UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Generative Connectionist Model of the Development of Rule Use in Children

Permalink

Journal

Authors
Marcovitch, Stuart
Zelazo, Philip David

Publication Date
2000

Peer reviewed

A Generative Connectionist Model of the Development of Rule Use in Children

Stuart Marcovitch (stuartm@psych.utoronto.ca)
Philip David Zelazo (zelazo@psych.utoronto.ca)
Department of Psychology, University of Toronto
Toronto, ON, M5S 3G3, Canada

Abstract

The cascade correlation algorithm (CASCOR), a generative connectionist model, was used to simulate age-related changes on the dimensional change card sort (DCCS), which has traditionally been used to evaluate the complexity of children's rule-use abilities. Like 2.5-year-olds, inexperienced networks behave as if following one rule; slightly more experienced networks (akin to 3-year-old children) behave as if following a pair of rules; and the most experienced networks (akin to 5-year-olds) behave as if following two pairs of rules. Analysis of the networks' activation levels revealed that mastery of simple rules is a necessary precondition for using higher order rules. The model also generated four novel predictions that can be tested in future research with children.

Introduction

Since its inception, artificial intelligence has made a large impact on the field of psychology. The infusion of computer generated models into psychological research has become increasingly common. In the past decade, connectionist models have become particularly influential as a research tool in psychology. Connectionist models benefit psychology in three ways: (a) successful simulation requires formalization of the assumptions of the model, (b) analyzing the solution of a connectionist network may provide insight into the psychological mechanisms used, and (c) the model may generate novel (and often counter-intuitive) predictions. In particular, connectionist modeling used in conjunction with empirical research has the potential to shed light on patterns of development across a wide range of cognitive domains. Researchers in developmental psychology have already employed connectionist models to simulate developmental phenomenon in a variety of cognitive tasks (e.g., McClelland & Jenkins, 1991; Schultz, Schmidt, Buckingham, & Mareschal, 1995; see Elman et al., 1996, for a comprehensive review). Often, the results of these simulations call into question contemporary explanations of cognitive development.

According to Cognitive Complexity and Control theory (CCC; Frye, Zelazo, & Palfai, 1995; Zelazo & Frye, 1997), developmental improvements on tasks assessing deliberate reasoning and intentional action can be attributed to the acquisition of increasingly complex rule systems. Specifically, CCC postulates that young children (2.5 years) can use one rule, slightly older children (3 years) can use a pair of rules, while the oldest preschoolers (5 years) can use two incompatible pairs of rules. Rule-based card sorting paradigms have been employed to illustrate the number of
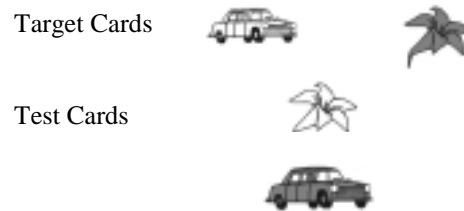

Figure 1: Stimuli for DCCS

rules that children can use. In these tasks, children are given cards that can be placed in one of two boxes based on a rule. For example, Zelazo, Reznick, & Piñon (1995) instructed 2.5-year-olds to sort pictures into categories such as things found inside the house versus things found outside. Typically, these children were able to sort the first card correctly, but then perseverated and sorted all subsequent cards in the same box. Thus, these results demonstrated that 2.5-year-old children could sort by one rule (e.g., if picture of things found inside the house then put card there), but not by a pair of rules (e.g., if picture of things found inside the house then put card here, but if picture of things found outside, then put card there).

