Simulating Stages of Human Cognitive Development with Connectionist Models

Thomas R. Shultz
Department of Psychology
McGill University
Montreal, Quebec, Canada H3A 1B1
shultz@hebb.psych.mcgill.ca

ABSTRACT

The psychological literature on stages of cognitive development was reviewed and found to contain support for the idea that stages represent ordinal, qualitative changes in organized knowledge structures. There was a lack of empirical support for the notions that stage transitions are abrupt and concurrent. All of these findings were found to be consistent with new connectionist models of cognitive development. A fundamental insight emerged from working with such models, namely, that stages result when a network solves part of a problem before solving all of the problem. Partial problem solving in connectionist networks is likely to occur under the following conditions: hidden unit herding, over-generalization, training pattern bias, and hidden unit recruitment.

1. INTRODUCTION

There are two major issues in cognitive developmental psychology: the question of what structures develop and the question of how developmental transitions occur. Although the two issues are equally important for a complete picture of development, perhaps less than 1% of the cognitive developmental literature has focussed on the problem of transition (Sternberg, 1984). The most commonly cited reasons for this lopsided focus are that: (a) issues of transition are still too difficult to solve, and (b) there should be common agreement on what structures develop before a concerted attempt at explaining transition becomes feasible.

Recent developments in machine learning have made it possible to think in detailed, mechanistic terms about transitions in cognitive development. Both connectionist (Chauvin, 1989; McClelland, 1988; Plunkett & Marchman, 1990) and symbolic level (Newell, 1990; Simon, Newell, & Klahr, in press) models of cognitive development are now starting to appear.

This paper examines connectionist models of human cognitive development with regard to how well they cope with the issue of stages. There is no question about whether connectionist algorithms are capable of learning complex input-output patterns from environmental feedback. But do these models acquire structures that correspond in realistic ways to what the literature purports to know about the stages through which children progress? Furthermore, do connectionist models add any interesting insights about the occurrence and properties of stages of development?

I start by reviewing the stage notion in developmental psychology, and then examine connectionist models of cognitive development for their ability to provide new insights into stages and cover existing stage phenomena.

2. STAGES IN COGNITIVE DEVELOPMENT

For most of the history of developmental psychology, associational and stage theorists offered conflicting metaphors for explaining the growth of thought. On the associational view, cognitive development occurred through the gradual accumulation of information. For the stage theorist, development occurred through the emergence of qualitatively different modes of thought. Later, I'll argue that more sophisticated associational devices, connectionist models, do capture some of the properties of stages.

The quintessential stage theorist in psychology is Piaget. For any particular cognitive achievement (number, space, time, causality, logic, etc.), Piaget proposed a series of stages by which the child approached a mature level of performance (Flavell, 1963). A great many developmental psychologists have followed in Piaget's footsteps.

3. ESSENTIAL ASPECTS OF STAGES

Flavell (1971) has analyzed the essentials of what Piaget seemed to mean by the concept of stages and reviewed the literature to determine whether these essentials are supported. He noted four principal implications of the stage concept. First, the change from one stage to another was to be *qualitative* rather than quantitative. It was not simply a matter of adding more information, but rather the emergence of a substantially different way of processing information. Second, stage changes were supposed to be *concurrent*. That is, there were simultaneous and similar changes across many domains at once. Third, the transition between stages was to be *abrupt* rather than gradual. The child is in stage n for prolonged period of time, followed by a relatively brief transition period, and then a prolonged period in stage n+1. Fourth, stages were highly *organized*, rather than containing independent, unrelated ideas. That is, cognitive items are interrelated so as to constitute an organized whole. Generally, this is taken to mean that there are two or more cognitive elements with one or more relations linking them.

Flavell (1971) omitted one additional, but widely held implication of stages. This is the notion of *ordinality*, which refers to the idea that children progress through stages in an order which is more or less invariant. That is, there is typically no skipping of stages and little regression to previous stages. A common corollary of ordinality has to do with the difficulty of learning much beyond the next stage. In general, one could expect no learning of stage n until the child was in stage n-1.

