Modeling developmental transitions in adaptive resonance theory

Maartje E.J. Raijmakers and Peter C.M. Molenaar

Department of Developmental Psychology, University of Amsterdam, The Netherlands

Abstract

Neural networks are applied to a theoretical subject in developmental psychology: modeling developmental transitions. Two issues that are involved will be discussed: discontinuities and acquiring qualitatively new knowledge. We will argue that by the appearance of a bifurcation, a neural network can show discontinuities and may acquire qualitatively new knowledge. First, it is shown that biological principles of neurite outgrowth result in self-organization in a neural network, which is strongly dependent on a bifurcation in the activity dynamics. Second, the effect of a bifurcation due to morphological change is investigated in an Adaptive Resonance Theory (ART) network. Exact ART networks with quantitative differences in network structure at the category level show qualitatively different dynamical regimes, which are separated by bifurcations. These qualitative differences in dynamics affect the cognitive function of Exact ART: Representations of learned categories are local or distributed.

Introduction

Constructivist theories of cognitive development are involved in at least two theoretical issues: the occurrence of discontinuities and the possibility of acquiring qualitatively new knowledge. According to the developmental theory of Piaget (e.g. Piaget & Inhelder, 1969), children show stages in their development during which their cognitive abilities are relatively stable. But it is predicted that during a transition between two stages, behavior shows discontinuous change and qualitatively new knowledge is acquired. Both issues have raised a lot of criticism. Brainerd (1978) argued that a discontinuity in behavior is not a testable prediction because it cannot be distinguished from a continuous acceleration. Fodor (1980) is one of the critics of the second claim because, as he argues, an explanatory model of the mechanism of qualitative change is impossible.

A solution has been found for testing discontinuities in behavior. Several authors (Preece, 1980; Van der Maas & Molenaar, 1992) model a discontinuity in behavior by a bifurcation, in particular the so-called cusp catastrophe (Thom, 1975). A bifurcation is mathematically defined as a qualitative change of the equilibrium behavior emerging from gradual change in a control parameter (see next section). Van der Maas and Molenaar (1992) derived testable predictions (catastrophe flags) from a bifurcation model (the so-called cusp model).

Empirical evidence of developmental transitions in behavior has been found in several domains: for example, Jansen and Van der Maas (2001) show that the transition from Rule I to Rule II on the balance scale task is a bifurcation. Evidence for bifurcations has also been found in the domain of language development (e.g. Dromi, 1987), motor development (e.g. Wimmers, Savelsbergh, Beek & Hopkins, 1998) and analogical reasoning (Hosenfeld, van der Maas & van den Boom, 1997).

The impossibility of acquiring qualitatively new knowledge is still debated (e.g. Laurence & Margolis, 2002), often in reaction to Fodor (1980, 1981). According to Fodor, learning is necessarily some sort of inductive inference involving hypothesis formation and confirmation. Therefore new concepts that a subject can learn have to be composed of innate primitive concepts. The addition of conceptual resources (by maturation) is the only way in which Fodor allows a system to become more powerful. We will refer to this as Fodor's scenario.

Connectionist models of learning and cognitive development have been proposed as systems that can acquire new concepts (e.g. Elman, 1996). After learning input–output relations, the hidden node activity in connectionist models is considered to represent new concepts. An objection against this claim is that the learned representations thus obtained still have to be compositional in the sense of Fodor's scenario (e.g. Bloom & Wynn, 1994). Moreover, current neural networks may not learn

Address for correspondence: Maartje Raijmakers, Department of Developmental Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands; e-mail: mraijmakers@fmg.uva.nl

concepts at all, but only associations that fully depend on the context of the learned examples (Marcus, 1998; Raijmakers, van Koten & Molenaar, 1996).

Generative neural networks, such as cascade correlation networks, model constructive growth principles by the addition of hidden nodes, as a function of learning. Quartz (1993) argues that these networks extend their hypothesis space during learning and hence escape Fodor's scenario (see also Mareschal & Shultz, 1996). Cascade correlation networks are presented as models of transitions in development on several Piagetian task domains: the balance scale (Shultz, Mareschal & Schmidt, 1994), distance space and velocity (Buckingham & Shultz, 2000) and seriation (Mareschal & Shultz, 1999). The constructive growth principles in generative neural networks would not seem to escape Fodor's scenario, however, but in fact constitute a particular instantiation of it (cf. Raijmakers, 1997, and Marcus, 1998, for further elaboration).

In what follows we will argue that modeling a transition by a bifurcation leads to a model of acquiring qualitatively new knowledge without the addition of resources, thus escaping Fodor's nativistic scenario. The outline of the article is as follows: First, it is explained why a bifurcation may lead to qualitatively new knowledge. Second, a constructivist model of neurite outgrowth is presented in which the development of network structure bifurcates. Third, in an Adaptive Resonance Theory neural network, we show how this growth-related bifurcation may have an effect on cognition.

Modeling transitions by a bifurcation

Modeling a developmental transition by a bifurcation provides testable predictions (catastrophe flags) to detect discontinuities, as we discussed in the introduction. Several researchers (including Fodor, 1981, p. 300) suggested that the occurrence of a bifurcation (often mentioned as self-organization) might be a mechanism by which a nonlinear dynamical system, such as a human neural network, acquires qualitatively new knowledge, escaping Fodor's scenario (cf. Bereiter, 1985; Lawson & Staver, 1989; Molenaar, 1986; Molenaar & Raijmakers, 2000).