The Dimensional Change Card Sort (DCCS; Frye et al., 1995; Zelazo, Frye, & Rapus, 1996) has also been used to reveal age-related changes in the number of rules children can use simultaneously. In the standard task, children are shown two *target cards* that differ on two dimensions, say color and shape (e.g., red car and blue flower). Children are presented with *test cards* that share one dimension with one target and the other dimension with the other target (e.g., red flower and blue car, see Figure 1). In the pre-switch phase, children are instructed to sort the test cards (i.e., match the test card to the appropriate target card) according to one rule (color or shape). After a predetermined number of pre-switch trials (e.g., 5, see Zelazo et al., 1996), children are asked to sort the same test cards by the other rule. So, the same test card will be sorted differently in the pre-switch and post-switch phases. On this task, 3-year-old children tend to pass the pre-switch phase, but fail the post-switch phase. This indicates that these children can sort by one pair of rules (e.g., in the color game, if it's red it goes here, but if it's blue it goes here), but not by two incompatible pairs of rules (e.g., if it's the color game, then if it's red it goes here, and if it's blue it goes here *but* if it's the shape game, then if it's a flower it goes here and if it's a car it goes here.) Five-year-old children tend to pass both the pre-switch and post-

switch phase, which illustrates that they can sort by two incompatible pairs of rules in the same context, and arguably requires the use of a higher order rule for selecting between pairs of rules.

The goal of the present study was to simulate the development of rule use in children using a generative connectionist model. Our study had three objectives: (a) to capture the age-related changes that are observed in children's sorting between the ages of 2.5 and 5 years, (b) to generate novel predictions, and (c) to explore what the internal structure of the connectionist networks reveals about the structuring of dimensions and features within the dimensions *vis à vis* success on the task.

In the present study, we used the cascade correlation learning algorithm (CASCOR; Fahlman & Lebiere, 1990) to simulate children's performance on the DCCS. Some researchers (e.g., Shultz, 1991) have suggested that CASCOR is appropriate in simulations of cognitive development because it embodies Piaget's principles of assimilation and accommodation. CASCOR is a generative algorithm that begins with connections between all the inputs and the output, but no hidden units. The model attempts to learn the training set in the constraints of this architecture, a phase akin to the Piagetian concept of assimilation. However, if the training set cannot be learned within a specific network architecture, hidden units are recruited as needed to increase computational power. Each hidden unit receives connections from all input units and all previously recruited hidden units. The restructuring of the network to create a more adaptive architecture is akin to the Piagetian concept of accommodation. One advantage of CASCOR is that the hidden unit chosen for recruitment is the one that will produce the lowest overall error. Consequently, the modified network is poised to solve the task at hand, and will do so more efficiently (using fewer hidden units) than networks with fixed architectures.

Training Phase

Age-related changes in the DCCS were simulated using CASCOR. The networks had 15 inputs. The first input determined the game that was to be played (color or shape). The next 12 input units determined the color and shape of the stimulus cards. Each card was coded across 4 attribute units (red, blue, car, flower). A value of 1.0 indicated the presence of an attribute while a value of 0.0 indicated the absence of the attribute. For example, the values {1.0, 0.0, 0.0, 1.0} indicated a red flower. The test card and the two target cards were each represented by specific configurations across the 12 units. The 14th and 15th units were context units, which determined if the network was learning in the *training* context {1.0, 0.0} or the *test* context {0.0, 1.0}. These context units were necessary to distinguish learning that occurred in the natural environment (training) from the laboratory environment (test). There was one output unit that returned a value ranging from -0.5 to 0.5. Matching to the first target card was assigned an output value of -0.5, whereas matching to the second target card was assigned an output value of 0.5. The target value that

was closest to the actual output value was considered the matching target.

In the training set, the network received a set of simple rules. The network was presented with the relevant game (e.g., color), a bidimensional test card (e.g., red flower), and two bidimensional target cards (e.g., a red car and a blue flower). For all the examples in the training set, the context units were set to the training context (i.e., 1.0, 0.0).