3.1 QUALITATIVE CHANGE

How do these five aspects of stages stack up against the empirical evidence from cognitive development? With regard to qualitative vs. quantitative changes, the bulk of the evidence seems positive (Flavell, 1971).

One of many examples of qualitative stage changes occurs in the child's learning about balance scales. Psychological assessments present the child with a rigid balance beam in which differing numbers of weights are placed on pegs at various distances to the left or right of a fulcrum. The child's task is to determine which side of the scale will go down when supporting blocks are removed. Siegler (1981) has indicated that children's performance on the balance scale progresses through four distinct rule based stages: (1) use weight alone to determine if the scale will balance, (2) emphasize weight, but consider distance (correctly) in the event that the weights to the left and right of the fulcrum are equal, (3) consider both weight and distance but get confused when one side has greater weight and the other has greater distance, (4) multiply distance by weight for each side and compare the products. Such stages emphasize, not simply the addition of distance information, but important differences in the way that weight and distance information is combined.

One qualification about qualitative change is viewing distance. Whether one regards a change as being primarily quantitative or qualitative depends on how closely one looks (Siegler, 1991). When viewed from afar, many changes look drastic and discontinuous; but when viewed from close-up, the same changes may look gradual and continuous. Because researchers have no direct window into the child's changing mental structures, their viewing distance is likely to be colored mainly by the grain size of their actual or assumed computational models. A symbolic level model, based say on rules or frames, will tend to yield more qualitative changes than, say, a connectionist model.

3.2 CONCURRENCE

The concurrence criterion demands simultaneous and similar changes across many domains. This has been a difficult question to address because of diagnostic problems, but the response from the literature has been resoundingly negative (Flavell, 1971).

For example, it has long been known that different types of conservation appear at different ages. Conservation refers to the ability to reason that two initially equal quantities do not change their equality relation over some physical transformation. An old finding is that conservation of number precedes conservation of solid-quantity, which in turn, precedes conservation of weight. More recently, 3- to 4-year-olds have been shown to conserve sets of small numbers, say below 5 items (Fuson, 1988). In contrast, other research has shown that most adults still hold erroneous conservation beliefs about area and perimeter quantities (Shultz, Dover, & Amsel, 1979). Piaget had referred to such variations by the term *d calages*, but he never explained them coherently.

3.3 ABRUPTNESS

Flavell's (1971) decision on whether abrupt transitions are common in stage progressions was negative. Instead, he concluded that change is more typically gradual and continuous. He noted that perfection of a cognitive item is not usually diagnostic of the end of a stage, but the appearance of new cognitive items is usually diagnostic of a stage transition. Thus, cognitive skills continue to improve beyond the stage in which they first appeared.

3.4 ORGANIZATION

The assessment of the organizational criterion has typically been positive (Flavell, 1971; Siegler, 1991). It has not made much sense to regard cognitive acquisitions as entirely independent of each other.

3.5 ORDINALITY

There is considerable agreement that, where stages of cognitive development occur, there is a strong tendency for children to proceed through them in a uniform order. However, there is typically some degree of skipping of stages and regression to earlier stages. An example of ordinality occurs in the 6-stage development of the concept of physical objects (Kramer, Hill, & Cohen, 1975).

Evaluation of the training corollary of ordinality is provided by a number of training studies that demonstrate that success depends critically on what the child currently knows. For example, it had been suspected that compensation (the realization that an increase in one physical dimension was compensated for by a decrease in another dimension) was a pre-requisite to learning about conservation. This was supported in an experiment in which trained compensators learned more about conservation than did either untrained compensators or trained non-compensators (Curcio, Kattef, Levine, & Robbins, 1972).