Dynamical systems are defined by equations that describe the evolution of their behavior (see Kelso, Ding & Schöner, 1992, for a general introduction). In the following example of a system that shows bifurcations, the behavioral variable is X, which depends on a parameter W, which is called the control parameter. For each value of W, X has one or more equilibrium values, as depicted in Figure 1. If a dynamical system is perturbed, it moves to a stable equilibrium state (solid line) and away from
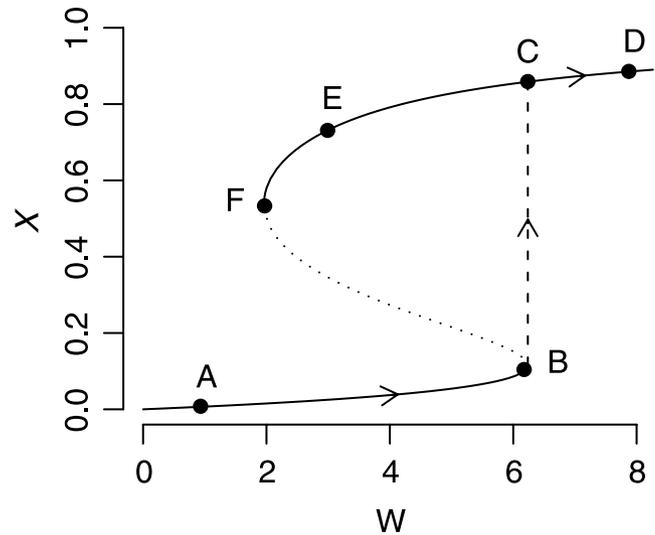
Figure 1 *Equilibrium behavior of a dynamical system with behavioral variable* X *as a function of parameter W. Solid lines are stable equilibrium states of* X*, dotted lines are unstable equilibrium states (see text for further explanation).*

an unstable equilibrium state (dotted line). Points F and B are bifurcation points, because at these points the number of equilibrium states changes under variation of W. More precisely, B and F are so-called fold bifurcation points, which are one particular kind of bifurcation (Guckenheimer & Holmes, 1993). In a fold bifurcation point, a stable and an unstable equilibrium appear or disappear simultaneously. If the system starts at A and W gradually increases from 1 to 8, then X moves from A to B to C to D. From B to C, the equilibrium value of X makes a discontinuous jump, which is one of the catastrophe flags.

According to Fodor's scenario, qualitatively new knowledge can only be learned by the addition of resources. However, if the behavior of the above system at point D differs qualitatively from that at A, then it becomes more powerful without adding resources. In dynamical systems, a qualitative change of the dynamics is triggered by a gradual change of a parameter, which is a resource that was present all the time. Moreover, the change of the parameter does not by itself determine the nature of the qualitative change, that is, the increase of power. In the following section we will discuss a model of biological growth of a neural network structure, which shows a bifurcation of activity dynamics.

Constructivism in biology

Biological models of neurite outgrowth show that constructive growth principles can be implemented by the
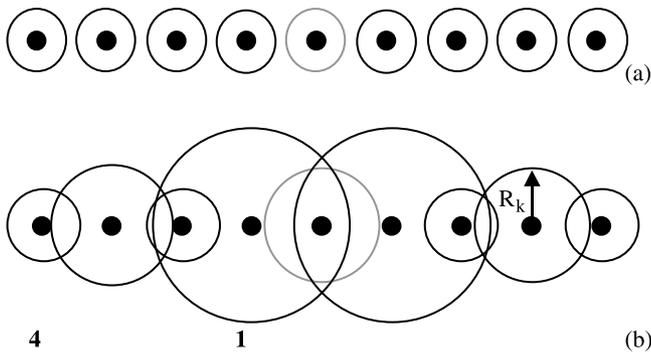
Figure 2 *A schematic view of a one-dimensional layer of eight excitatory (black lines) and one inhibitory (gray line) neurons. (a) The model starts with non overlapping neuritic fields, that is disconnected neurons. (b) In equilibrium, a pattern of small and large overlapping neuritic fields has resulted, which means that neurons are connected with different weights. Replication of results in Van Ooyen* et al. *(1995).*

Figure 3 *The time course of the neuritic field size of two neurons, 1 and 4, and of the average activity of excitatory neurons of the model displayed in Figure 2. Note the initial overshoot of the neuritic field sizes of the neurons and the abrupt change of the average activity. Replication of results in Van Ooyen* et al. *(1995).*

evolution of a fixed number of variables. Van Ooyen, Van Pelt and Corner's (1995) model is called a constructivist model because the development of network morphology is directly dependent on neuron activity and external influences upon it. The mathematical basis of the model consists of evolution equations for neuronal activity of individual neurons and the outgrowth of the neuritic fields of individual neurons, which are mutually dependent. The network is a so-called shunting neural network (SNN; Grossberg, 1988). In an SNN excitatory inputs drive the membrane potential towards a finite maximum, while inhibitory inputs drive the membrane potential towards a finite minimum. In the model discussed below, the neurons are placed on a circle such that connections grow first between neighbors and then with more distant neurons.