The network updated its weights based on a supervised learning algorithm. The network's output was compared to the expected output (i.e., in the color game, a red flower should be matched to the red car), and the weights were updated using the quickprop algorithm (Fahlman, 1988) and batch learning (i.e., the weights were updated after each epoch, as opposed to each example). Quickprop is a weight adjustment algorithm that is much quicker than backprop because it uses second-order (curvature) information as well as first-order (slope) information when adjusting weights, whereas backprop is restricted to slope information. Slope information indicates the direction of change; curvature information provides an index of the change in slope, which is used to determine the magnitude of weight change (Mareschal & Shultz, 1996; also see Fahlman, 1988, for more details).

In the training phase of the simulation, all possible training combinations were used. That is, 2 games (color or shape) X 4 test cards (red flower, blue flower, red car, blue car) X 4 target combinations (red flower, blue flower, red car, blue car for target 'A'; target 'B' differed from target 'A' on both dimensions), which yielded 32 training examples. Because the preliminary goal was to simulate data that were averaged over groups of children, a cross-sectional design was implemented as per previous studies of the DCCS with children (e.g., Zelazo et al., 1996). Twenty networks were trained in each of 5 conditions that differed on the number of epochs of training that the network experienced. The conditions were 50, 75, 100, 150, and 225 epochs.

Test Phase

After various amounts of exposure to the training set, training was halted so that the network could be tested. Testing consisted of changing the training set to five examples (pre-switch trials) that correspond to the five trials of the pre-switch phase of the DCCS. In all five trials, the network was presented with the same game (i.e., shape), the same two target cards (i.e., target 'A' was a red flower, target 'B' was a blue car), and the context nodes were set to the test context (i.e., 0.0, 1.0). The two possible test cards were presented (i.e., red car and blue flower) on alternate trials with one test card presented three times and the other test card presented twice. The network updated its connection weights after each pre-switch trial. After the fifth pre-switch trial, the network was tested on two post-switch trials. These were equivalent to the pre-switch trials, except now the network was asked to sort by the other dimension (e.g., color). The output revealed how the network sorted each of the two test cards. Because weights were *not*

updated in the post-switch phase, two post-switch trials were sufficient for the appropriate categorization of the network.

The network outputs were categorized into one of four categories based on criteria used with children (e.g., Zelazo et al., 1996):

(1) Fail Pre-Switch - The network incorrectly sorted on two or more pre-switch trials.
(2) Fail Post-Switch (same box) – The network passed the pre-switch phase, but incorrectly sorted on one of the two test trials in the post-switch phase (i.e., the network put all of the cards in the same box).
(3) Fail Post-Switch (perseveratively) – The network passed the pre-switch but incorrectly sorted both test cards in the post-switch phase (i.e., the network perseverated on the two original rules).
(4) Pass Post-Switch – The network correctly sorted both test cards in the post-switch phase.

Results

The CASCOR network began with the 15 input units and the one output unit. Although the network did not initially contain hidden units, these were recruited as needed through the progression of the simulation. The number of hidden units recruited was noted. The number of networks in each of the four classifications is displayed in Table 1.

Table 1: Performance of CASCOR networks on DCCS

	Categorization of Network			
No. of Epochs	FPre	FPost Box	Fpost Pers	Pass
50	12 (1*)	2 (1*)	0	6 (1*)
75	5 (5*)	7 (7*)	5 (5*)	3 (2*)
100	10 (10*)	2 (2*)	5 (5*)	3 (3*)
150	2 (2**)	1 (1*)	4 (2*, 2**)	13 (3*, 10**)
225	1 (1**)	0	0	19 (19**)

Note. FPre = Fail Pre-Switch; Fpost Box = Fail Post-Switch (same box); FPost Pers = Fail Post-Switch (perseveratively); Pass = Pass Post-Switch. The number of hidden units recruited by the networks is represented by asterisks (*). For example, 3* means three networks recruited one hidden unit, while 10** means 10 networks recruited 2 hidden units.