3.6 SUMMARY OF EMPIRICAL EVIDENCE

Thus, the developmental literature supports the stage notions of qualitative change, organization, and ordinality, but not those of concurrence or abruptness. It could be said that stages in cognitive development are characterized by qualitative and ordinal changes in organized knowledge structures.

4. INSIGHTS INTO STAGES FROM CONNECTIONIST MODELS

Early in our own connectionist simulations, the issue became one of how to get, or when to expect, multiple stages. Investigation of this issue led to what is so far our most fundamental new insight about stages -- the idea that stages result from solving part of a problem before solving all of the problem. Because there are several different ways that partial solutions can occur, this basic insight then unravelled into hypotheses about different sources of stages: hidden unit herding, over-generalization, training pattern bias, and hidden unit recruitment.

4.1 HIDDEN UNIT HERDING

The phenomenon of hidden unit herding was discovered by Fahlman and Lebiere (1990) in their study of the relative slowness of the back-propagation learning algorithm. During learning, each hidden unit is trying to become a feature detector to contribute to the overall solution. This process is complicated by the fact that all of the hidden units are changing simultaneously. Instead of each unit moving quickly and decisively to fill some useful role, hidden units move in a herd to deal with the current largest source of error.

A good example of hidden unit herding is that of inverted U-shaped development in learning the past tense of English verbs. Developmental psycholinguistic research has suggested that correct performance on irregular verbs is followed by incorrect over-regularization of irregular verbs (for example, saying *runned* for *ran*) and then by correct performance on the irregulars. Plunkett and Marchman (1990) have constructed back-propagation networks that simulate these psychological regularities. Plunkett and Marchman found that they could simulate the inverted U effect by manipulating the type and token frequencies of regular and irregular verb forms. To obtain over-regularization errors, the input to their networks had to conform to biases that also occur in spoken English. The regular verb forms had to be high in type frequency and low in token frequency. And the irregular verb forms had to be low in type frequency and high in token frequency. As well, there was a gradual expansion of the vocabulary to which the network was exposed.

In terms of hidden unit herding, one could imagine that learning the irregular forms is one subtask, and learning regular forms is another subtask. Most of the initial error was probably due to the high token frequency irregular forms. But as those forms were learned and vocabulary expanded, the largest remaining error came from the high type frequency regular forms. Coping with that error produced over-regularization errors on the irregular forms through retroactive interference. Eventually, the hidden unit responsibilities were sorted out so as to reduce most of the error. Although this interpretation has not been conclusively demonstrated by analyses of network structure, it is likely because back-propagation is so prone to hidden unit herding.

4.2 OVER-GENERALIZATION

Over-generalization is treated separately from hidden unit herding because it can occur without hidden units. Over-generalization can be based on differentially early solution of subtasks, for whatever the reason. If two or more subtasks share a common distributed representation, and some of these subtasks are more difficult than others, then it might well happen that the early developing solutions get over-generalized to the more difficult subtasks.

An example occurred in our simulation of the development of the concepts of potency and resistance in causal reasoning (Shultz, Zelazo, & Strigler, 1991). Potency refers to the strength of the causal system; resistance is the degree of difficulty in producing an effect. Estimates of these two concepts are combined to predict magnitude of the effect. There are two common combination rules for simple physical effects. One is based on subtraction (effect size = potency - resistance), the other on division (effect size = potency / resistance).

Psychologically, it has been found that a subtraction rule is learned first, and before a division rule is learned, children over-generalize the subtraction rule to division problems (Zelazo & Shultz, 1989). Cascade-Correlation networks simulated these and other psychological results (Shultz, Zelazo, & Strigler, 1991). Because these networks learned as perceptrons, without recruiting any hidden units, the results cannot be attributed to hidden unit herding. The networks learned subtraction first, and then temporarily over-generalized subtraction to still unlearned division problems. The stage of over-generalization was eventually replaced by a more mature stage in which subtraction was restricted to subtraction phenomena and division applied to division phenomena.

