphery, the eye only returns a block ID. The hands are able to pick up, drop, rotate, turn over, fit, and disassemble blocks.

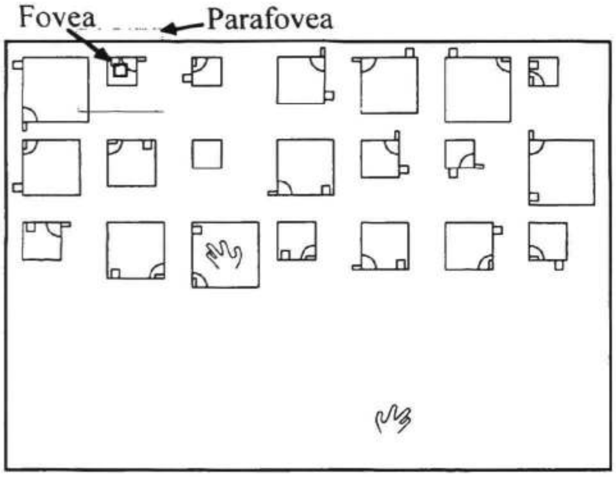

Figure 2: The Tower of Nottingham simulation.

The cognitive aspects of the models are based on the ACT-R cognitive architecture (Anderson, 1993). The adult model contains 264 ACT-R rules including task knowledge and knowledge about how to direct the eye and the hands. Within the model, all blocks and block features have an associated activation level. When several of the same rule are instantiated in ACT-R, the one with the highest activation is selected. In general, rules fire whose conditions have the most active blocks and block features in them. The activation levels are subject to *decay* each cycle. When they fall below a specified level (the *retrieval threshold*) they can no longer be matched in rule conditions. Activation is raised based on the current goals of the model and by what blocks the model is currently focusing on.

There are two learning mechanisms in the model. First, a simple method of increasing the chances of choosing to fit blocks by specific features if a previous fit using the same features was perceived to be a success. Success is determined by the blocks in the construction being flush on their outer edges and having their quarter circles aligned (this is consistent with adult data on the task). On some occasions the model may believe a successful construction has been made when in fact it has not (e.g. aligning the quarter circles of blocks A and B but leaving the blocks unconnected). This learning mechanism helps the model to fit adult data (Jones & Ritter, 1997). Second, the model learns task specific knowledge based on the perceived success. If, for example, the model fitted a peg in a hole, and deemed this successful, then new declarative knowledge is acquired noting that fitting pegs in holes is a task appropriate behaviour. This can then be used to direct later fit attempts.

The model contains working memory (in ACT-R) and visual memory (in the Nottingham Interaction Architecture). Working memory contains all blocks and block features that are active enough to be matched by rules (i.e. their activation is above retrieval threshold). The number of elements in working memory varies based on the activation of the blocks and block features. Visual memory allows the model to remember the details of some of the blocks that have been looked at previously even though they are now in the periphery. Visual memory is static (it is set at seven items), and complements working memory because blocks in visual memory that are not in working memory can also be matched in conditions of rules.

Comparison of the Model to Adult Data

We use the ACT-R default timing of 50 ms per production firing. For eye movements (saccades), this timing is consistent, but we increase it to 250 ms for productions involving fixations (Baxter & Ritter, 1996). For motor actions (fitting, rotating, moving, and disassembling blocks) we increase the timing to 550 ms (Jones & Ritter, 1997). Thus, the time predictions made by the model are absolute predictions; they are not scaled or otherwise adjusted in the analyses and graphs reported below.

Production firing latencies in ACT-R also take into account activation of memory elements. In order for the influence of memory elements on production firing latencies to be negligible (since we manipulate activation as part of the learning mechanism), the base level activation of memory elements was set to 10.0 (the default is 0.0). Other applicable ACT-R parameters (e.g. retrieval threshold) were set to the suggested default settings.

The adult model begins with the initial knowledge of the task that subjects had, such as blocks of the same size go together, pegs go in holes, checking that intended constructions are flush on their outer edges, and so on. This knowledge was taken from video analysis of adult behaviour (e.g. adults produce 33 peg-pair or hole-pair constructions, 13 two-peg or two-hole constructions, and never consider producing constructions that are not flush on their outer edges).