Figure 2 shows a simple example of Van Ooyen *et al.*'s model. The neuritic field of each neuron, which is shown by a circle, is the range of its interactions with other neurons. The neuritic field of neuron k is modeled by one real-valued parameter, its radius, R_k. The connection weight between two neurons k and l equals the overlap of their neuritic fields.

In the model, initially disconnected cells (Figure 2a) organize themselves into a network of connected cells (Figure 2b) under the influence of intrinsic activity. Several characteristic phenomena of brain development emerge from the dynamics of this relatively simple model (see Van Ooyen *et al.*, 1995), such as a transient overproduction (overshoot) in the number of connections. The bifurcation diagram of Figure 1 shows the dependence of the average activity, X, on the average connection weight, W. Due to their mutual dependence, in time, X and W
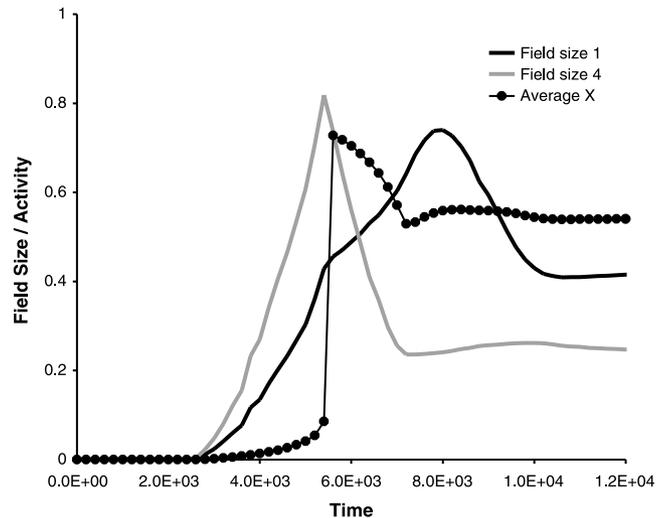
move, approximately, from A to B to C to E. The effect of the bifurcation is illustrated in Figure 3, which shows that the average activity of excitatory neurons abruptly increases with gradual change of connection weights.

The dynamics of the model are defined on a low level, that is, the interaction of individual cells. Morphological and functional properties of the model, such as recovery of network structure after cell death and a critical period for the elimination of neurons, appear as emergent properties of the system. The fact that the outcomes of developmental processes are similar under normal circumstances does not imply that the network structure and the functional properties have been completely laid down in the mathematical description of the network.

Bifurcations in Adaptive Resonance Theory

The biological model of morphological growth discussed here does not show functional, cognitive behavior. However, in some neural network models for cognition, such as Adaptive Resonance Theory (ART; Grossberg, 1976), the SNN implements a content addressable memory. In ART, the SNN maps inputs to category information. The category information is stored in long-term memory, which is recalled directly on the basis of the content of input. We will show how the occurrence of bifurcations in the SNN leads to qualitative changes of ART on a cognitive level. First, we will briefly introduce an ART model.

Exact ART

Adaptive Resonance Theory defines network principles for unsupervised learning of patterns. As in a statistical clustering algorithm, patterns that are alike according to some distance measure are clustered in one category. In ART, each category is represented by a vector, which describes the common factors of its elements. An ART network constructs categories online during the presentation of input vectors, and is able to classify known and unknown input vectors. Each sequence of input vectors will result in a stable clustering. Moreover, an ART network remains adaptable to unknown input vectors and a very long or frequent presentation of one pattern will not remove some of its learned categories. There exist many implementations of Adaptive Resonance Theory: ART 1 (Carpenter & Grossberg, 1987a) is a classifier of binary vectors; ART 2, ART 3 and fuzzy ART are classifiers of analog vectors (Carpenter & Grossberg, 1987b, 1991; Carpenter, Grossberg & Rosen, 1991). Unsupervised ART networks are extendable to ARTMAP networks for supervised learning (e.g. Carpenter, Milenova & Noeske, 1998). See Levine (2000) for an overview. Here we will describe Exact ART (Raijmakers & Molenaar, 1997) because it actually implements the content addressable memory as an SNN. Exact ART clusters analog input vectors.

The Exact ART model consists of several modules (Figure 4): an input level (F0), a feature level (F1), a category level (F2) and a reset system. Between F0 and F1 only bottom-up, non-adaptable connections exist. F1

Figure 4   *General lay-out of Exact ART. Black bars are activities at the different levels. The gray lines are connections between levels. Between F0 and F1 only bottom-up, non-adaptable connections exist. F1 and F2 are mutually, fully connected by bottom-up connections (dark gray) and top-down connections (light gray). Only outgoing connections of one F1-node and one F2-node are drawn.*

and F2 are mutually, fully connected by bottom-up, adaptable connections (dark gray in Figure 4) and top-down, adaptable connections (light gray in Figure 4).

ART is based on the idea of adaptive resonance feedback between two layers of nodes, F1 and F2, as developed in Grossberg (1976). The following sequence of events occurs after presentation of an input pattern. Each step in the description corresponds with a diagram in Figure 5.