4.3 TRAINING PATTERN BIAS

There can be environmental biases in training patterns that influence what aspects of a problem are mastered first. These biases can be either sequential (Rumelhart & McClelland, 1986) or continuous (Plunkett & Marchman, 1990). Connectionist architectures are highly sensitive to these differential frequencies in training patterns.

Examples can be found in simulations of the balance scale phenomena discussed earlier. In order to capture rules 1 (use of weight alone) and 2 (use of weight and also distance when weights are equal), it is useful to bias the training patterns in favor of equal distance problems

(McClelland, 1988; Shultz & Schmidt, 1991). Without such bias, many networks skip stages 1 and 2 and move directly into stages 3 and 4.

The importance of training pattern bias can also be seen in the simulations of learning the past tense of verb forms discussed earlier (Plunkett & Marchman, 1990). The stages in that simulation could more accurately be attributed to a combination of training bias and hidden unit herding.

An example of sequential training pattern bias is provided by Elman's (1989) model of the acquisition of a phrase structure grammar. His recurrent, back-propagation network was able to learn this grammar only when the proportion of complex sentences (with relative clauses) was increased gradually over training. Early training on simple sentences seemed to create a solid foundation for later learning of complex sentences. Speech directed to the young child is often simplified, syntactically and in other ways (Snow & Ferguson, 1977). Perhaps this sort of sequentiality of complexity could generate syntactic stages.

4.4 HIDDEN UNIT RECRUITMENT

The recruitment of hidden units as needed during learning is a technique used by Cascade-Correlation (Fahlman & Lebiere, 1990) and a few other connectionist algorithms. These networks generally recruit hidden units that correlate highly with any existing error signals. Such recruitment then enables further error reduction as output-side weights continue to be adjusted. The installation of new hidden units in the network represents an underlying qualitative difference in network structure; it now possesses higher level feature detectors that it previously lacked. Consequently, the network is now poised to solve new aspects of the overall problem.

An example can be cited from our Cascade-Correlation model of balance scale phenomena (Shultz & Schmidt, 1991). We found that hidden unit recruitment often allowed a quick jump in the network's performance from the present stage to the next stage. In 16 runs, 24 hidden units were recruited and 13 of them were associated with rapid stage progressions: 5 to rule 4, 7 to rule 3, and 1 to rule 2.

Another example comes from Cascade-Correlation simulations of the learning of the personal pronouns *me* and *you* (Shultz, Buckingham, & Oshima-Takane, 1991). These pronouns are particularly difficult to learn because their referent shifts with conversational role and because correct models of their use are not provided in speech directly addressed to the child. It was found that Cascade-Correlation networks could learn these pronouns, in a way that simulated some psychological phenomena, but only by recruiting at least one, and usually two, hidden units.

5. STAGES REVISITED

My earlier review of the psychological status of the various implications of the stage concept had suggested that qualitative changes, organization, and ordinality were supported, but that concurrence and abruptness were not. How do these findings fit with current connectionist models of development?

Simulations of the learning of verb past tense, potency and resistance in causal prediction, balance scale phenomena, and personal pronouns all involved qualitative changes in network performance. Although rules descriptive of qualitatively different performance levels were not explicitly represented, the networks behaved as if their knowledge was qualitatively different than it had been in the previous stage. As in the psychological literature, some of these changes at relatively high levels of description could be revealed, on closer inspection, to be due to small underlying quantitative changes in weight adjustment. Other qualitative stage changes could be attributed to underlying qualitative changes in the network structure, caused by recruitment of hidden units.

The absence of concurrence across domains also falls naturally out of connectionist models, since they rely so heavily on particular experiences. Because of the way the environment or the network is structured, stages may be found in one domain and not in other domains, or may be found across domains at different times. In general, as in the psychological literature, concurrence would be the lucky exception, rather than the norm.