The adult subjects (N=5, from Jones & Ritter, 1997) had completed the task once. The ACT-R activation noise parameter was set to 0.05 (the default) so that noise is added, on each cycle, to the activation of constructions and features. This causes behaviour to vary across trials. We therefore compare 5 runs of the model to the 5 adult subjects.

The adult model fits the adult subjects reasonably well as shown in Table 1. The model also fits the adult subjects on other measures, such as strategies, but we do not consider these other measures when comparing to seven year olds later on. The reliable differences we report (using t-tests) for times and errors between the model and subjects are for aiding the reader in understanding the variances in familiar terms. They are not offered in any way as a proof of the model.

We include layer-by-layer analyses because there is within-task learning. Figure 3 compares the times to complete each layer for the adult model and the adult subjects ($r^2=0.98$; RMS error=4.5 s). The construction attempts per layer by the adult model is also similar to adult subjects ($r^2=0.79$; RMS error=0.15), although we think this could be improved.

Table 1: Mean measures (standard deviations) for adult model and adult subjects. Reliable differences between model and subjects ($p<.05$) are indicated with an *.

Measure	Adult Subjects	Adult Model
Completion time	80.6 s* (13.3)	97.6 s (8.0)
Number of errors	0.2 (0.5)	1.0 (1.2)

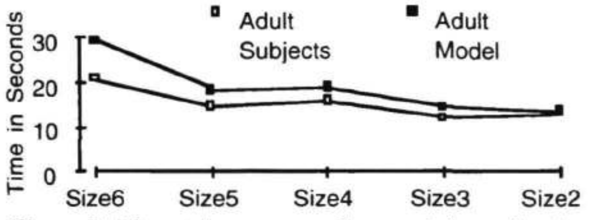

Figure 3: Time taken to complete each layer for the adult model and the adult subjects.

An exact fit on every measure is not essential because we are more interested in how the behaviour changes as a result of our alterations. We can therefore break the model to see to what extent we can account for the behaviour of younger children (seven year olds) on the Tower of Nottingham.

Creating a Child Model

In order to begin examining how problem solving could change with development, we created three new versions of the adult model. Each new version is created by making one or more modifications to the original adult model. The modifications are perhaps the most plausible based on developmental literature and our knowledge of children's performance on this task. They are also useful explorations in how to implement development within a fixed architecture. We: (a) altered the accuracy of size information in the parafovea; (b) reduced the working memory capacity; (c) combined the two. There are further changes that should be explored as well, such as basic processing speed and fovea size. Removing or changing knowledge is also possible, and an initial attempt is reported in Jones and Ritter (1998).

We will compare the modified models against seven year olds (N=5). The children were assisted on their first attempt at completing the Tower (contingently tutored: Wood & Middleton, 1975), and so we re-analysed the performance of their second, unsupported attempt. Children's behaviour on the task is more varied than adults. One seven year old in our analyses takes a lot longer to produce the size 5 layer than the other children. We leave this child in our analyses, but note that further children will have to be analysed so that the full range of seven year old behaviour can be identified and examined.

Reduced Parafovea Accuracy (RPA) Model

Why Children find it difficult to select blocks by size in the Tower of Nottingham task (Murphy & Wood, 1981). The seven year olds attempted 1.8 constructions, on average, involving different sized blocks; the adults do not attempt any constructions involving different sized blocks. Reducing the accuracy of the parafovea is one manipulation that may cause this effect.

How The parafovea noise parameter for size in the Nottingham Interaction Architecture was set to 30% (from 0%). This represents a 30% chance that a block in the parafovea will have a perceived size different to its actual size.

Predicted Effect The increased noise for block size should lead to incorrect constructions attempts involving blocks of different sizes. These errors should also increase the time taken to construct each layer.

Table 2: Comparison between the RPA model and the seven year old subjects. Reliable differences between model and subjects (p<.05) are indicated with an *.