1. In F0, the input vector is preprocessed, such that contrast is enhanced, activity is normalized and noise is suppressed.
2. The resulting activity vector is bottom-up input for the feature level, F1.
3. The activity vector in F1 is multiplied by bottom-up, adaptable weights and triggers an activity pattern at the category level, F2. In its standard form, F2 is an SNN that results in a winner-takes-all competition between F2-nodes. This means that the F2-node with the largest input is the only node with sustained activity. The one, active F2-node is the recognized category. The category is represented by its connected bottom-up weight vector (In Figure 5–3, the thickness of the line denotes the size of the weight). In a standard F2-layer, categories have a local representation, because they are represented by the weight vector of one F2-node.
4. The active F2-node also sends activity down, which is the expected activity pattern generated by the active category. This top-down input for F1 is the F2 activity vector multiplied by top-down weights. F1 now receives two input vectors: the bottom-up input from F0 and the top-down expectation from F2. The resulting F1 activity consists of the matching elements of both vectors. Resulting F1 activity is normalized such that the sum of F1-node activities is 1.
5. The F0 and the F1 vector are compared by calculation of the match, which is a scalar between 1 and 0.[1] A match larger than the predefined vigilance parameter ρ results in resonant feedback between F1 and F2 (continue with step 8). A match smaller than ρ is insufficient and results in a reset of the selected category (continue with step 6).
6. The reset system generates activity that inhibits all nodes of F1 and F2 shortly, such that activities become 0. Only recently active F2-nodes are suppressed for a longer period.

---

[1] The match, R, is calculated as follows: $R = 1 - \frac{1}{2} \sum_{i=1}^{N} |u_i - x_i| \cdot u_i$ is an element i of the activity vector in F0 and x_i is an element i of the activity vector in F1. Both vectors have length N.
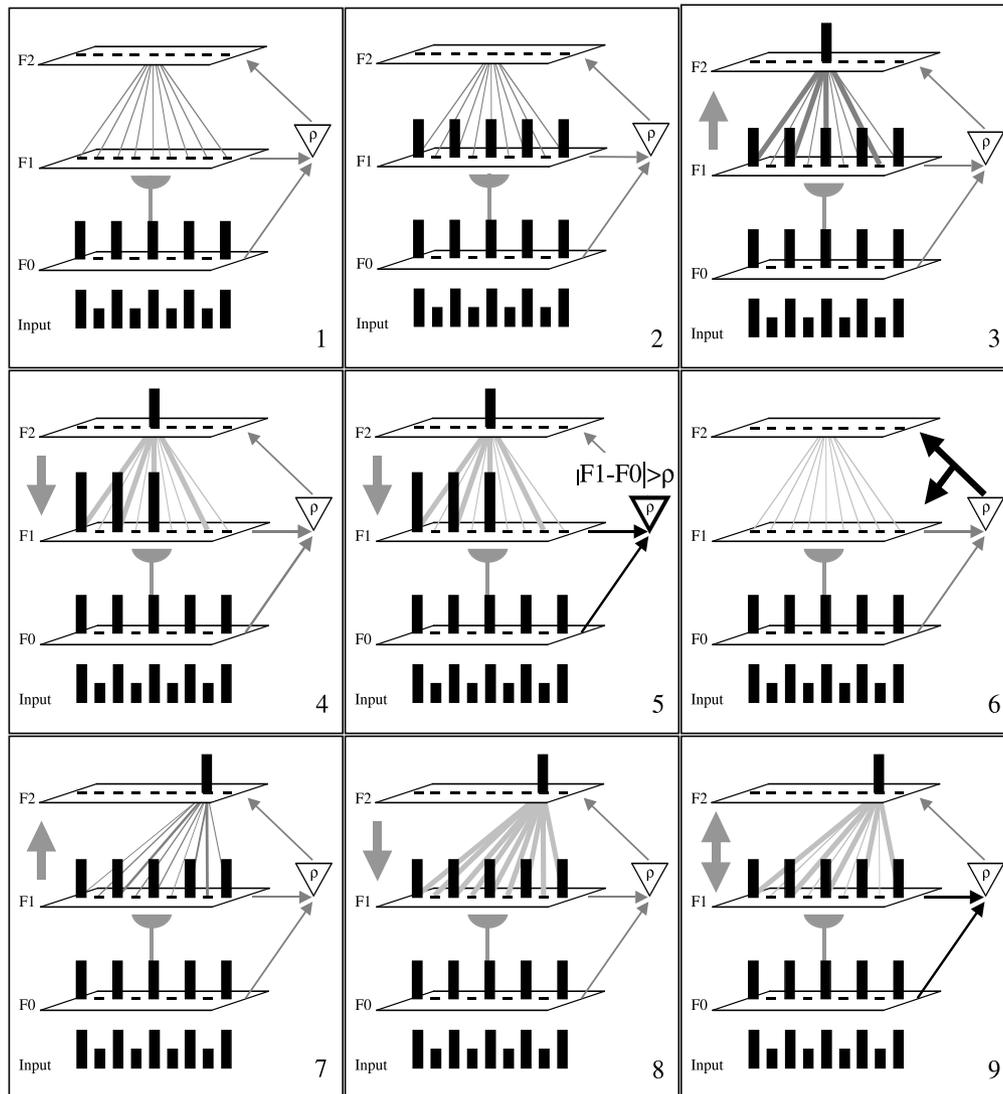
Figure 5   *Each diagram shows a phase of the classification and learning process of Exact ART. Separate steps are described in the text.*

7. As in steps 2 to 5, F1 becomes activated by F0 and competition takes place in F2. F2-nodes that caused a mismatch before, are now excluded from the competition. The search for an appropriate F2-node continues until a matching F2-node is found or a blank F2-node is selected, which always results in resonant feedback. The search is a relatively fast process, so that the weights between F1 and F2 get no time to adapt.