In the 50-epoch condition, 12 out of 20 (60%) of the networks failed the pre-switch phase. For the slightly more experienced network in the 75-epoch condition, 15 out of 20 (75%) of the networks passed the pre-switch phase. Furthermore, 12 out of 15 (80%) of those networks went on to fail the post-switch phase. In the 225-epoch condition, 19 out of 20 (95%) of the networks passed the pre-switch phase. All of those networks (100%) went on to pass the

post-switch phase. Overall, this pattern of results mirrored the pattern found in the empirical literature. Namely, the youngest children tend to fail the pre-switch phase, indicating failure to use a single pair of rules systematically. The slightly older children pass the pre-switch phase but fail the post-switch phase. Finally, the oldest children tend to pass both the pre-switch and post-switch phases, arguably indicating that they were capable of using a higher order rule for selecting between two incompatible pairs of rules.

Table 2: Number (and row percentages) of networks in each classification based on the number of hidden units.

	Categorization of Network			
No. of Hidden Units	FPre	FPost Box	Fpost Pers	Pass
0	11 (61%)	1 (6%)	0	6 (33%)
1	16 (33%)	11 (23%)	12 (25%)	9 (19%)
2	3 (9%)	0	2 (6%)	29 (85%)

Note. FPre = Fail Pre-Switch; Fpost Box = Fail Post-Switch (same box); FPost Pers = Fail Post-Switch (perseveratively); Pass = Pass Post-Switch.

The number of hidden units the network recruited seems to be related, albeit imperfectly, to performance on the DCCS. Table 2 displays the classification of networks across all five conditions based on the number of hidden units. A chi-squared analysis revealed a relation between the number of hidden units and the DCCS classification, χ^2 (6, $N = 100$) = 49.40, $p < 01$. The majority of networks with no hidden units fail the pre-switch phase, while the majority of networks with two hidden units pass both the pre-switch and post-switch phases. Networks with one hidden unit tend to be transitional and distributed across all four conditions. Thus, it can be argued that by acquiring more sophisticated internal representation (measured by the number of hidden units), more complex rules can be solved.

The current findings are congruent with Siegler's (1996) notion that cognitive development is driven by changes in strategy selection. According to this notion, children typically have a number of strategies available to them to solve any task. With age, the likelihood of selecting more appropriate strategies increases. However, even at older ages, children sometimes select inappropriate strategies. In the current simulations, increases in the number of hidden units may correspond to increases in the likelihood of selecting a more appropriate strategy. For example, networks with two hidden units usually adopt the most appropriate strategy (85% of the time), but occasionally adopt a less-appropriate strategy.

In addition to capturing the general pattern of age-related changes on the task, the simulations offer several predictions that raise interesting questions for future empirical work:

(1) In networks that passed the pre-switch phase but failed the post-switch phase, there was a developmental increase in the proportion that failed perseveratively (as opposed to sorting cards in the same box). In the four network conditions where these types of errors occur, the proportions that failed perseveratively were 0%, 42%, 71%, and 80%, for 50, 75, 100, and 150 epochs respectively. We expect a similar increase with children.

(2) The proportion of networks that passed the pre-switch phase followed a U-shaped developmental trajectory. The proportions in the network conditions were 40% at 50 epochs, 75% at 75 epochs, **50%** at 100 epochs, 90% at 150 epochs, and 95% at 225 epochs respectively. It is predicted that children will follow a similar U-shaped trajectory.

(3) The unexpected decrease in the proportion of networks that pass the pre-switch phase occurred in the same condition (100 epochs) as when the networks began to fail the post-switch phase perseveratively as opposed to putting the cards in the same box. Arguably, this occurred because the networks are beginning to categorize both dimensions simultaneously. This will lead to a decrease in performance in the pre-switch phase (sorting is more likely to be based on the wrong dimension), and an increase in perseverative errors in the post-switch phase (more likely to sort the cards according to the dimension that was previously correct). It is predicted that careful analyses of children's performance will reveal similar trends.