Measure	RPA Model	7 yo Subjects
Completion time	109.6 s* (11.8)	214.4 s (95.8)
Number of errors	1.2 (1.3)*	7.6 (2.4)

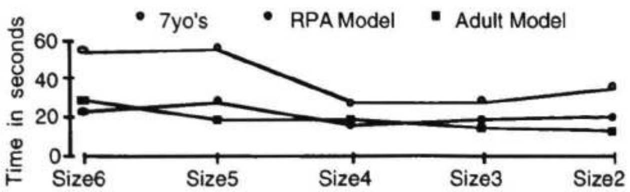

Figure 4: Time taken to complete each layer for the RPA model and the seven year olds.

Effect Table 2 shows the summary statistics for the RPA model and the seven year old subjects. Figure 4 shows time comparisons on a layer by layer basis.

The results go against our main prediction that there will be a greater number of constructions made involving blocks of different sizes (neither the original adult model or the RPA model produce any). In hindsight, this happens because when picking up a block, the model fixates upon it. At this point the block is in the fovea, the correct size is returned, and if the block is the wrong size it is replaced. The time taken to complete the task increases because incorrect blocks are picked up (but are not assembled). This can be quantified in the number of fovea fixations (105.0 for the RPA model versus 92.4 for the adult model) and the number of blocks picked up (36.8 versus 31.0). The RPA model correlates well with the seven year olds (r^2=0.93) in time taken to construct each layer, but the timings are well below that of the subjects (RMS error=21.1 s).

Reducing the parafovea size accuracy makes an interesting prediction. It suggests that seven year olds either do not examine the block again once they have decided to pick it up or their fovea vision is not as accurate as adults. The visual strategies employed by adults may differ from those of children.

Reduced Working Memory Capacity (RWMC) Model

Why Several developmental theories suggest working memory capacity may influence task performance (e.g. Case, 1985; Halford, 1993). On the Tower of Nottingham, children have been noted to search with replacement (D.Wood, personal correspondence, 1996), a characteristic which may well be linked to working memory if children forget which blocks they have tried to fit. This behaviour would lead to more incorrect constructions using the same blocks. On the Tower, seven year old children fit the same blocks together an average of 3.7 times, whereas this behaviour never occurs in adults.

How Our model provides an easy way to manipulate working memory to see how it influences performance. We implemented a reduced working memory capacity by altering three parameters (the first two are parameters in ACT-R and the third is a parameter in the Nottingham Interaction Architecture): (a) Increasing the retrieval threshold requires constructions to be higher in activation in order to be matched by rules. (b) Increasing the decay rate means constructions are forgotten more quickly. (c) Reducing the number of visual memory items means that visual memory provides less support to working memory.

In order to explore the effect of these parameters we used two sets of values. First (the RWMC1 model), we increased the retrieval threshold from 0.0 to 2.5, increased the decay rate from 0.05 to 0.15, and reduced the number of items in visual memory from 7 to 3. Second (the RWMC2 model), we further increased the retrieval threshold and decay rate (to 4 and 0.225 respectively) and further reduced the number of visual memory items (to 1). The ACT-R working memory mechanisms that we manipulate have also been used by Lovett, Reder and Lebiere (1997) in their ACT-R model of working memory differences.

Predicted Effect Less working memory should lead to more search with replacement—the same pairs of blocks should be fit together more often. This should mean the task will take longer and involve more errors.

Effect Table 3 shows summary statistics for the RWMC models and seven year old subjects. Figure 5 shows time comparisons on a layer by layer basis.

Reducing the working memory capacity in the adult model does not lead to fitting the same blocks together more often (0 for both RWMC models). However, when we include "mental" fits (checking that intended constructions are flush on their edges without actually fitting the blocks), fitting the same blocks is considered

Table 3: Comparison between the RWMC models and seven year old subjects. Reliable differences between model and subjects (p<.05) are indicated with an *.

Measure	RWMC Model 1	RWMC Model 2	7 yo Subjects
Completion Time	117.6 s (27.5)	193.8 s (59.9)	214.4 s (95.8)
Number of Errors	0.8 (0.8)*	1.4 (0.9)*	7.6 (2.4)

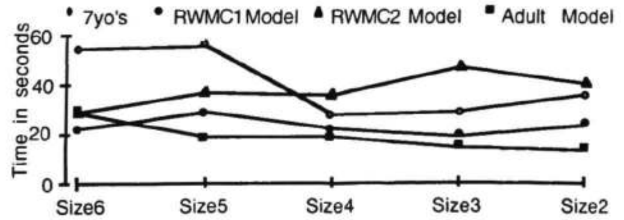

Figure 5: Time taken to complete each layer for the RWMC models and the seven year olds.