8. As a result of resonant feedback between F1 and F2, activity patterns in F1 and F2 are sustained for a relatively long period (compared to the search process), meaning that the input vector is stored in short-term memory (STM).

9. Resonant feedback gives weights of connections between F1 and F2 time to adapt, such that F1 activity patterns are stored in long-term memory (LTM).

Equations of bottom-up and top-down weight changes are given by:

$$
\begin{aligned}
\text{Top-down} &: \Delta z_{ji} = \alpha f(y_j)(f(x_i) - z_{ji}) \\
\text{Bottom-up} &: \Delta z_{ij} = \alpha f(y_j)(f(x_i) - z_{ij})
\end{aligned}
\tag{1}
$$

Δz_{ji} denotes the change in weight from F2-unit j to F1-unit i; Δz_{ij} denotes the change in weight from F1-unit i to F2-unit j; y_j denotes the activity of F2-node j; x_i denotes the activity of F1-node i; $f()$ is a threshold

function; α denotes the learning rate. Bottom-up weights are initialized at small random values, top-down weights are initialized at maximum values (i.e. 1.0). According to Equation 1, weights are only updated if they are connected to active F2-units. Moreover, weights are normalized because F1 activity is normalized and during the update weight z_{ji} and z_{ij} become equal to x_i (see also Equation 3 from Munakata and Pfaffly, this issue). Although from the above description, the classification and learning processes appear to be sequential and stepwise, the actual implementation is continuous and parallel in Exact ART (but not in most other ART implementations). Exact ART is fully defined by the evolution equations of all variables (such as Equation 1) and the initial values of all variables. For example, the sequential search process is realized as transient behavior in the Exact ART model (i.e. behavior of the network before the network is in equilibrium). So-called fast learning takes place if after presentation of each input vector, the network equations are calculated until all variables reach equilibrium, such that Equation 1 also reaches equilibrium. Further details are described in Raijmakers and Molenaar (1997).

From local to distributed representations

In its standard form, the category level of ART models is implemented by winner-takes-all dynamics in an SNN, which means that only one F2-node is active at a time. As a consequence, the representations of categories are local: each category is represented by one weight vector and each weight is part of the representation of only one category. However, the SNN appears to have several dynamical regimes under variation of the range of connections, as we saw in the biological model of neurite outgrowth. Raijmakers, van der Maas and Molenaar (1997) performed a numerical bifurcation analysis to systematically map all relevant bifurcations and dynamical regimes of the SNN as it is implemented in Exact ART. They varied structural parameters of the SNN: the range of excitatory connections (σ_e) and the range of inhibitory connections (σ_i). The strength of inhibitory and excitatory connections between units within F2 is defined by a Gaussian function with parameter σ and d: σ denotes the width of the function (σ denotes the standard deviation in a Normal distribution function); d denotes the size of the surface under the graph (d equals 1 in a Normal distribution function); d_e and d_i denote the sum of connection weights of, respectively, excitatory and inhibitory connections per neuron within F2 (see Figure 6). Parameters σ_e, σ_i, d_e and d_i can be interpreted as properties of cortical networks that change during postnatal development (Purves, 1994).
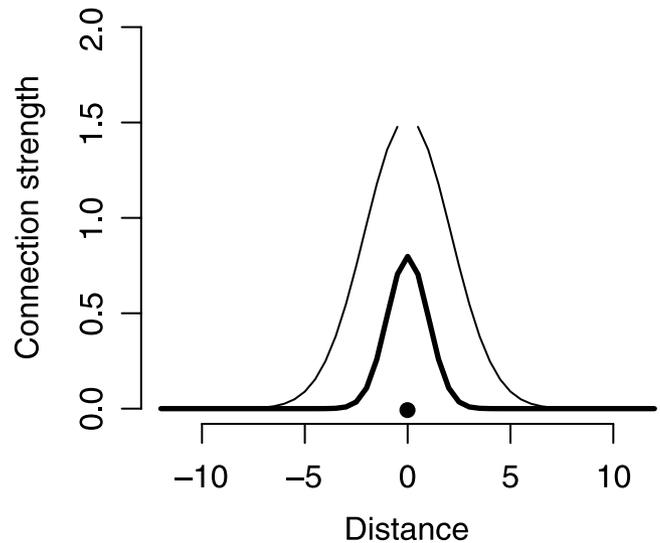
Figure 6 *Graphs of the strength of excitatory (bold line) and inhibitory (thin line) connections between units within F2. The horizontal axis represents the distance between two neurons. The vertical axis represents the connection strength of two neurons. Note that there are no self-inhibitory connections* (σ_e, σ_i, d_e, d_i) = (1, 2.1, 1, 6.48).

Numerical bifurcation analysis shows that fold bifurcations (see Figure 1; Guckenheimer & Holmes, 1993) in activity dynamics appear with σ_i and σ_e as control parameters. Figure 7 shows different activity patterns of a 25-node SNN with variation of σ_i.

The most striking difference between the activity patterns in Figure 7 is the number of simultaneously active F2-nodes. The winner-takes-all dynamics corresponds with one active node, as in Figure 7j, and results in local representations. If more nodes are active simultaneously distributed representations may also appear, as we show by the following small simulation study. During this study Exact ART learns and classifies a small number of input patterns, during which connections within F2 are fixed.