(4) Although 60% of the networks at 50 epochs failed the pre-switch phase, those that passed tended to pass the post-switch phase (6 out of 8, 75%). It is predicted that the youngest children (2.5-year-olds) who are able to pass the pre-switch phase will succeed in the post-switch phase. Perhaps these children have learned to sort a pair of rules, but fail to link the rules in the pre-switch to the rules in the post-switch. As a result, the post-switch phase is treated independently of the pre-switch phase, with a consequent absence of proactive interference.

Analysis of Network Activations

A primary benefit of connectionist simulations to cognitive psychology is the ability to analyze the internal representations of the networks. To that end, cluster analyses were carried out on the activations of the hidden units and the output node in the networks for each of the training examples. Figure 2 displays graphically the results from the analysis of one randomly selected network in the 225-epoch condition[1] (i.e., after the network had learned to sort successfully on both pre-switch and post-switch trials). Each training example is represented by a string of seven letters. The first letter denotes which game the network is required to play. The next six letters denote the test card, the first target and the second target respectively. Training examples that are clustered together elicit similar activation levels from the hidden units and the output. Because the features of the first target card necessarily determine the features of the second target card (e.g., red flower is always paired with blue car), only the first target card is discussed in the analysis.

As can be seen from Figure 2, group A contains all of the examples that have flowers both in the test card and in the target card. In contrast, all training examples that have cars in the test card and in the target card are in group B. Thus, the network appears first to discriminate, at least partially, on the basis of the shape dimension.

Group A (the flower group) can further be separated into 2 subgroups, C and D. Of all the test cards in group A, subgroup C contains all of the blue test cards, whereas subgroup D contains most of the red test cards (75%). Similarly, group B (the car group) can be further separated into subgroups E and F. Of all the test cards in group B, most of the blue test cards (80%) are in subgroup E, whereas most of the red test cards (75%) are in subgroup F. Therefore, once the shape dimension is established, the network appears to discriminate on the basis of color.

Correct performance on the DCCS requires more than successful categorization of the stimuli by the appropriate dimension. It is also necessary to categorize the stimuli by the type of game that is to be played. In Figure 2, all branches labeled G indicate the six places where this occurs. Based on the network's activation levels, we can speculate that success on the DCCS may first involve categorizing the stimuli by one dimension. Once this categorization has been established, the stimuli are then categorized by the other dimension. Only when both dimensions are appropriately categorized can a higher order rule that discriminates between the two dimensions, such as the type of game, be considered. This interpretation is consistent with CCC theory (Frye at al., 1995; Zelazo & Frye, 1997). For example, Zelazo (1999) suggested that success on the pre-switch phase of the DCCS requires the conjunction of two simple rules into a contrastive pair of rules. Each pair of rules must then be mastered before a higher order rule controlling their selection can be evoked. Without this higher order rule, children will select the rule that is most strongly associated with the given context (i.e., fail perseveratively on the post-switch phase).

Conclusions

In conclusion, the CASCOR simulations were successful in its three goals. First, the age-related changes on the DCCS task were simulated. Namely, inexperienced networks failed the pre-switch phase, slightly more experienced networks

---

[1] Cluster analyses on less experienced nets revealed similar patterns as the 225-epoch condition. However, the results were