3.5 times for the RWMC1 model, and 11.8 times for the RWMC2 model, compared with 2.4 times for the adult model.

Although construction time increases, errors remain fairly constant because of the mental fits, since potential constructions must be flush on their outer edges (since adults never produce any constructions that are not). On a layer by layer basis, the RWMC models have clear differences from the seven year old data (RMS error of 20.4 s and 17.0 s for the RWMC1 and RWMC2 models respectively).

Reducing the working memory capacity has allowed the model to move closer to the seven year old data for overall time taken, but at the same time reducing the fit on a layer-by-layer basis. The RWMC models (particularly RWMC2) do not show a steady decline in time for each subsequent layer constructed. We suggest this is because the model does not employ a strategy of gathering same sized blocks together (subjects often do). This does not affect behaviour for the adult model. However, when working memory is reduced it means that the further into the task, the more spread out the blocks become, and the greater the reliance is upon working memory to remember block details. This increases visual search, which is why the time rises for later layers (there are 137 more fixations for the last three layers in the RWMC2 than for the adult model).

Reduced Parafovea Accuracy and Reduced Working Memory Capacity (RPA-RWMC) Model

Why We have shown that manipulating single variables (at least these variables) is not enough to fit the behaviour of seven year olds on the Tower of Nottingham. Altering more than one variable at a time could lead to more interesting behaviour as it allows interaction between variables.

How We included both changes—the reduced parafovea accuracy and reduced working memory capacity. For the working memory changes, we used the average of the alterations in RWMC1 and RWMC2 (retrieval threshold of 3, decay rate of 0.2, and 2 visual memory items).

Table 4: Comparison between the RPA-RWMC model and 7 yo subjects. Reliable differences between model and subjects ($p<.05$) are indicated with an *.

Measure	RPA-RWMC Model	7 yo Subjects
Completion time	148.2 s (31.4)	214.4 s (95.8)
Number of errors	1.6 (1.5)*	7.6 (2.4)

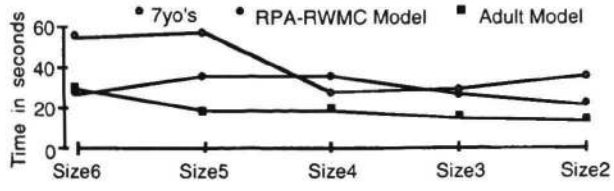

Figure 6: Time taken to complete each layer for the RPA-RWMC model and the seven year olds.

Predicted Effect The behaviour of the model should be consistent with that of the independent reduced parafovea accuracy and reduced working memory capacity models, although we expect that the effects will interact.

Effect Table 4 shows summary statistics for the RPA-RWMC model and the seven year old subjects. Figure 6 shows time comparisons on a layer by layer basis. The RPA-RWMC model takes longer than the adult model, but not long enough. It does not produce enough errors either. When examined on a layer-by-layer basis, the RPA-RWMC model's timings do not fit the seven year olds ($r^2=0.19$; RMS error=17.1 s).

The interaction of the reduced parafovea accuracy and reduced working memory has not lead to a substantially different model over and above the respective individual models. Further modifications and interactions will need to be explored. Particular emphasis should be aimed at alterations that increase the errors that the model makes.

Summary

We took an initial adult model and modified it to simulate a younger problem solver: perceptually (reducing parafovea accuracy), cognitively (reducing working memory capacity), and in combination. These changes impaired the performance of the model to differing degrees and in different ways. None of the alterations was sufficient to produce behaviour similar to seven year olds.

These results suggest that simple changes may not be enough. The most modified adult model is still some distance from appearing like a younger problem solver. However, the changes have highlighted behaviours of the model that should be investigated further, such as the fixation strategy when a block is picked up, and the strategy of collecting together all same-size blocks. The modifications also indicate that the two learning mechanisms used may not be sufficient to model children: for most of the modifications made, the layer-by-layer trend does not match the children.