Exact ART is repeatedly presented the following sequence of input patterns A, B, C, A, D, which are shown in Figure 8. The study is constructed after Carpenter and Grossberg (1987b) to test for some important characteristics of the model with regard to stability and plasticity (for the rationale of this study see Carpenter & Grossberg, 1987b, and Raijmakers & Molenaar, 1997). The match criterion, ρ, is set to .7 such that Exact ART with a winner-takes-all dynamics would learn four distinct categories to represent the input patterns.

In this study σ_i is relatively small, such that more nodes are active simultaneously ($\sigma_i = 1.3$). Figure 9 shows equilibrium patterns of F2. With six or seven simultaneously active F2-nodes, the total number of nodes in F2 is too small to form four totally distinct representations.

$\sigma_i = .7$ $\sigma_i = 1.2$ $\sigma_i = 1.6$ $\sigma_i = 2.0$

a   **b**   **c**   **d**

$\sigma_i = 2.7$ $\sigma_i = 3.0$ $\sigma_i = 3.15$ $\sigma_i = 3.35$

e   **f**   **g**   **h**

$\sigma_i = 5.0$ $\sigma_i = 8.15$ $\sigma_i = 10.0$

i   **j**   **k**
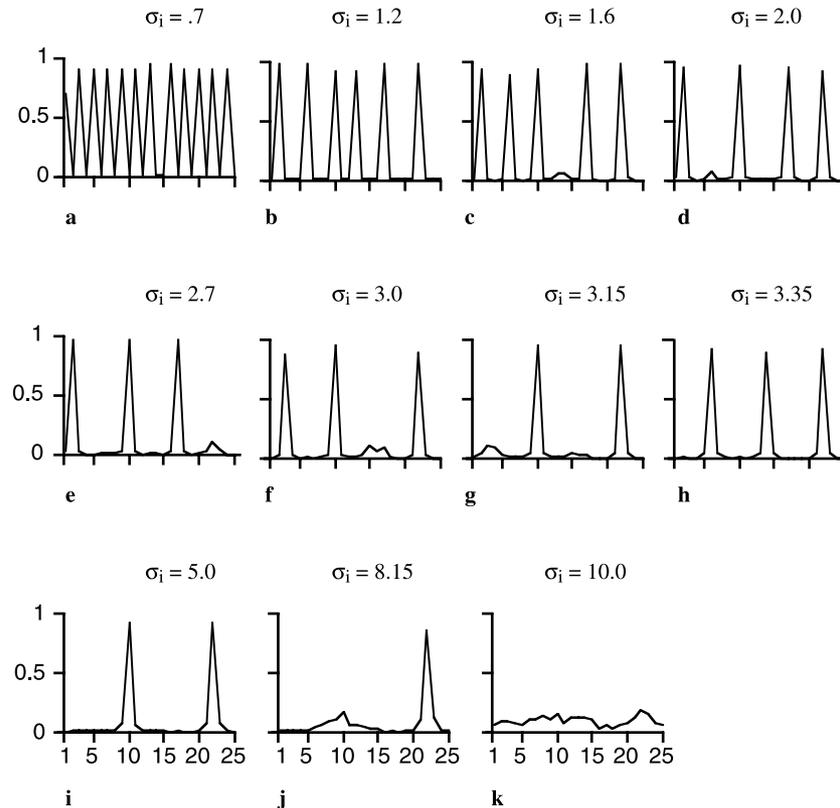
Figure 7 *Activity patterns for different values of σ_i, ranging from .7 in Figure 5a to 10.0 in Figure 5k ($d_e = 1$, $d_i = 24$, $\sigma_e = 1.02$). In general, the number of active nodes decrease with increasing σ_i: i.e. 12, 6, 5, 4, 3, 3, 2, 3, 2, 1 and 0, Figure (h) being an exception. More details can be found in Raijmakers et al. (1997).*

A   **B**   **C**   **D**
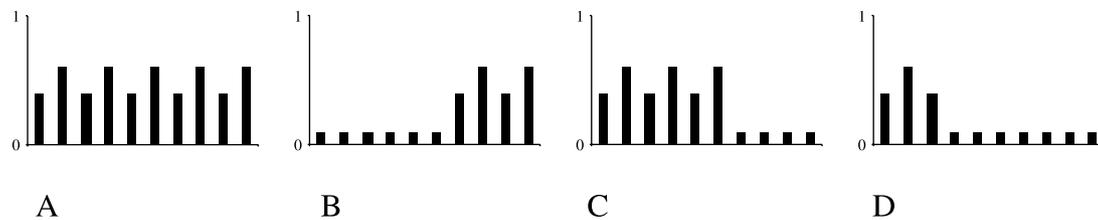
Figure 8 *Input patterns A, B, C and D of a small simulation study (see also Carpenter & Grossberg, 1987b). Horizontal axis denotes node index, vertical axis denotes the size of activity.*

Consequently, some F2-nodes are active during the presentation of several input patterns, such that distributed representations are learned. The representations are stable and the match between input, i.e. F0, and learned representation, i.e. F1, is sufficient for each input pattern, although it is no longer perfect; the match is .99, .98, .73 and .81 for inputs A, B, C and D, respectively (calculation of the match is explained in footnote 1). As expected, learned category representations overlap: Representations of input patterns C and D share five out of seven F2-nodes. As can be verified in Figure 8, of all pairs, input patterns C and D are most alike. It appears that Exact ART learns distributed representations where it is forced to do so. These distributed representations are unique, but, in contrast to local representations, they represent similarities between patterns.[2]