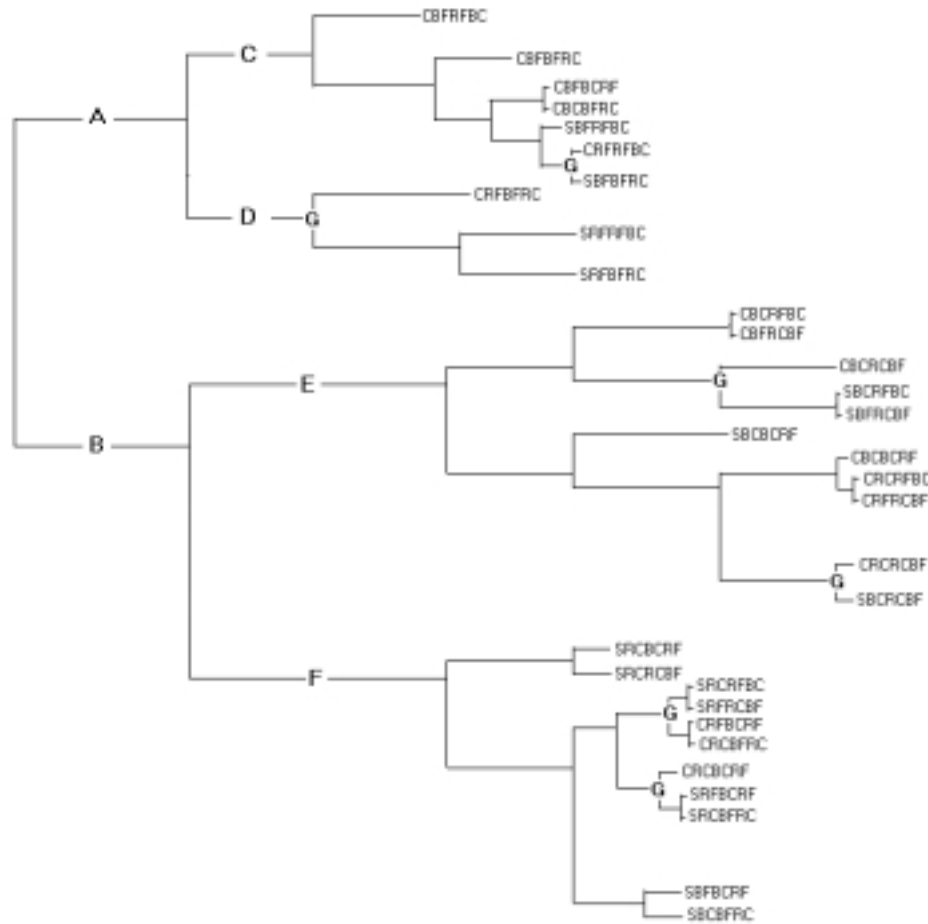more variable. It appears that experience stabilizes the clustering structure.

Figure 2: Cluster analysis on hidden and output unit activations of a randomly selected network in the 225-epoch condition.

passed the pre-switch phase but failed the post-switch phase and the most experienced networks passed both the pre-switch and post-switch phases. Second, novel predictions were generated and will be tested in future research. These include (1) an age-related increase in the number of children who fail the post-switch phase perseveratively (as opposed to sorting all the test cards in the same box), (2) a U-shaped developmental curve depicting performance on pre-switch trials, and (3) those very young children who pass the pre-switch phase will also pass the post-switch phase due to a relative lack of proactive interference. Third, cluster analyses on the hidden and output unit activations suggest that the formation of a higher order rule requires that the stimuli can be appropriately categorized by the appropriate dimensions. Further empirical research, coupled with modifications to modeling, hopefully will lead to an increased understanding of the mechanisms involved in the development of children's flexible rule use.

Acknowledgements

References

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fahlman, S. E. (1988). Faster-learning variations on back-propagation: An empirical study. In D. S. Touretzky, G. E. Hinton, & T. J. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School*. Los Altos, CA: Morgan Kaufmann.

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in neural information processing systems,* Vol. 2 . Los Altos, CA: Morgan Kaufmann.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483-527.

Mareschal, D., & Shultz, T. R. (1996). Generative connectionist networks and constructivist cognitive development. *Cognitive Development, 11,* 571-603.

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: Erlbaum.

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*. San Mateo, CA: Morgan Kaufmann.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Zelazo, P. D. (1999). Language, levels of consciousness, and the development of intentional action. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention*. Mahwah, NJ: Erlbaum.

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. Stamenov (Ed.), *Language structure, discourse, and the access to consciousness*. Amsterdam & Philadelphia: John Benjamins.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11,* 37-63.

Zelazo, P. D., Reznick, J. S., & Piñon, D. E. (1995). Response control and the execution of verbal rules. *Developmental Psychology, 31,* 508-517.