Although the changes have not enabled the modified models to fit the seven year old data, there are many other modifications that we can make to both the Nottingham Interaction Architecture and the model (such as changes in processing speed). Making these changes and examining their effects is now straightforward.

We will also have to look further at combinations of modifications. Interactive effects should reveal more about performance at different ages, although independent changes are still useful for our understanding of the effects they have.

This work will eventually lead to models of how five and seven year olds solve the Tower that are based on the adult model. We hope that these models will be able to explain individual differences within age groups as well as to explain the progression between ages (in terms of differences between the models rather than transition mechanisms, for the moment). In both cases, we should be able to highlight the knowledge differences or architectural changes that lead to the differences in behaviour. Further learning mechanisms are also required. The models must learn on a layer-by-layer basis but the learning mechanisms must not be impeded by the modifications that are made to the models (as they seem to be for the modifications reported).

We are now in a position to look at how problem solving changes across development on this task. We have a cognitive model that performs the task. We can add and remove knowledge from the cognitive model and we can modify the architecture to represent developmental changes in cognition (the cognitive model based in ACT-R) and perception (the Nottingham Interaction Architecture). This type of work will help us more directly answer the question "What develops?".

Acknowledgements

Josef Nerb wrote the first tower simulation. Heather Wood provided the seven year olds' data and the Tower picture. We thank David Wood for useful discussions, and Richard Young and Fernand Gobet for useful comments.

References

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: LEA.

Baxter, G. D., & Ritter, F. E. (1996). Designing abstract visual perceptual and motor action capabilities for use by cognitive models (Tech. Report No. 36). ESRC CREDIT, Psychology, U. of Nottingham.

Baxter, G. D., Ritter, F E., Jones, G., & Young, R. M. (1997). Exploiting user interface management systems in cognitive modelling. Manuscript submitted for publication.

Case, R. (1985). *Intellectual development: A systematic reinterpretation.* New York: Academic Press.

Halford, G. S. (1993). *Children's understanding: The development of mental models.* Hillsdale, NJ: LEA.

Jones, G., & Ritter, F. E. (1997). Modelling transitions in children's development by starting with adults. In *Proceedings of the European Conference on Cognitive Science (ECCS '97).* 62-67. Manchester, UK: AISB.

Jones, G., & Ritter, F. E. (1998). Initial explorations of modifying architectures to simulate cognitive and perceptual development. In *Proceedings of the Second European Conference on Cognitive Modelling,* 44-51. Nottingham, UK: Nottingham University Press.

Klahr, D., & Wallace, J. G. (1976). Cognitive development: An information processing view. Hillsdale, NJ: LEA.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1997). Modeling individual differences in a digit working memory task. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society.* 460-465. Mahwah, NJ: LEA.

McClelland, J. L., & Jenkins, E. (1991). Nature, nurture and connections: Implications of connectionist models for cognitive development. In K. VanLehn (Ed.), *Architectures for intelligence,* Hillsdale, NJ: LEA.

Murphy, C. M., & Wood, D. J. (1981). Learning from pictures: The use of pictorial information by young children. *Journal of Experimental Child Psychology, 32*, 279-297.

Myers, B. A., et al. (1990). Garnet: Comprehensive support for graphical, highly-interactive user interfaces. *IEEE Computer, 23*(11), 71-85.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Siegler, R.S. (1986). *Children's thinking: What develops?* Englewood Cliffs, NJ: Prentice-Hall.

Siegler, R. S. & Shipley, C. (1995). Variation, selection and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence.* Hillsdale, NJ: LEA.

Simon, H. A. (1962). An information processing theory of intellectual development. *Monographs of the Society for Research in Child Development, 27* (2, Serial No. 82).

Wood, D., & Middleton, D. (1975). A study of assisted problem solving. *British Journal of Psychology, 66*(2), 181-191.

Young, R. M. (1973). Children's seriation behavior: A production-system analysis. Unpublished doctoral dissertation, Carnegie-Mellon University.