[2] ART3 and Distributed ARTMAP can also form a distributed code (Carpenter & Grossberg, 1991; Carpenter *et al.*, 1998). As is pointed out by Carpenter (2001), if distributed codes are learned, there is a trade-off between adaptability and stability of learned categories, which does not occur in local coding.
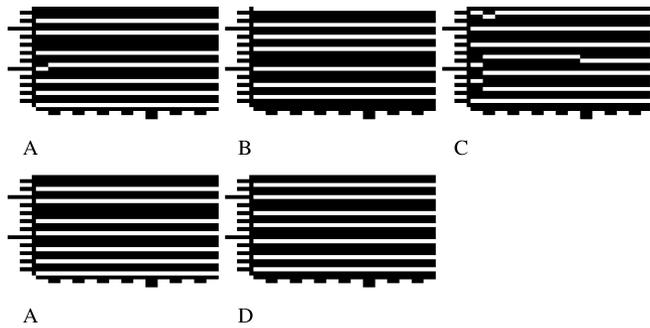
Figure 9   *Activity of F2-nodes, y_j, during presentation of input patterns at successive trials. The sequence of presentation of input is A, B, C, A, D, which sequence is repeated for 15 trials. Each figure shows, separately for each input, the resulting values of y_j, j = 1 to 25 at the end of each trial. The horizontal axis of each figure denotes the trial (1 to 15), the vertical axis denotes node index j (1 to 25). The hue represents the activity of F2-node j (black means $y_j = 0$, white means $y_j = 1$).*
$(\sigma_e, \sigma_i, d_e, d_i) = (.4, 1.3, 1, 24)$.

Conclusion

In the present article, we argue that modeling a developmental transition by a bifurcation may solve two main criticisms of constructivist theories of cognitive development. First, as was shown elsewhere (Van der Maas & Molenaar, 1992), a bifurcation model provides testable criteria to detect discontinuities. Second, simulation models of developmental mechanisms that show bifurcations, such as some neural network models, may learn qualitatively new knowledge, that is, they may escape Fodor's scenario. To our knowledge, present neural network models of developmental transitions do not include such a bifurcation. Furthermore, it appears to be extremely difficult to define a neural network that learns a nontrivial, developmentally relevant task by a bifurcation, because bifurcations involve instabilities. Moreover, transitions in cognitive development such as the balance scale task involve higher cognitive functioning, such as the application of higher-order rules. Most neural networks, such as feed-forward PDP networks, are not naturally suited for modeling higher cognitive tasks (Raijmakers *et al.*, 1996). As an exception, ART networks have some promising applications of higher-order rules, among other things because they include the sequential search process that is explained in the section Exact ART (Anumolu, Bray & Reilly, 1997; Levine, 1995; Tan, 1995).

The numerical bifurcation analysis of the SNN and the presented simulation study of Exact ART together show that bifurcations in the dynamics of the category level F2 result in the learning of qualitatively different representations in Exact ART, namely local representations and distributed representations. This is one import-

ant condition for learning qualitatively new knowledge. To show that Exact ART can truly escape from Fodor's scenario, two additional conditions should be fulfilled. First, the distributed representations should be more powerful than the local representations. This is not immediately clear, but the memory capacity and generalization ability of the network are certainly affected by the nature of the representations (Földiák & Young, 1995). Second, maturation (the change of parameters σ_i, σ_e, d_i and d_e) should be modeled as a process of change in addition to the learning process. In our view (cf. Quinlan, 1998), the extension of a neural network with a biologically plausible model of morphogenesis is an interesting innovation for neural network models of cognitive development. One way of doing this is to incorporate Van Ooyen *et al.*'s (1995) evolution equations of connection growth in Exact ART. With bifurcations at the category level of Exact ART, we showed a possible effect of growth dynamics on the cognitive level, namely the occurrence of qualitatively different dynamical regimes, which lead to qualitatively different, local and distributed, representations.

Acknowledgement

References

Anumolu, V., Bray, N.W., & Reilly, K.D. (1997). Neural network models of strategy development in children. *Neural Networks*, **10**, 7–24.

Bereiter, C. (1985). Toward a solution of the learning paradox. *Review of Educational Research*, **55**, 201–226.

Bloom, P., & Wynn, K. (1994). The real problem with constructivism. *Behavioral and Brain Sciences*, **17**, 707–708.

Brainerd, C.J. (1978). The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences*, **1**, 173–213.

Buckingham, D., & Shultz, T.R. (2000). The developmental course of distance, time, and velocity concepts: a generative connectionist model. *Journal of Cognition and Development*, **1**, 305–345.

Carpenter, G.A. (2001). Neural-network models of learning and memory: leading questions and emerging framework. *Trends in Cognitive Sciences*, **5**, 114–118.

Carpenter, G.A., & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing*, **37**, 54–115.

Carpenter, G.A., & Grossberg, S. (1987b). ART 2: self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.

Carpenter, G.A., & Grossberg, S. (1991). ART 3: hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. In G.A. Carpenter & S. Grossberg (Eds.), *Pattern recognition by self-organizing neural networks* (pp. 453–499). Cambridge, MA: MIT Press.