Lack of abruptness in the appearance of stages is also consistent with connectionist models, since most of their adjustments are rather small changes in connection strengths. Often these small adjustments need to accumulate to cause any large difference in the network's performance. And when transitions occur, they are often tentative, just as with children. Algorithms that recruit new hidden units, such as Cascade-Correlation, can occasionally generate some rather abrupt changes, but these transitions are not concurrent or broad based. Rather, they are restricted to a particular domain.

Organizational properties, claimed to be common to children's cognition, are also inherent in connectionist networks. Although there is not full connectivity in the feed-forward networks discussed here, there is ordinarily full uni-directional connectivity from one level of the network to the next level.

Ordinality in the appearance of stages has been found in connectionist models, particularly in those concerned with the balance scale. The typical pattern was to progress through all four stages in the correct order, with some minimal skipping and regression (Shultz & Schmidt, 1991). The absence of skipping means that there was no learning of stage n until the network was firmly in stage n-1.

6. CONCLUSION

I conclude that connectionist modelling of stages is so far quite consistent with major regularities in the psychological literature on cognitive development. These models produce those aspects of stages that are seen in children, avoid those aspects that are not seen in children, and suggest a number of novel theoretical insights into why stages arise and when one can expect to find them. Connectionist networks are worth exploring further as models of transition mechanisms in cognitive development.

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Thanks to Laura Petitto and William Schmidt for comments and discussion.

References

Chauvin, Y. (1989). Toward a connectionist model of symbolic emergence. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 580-587). Hillsdale, NJ: Lawrence Erlbaum.

Curcio, F., Kattef, E., Levine, D., & Robbins, O. (1972). Compensation and susceptibility to conservation training. *Developmental Psychology*, **7**, 259-265.

Elman, J. L. (1989). Representation and structure in connectionist models. Technical Report #8903, Center for Research in Language, University of California, San Diego, CA.

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

Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. Princeton, NJ: Van Nostrand.

Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, **2**, 421-453.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer Verlag.

Kramer, J. A., Hill, K. T., & Cohen, L. B. (1975). Infants' development of object permanence: A refined methodology and new evidence for Piaget's hypothesized ordinality. *Child Development*, **46**, 149-155.

McClelland, J. L. (1988). Parallel distributed processing: Implications for cognition and development. Technical Report AIP-47, Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA.

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

Plunkett, K., & Marchman, V. (1990). From rote learning to system building. Technical Report #9020, Center for Research in Language, University of California, San Diego, CA.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 2. Cambridge, MA: MIT Press.

Shultz, T. R., Buckingham, D., & Oshima-Takane, Y. (1991). A connectionist model of the learning of personal pronouns in English. Manuscript submitted for publication.

Shultz, T. R., Dover, A., & Amsel, E. (1979). The logical and empirical bases of conservation judgments. *Cognition*, **7**, 99-123.

Shultz, T. R., & Schmidt, W. C. (1991). A Cascade-Correlation model of balance scale phenomena. Manuscript submitted for publication.

Shultz, T. R., Zelazo, P. D., & Strigler, D. (April, 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, WA.

Siegler, R. S. (1981). Developmental sequences between and within concepts. *Monographs of the Society for Research in Child Development*, **46** (Whole No. 189).

Siegler, R. S. (1991). *Children's thinking*. Englewood Cliffs, NJ: Prentice-Hall.

Simon, T., Newell, A., & Klahr, D. (in press). Q-Soar: A computational account of children's learning about number conservation. In D. Fisher & M. Pazzani (Eds.), *Computational approaches to concept formation*. Los Altos, CA: Morgan Kauffman.

Snow, C. E., & Ferguson, C. A. (Eds.) (1977). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

Sternberg, R. J. (1984). *Mechanisms of cognitive development*. New York: Freeman.

Zelazo, P. R., & Shultz, T. R. (1989). Concepts of potency and resistance in causal prediction. *Child Development*, **60**, 1307-1315.