Carpenter, G.A., Grossberg, S., & Rosen, D.B. (1991). Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, 759–771.

Carpenter, G.A., Milenova, B.L., & Noeske, B.W. (1998). Distributed ARTMAP: a neural network for fast distributed supervised learning. *Neural Networks*, **11**, 793–813.

Dromi, E. (1987). *Early lexical development*. Cambridge: Cambridge University Press.

Elman, J.L. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.

Fodor, J.A. (1980). On the impossibility of acquiring more powerful structures. In M. Piattelli-Palmarini, J. Piaget & N. Chomsky (Eds.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 142–162). Cambridge, MA: Harvard University Press.

Fodor, J.A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.

Földiák, P., & Young, M.P. (1995). Sparse coding in the primate cortex. In M.A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 895–898). Cambridge, MA: MIT Press.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: feedback, expectation, olfaction and illusions. *Biological Cybernetics*, **23**, 187–202.

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, **1**, 17–61.

Guckenheimer, J., & Holmes, P. (1993). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields* (4th edn.). New York: Springer.

Hosenfeld, B., van der Maas, H.L.J., & van den Boom, D.C. (1997). Indicators of discontinuous change in the development of analogical reasoning. *Journal of Experimental Child Psychology*, **64**, 367–395.

Jansen, B.R.J., & Van der Maas, H.L.J. (2001). Evidence for the phase transition from rule I to rule II on the balance scale task. *Developmental Review*, **21**, 450–494.

Kelso, J.A.S., Ding, M., & Schöner, G. (1992). Dynamic pattern formation: a primer. In A.B. Baskin & J.E. Mittenthal (Eds.), *Principles of organization in organisms*. SFI Studies in the Sciences of Complexity, in Proceedings Vol. XII. Addison-Wesley.

Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition*, **86**, 22–55.

Lawson, A.E., & Staver, J.R. (1989). Toward a solution of the learning paradox: emergent properties and neurological principles of constructivism. *Instructional Science*, **18**, 169–177.

Levine, D.S. (1995). Learning and encoding higher-order rules in neural networks. *Behavior Research Methods Instruments & Computers*, **27**, 178–182.

Levine, D.S. (2000). *Introduction to neural and cognitive modeling* (2nd edn.). Mahwah, NJ: Erlbaum.

Marcus, G.F. (1998). Can connectionism save constructivism? *Cognition*, **66**, 153–182.

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

Mareschal, D., & Shultz, T.R. (1999). Development of children's seriation: a connectionist approach. *Connection Science: Journal of Neural Computing, Artificial Intelligence and Cognitive Research*, **11**, 149–186.

Molenaar, P.C.M. (1986). On the impossibility of acquiring more powerful structures: a neglected alternative. *Human Development*, **29**, 245–251.

Molenaar, P.C.M., & Raijmakers, M.E.J. (2000). A causal interpretation of Piaget's theory of cognitive development: reflections on the relationship between epigenesis and nonlinear dynamics. *New Ideas in Psychology*, **18**, 41–55.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child* (8th edn.). New York: Basic Books.

Preece, P.F.W. (1980). A geometric model of Piagetian conservation. *Psychological Reports*, **46**, 143–148.

Purves, D. (1994). *Neural activity and the growth of the brain*. Cambridge: Cambridge University Press.

Quartz, S.R. (1993). Neural networks, nativism, and the plausibility of constructivism. *Cognition*, **48**, 223–242.

Quinlan, P.T. (1998). Structural change and development in real and artificial neural networks. *Neural Networks*, **11**, 577–599.

Raijmakers, M.E.J. (1997). Is the learning paradox resolved? *Behavioral and Brain Sciences*, **20**, 573–574.

Raijmakers, M.E.J., & Molenaar, P.C.M. (1997). Exact ART: a complete implementation of an ART network. *Neural Networks*, **10**, 649–669.

Raijmakers, M.E.J., van der Maas, H.L.J., & Molenaar, P.C.M. (1997). Numerical bifurcation analysis of distance-dependent on-center off-surround shunting neural networks. *Biological Cybernetics*, **75**, 495–507.

Raijmakers, M.E.J., van Koten, S., & Molenaar, P.C.M. (1996). On the validity of simulating stagewise development by means of PDP networks: application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science*, **20**, 101–136.

Shultz, T.R., Mareschal, D., & Schmidt, C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning*, **16**, 57–86.

Tan, A.H. (1995). Supervised Adaptive Resonance Theory and rules. In L.C. Jain, B. Lazzerini & U. Halici (Eds.), *Innovations in ART neural networks* (pp. 55–86). Heidelberg: Physica-verlag.

Thom, R. (1975). *Structural stability and morphogenesis*. Reading, MA: Benjamin.

Van der Maas, H.L., & Molenaar, P.C. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review*, **99**, 395–417.

Van Ooyen, A., Van Pelt, J., & Corner, M. A. (1995). Implications of activity dependent neurite outgrowth for neuronal morphology and network development. *Journal of Theoretical Biology*, **172**, 63–82.

Wimmers, R.H., Savelsbergh, G.J.P., Beek, P.J., & Hopkins, B. (1998). Evidence for a phase transition in the early development of prehension. *Developmental Psychobiology*, **32**, 235–248.