Mental Tracking: A Computational Model of Spatial Development

Kazuo Hiraki
Presto, JST / ETL, MITI
1-1-4 Umezono Tsukuba-shi
Ibaraki, 305 Japan
khiraki@etl.go.jp

Akio Sashima and **Steven Phillips**
ETL, MITI
1-1-4 Umezono Tsukuba-shi
Ibaraki, 305 Japan
{sashima,stevep}@etl.go.jp

Abstract

Psychological experiments on children's development of spatial knowledge suggest experience at self-locomotion with visual tracking as important factors. Yet, the mechanism underlying development is unknown. We propose a robot that learns to *mentally track* a target object (i.e., maintaining a representation of an object's position when outside the field-of-view) as a model for spatial development. Mental tracking is considered as prediction of an object's position given the previous environmental state and motor commands, and the current environment state resulting from movement. Following Jordan and Rumelhart's (1992) forward modeling architecture the system consists of two components: an inverse model of sensory input to desired motor commands; and a forward model of motor commands to desired sensory input (goals). The robot was tested on the "three cups" paradigm (where children are required to select the cup containing the hidden object under various movement conditions). Consistent with child development, without the capacity for self-locomotion the robot's errors are self-center based. When given the ability of self-locomotion the robot responds allocentrically.

1 Introduction

This research challenges the traditional approach of theory construction in cognitive development by using the framework of robot learning. Traditionally, researchers in cognitive development (e.g., developmental psychologist) have focused on general and abstract descriptions of experimental data as explanations for their observations. However, developmental psychology is intrinsically limited with respect to the question *"how does development occur?"*, because of difficulties in the methodology (e.g., scientists should not open an infant's head to check for internal representations, and should not control their everyday experiences). Instead of real infants, we need a substitute that can be used for testing the theory and controlling conditions without ethical limitation. Consequently, the requirement of a computer simulation can no longer be ignored.

Over the past few decades several studies have been conducted on computational models of cognitive development. For example, Klahr and Wallance developed a computer model of acquisition of number conservation[1] using *self-modifying production system* [Klahr and Wallance, 1976]. Drescher proposed a *schema mechanism* to elaborate and test Piaget's theory from a constructivist's perspective [Dresher, 1991]. However, what is lacking in these approaches is an account of the interaction between children and environment. Consequently, models based on these approaches sometimes lack realism. We should pay more attention to the dynamics of cognitive development in the real world.

In contrast to these approaches, we propose using autonomous robots as the subject of cognitive development, and constructing computer programs by which robots can develop or learn analogously to infants. The advantage of using robots is twofold. First, we can utilize a robot's vision sensors and actuators as the inputs and outputs of the model. This forces us to use the same input stimuli and action goals as those of the infant, whereas the input and output representations of a computer simulation must be assumed. Second, we can construct a theory absorbing *activeness* in cognitive development. Recently, researchers have emphasized the importance of activeness (i.e., mobility) of infants during development [Thelen and Smith, 1994]. However, the theory derived from this stream needs

[1] Conservation is a term introduced by Piaget for the child's understanding that quantitative aspects of a set of materials are not changed or affected by transformations of the display itself.

to be tested and refined in more detail. We believe that using a robot leads us to a more concrete theory. More recently, Elman et. al. published an exciting book on development from a connectionist perspective [Elman *et al.*, 1996]. We follow their approach, but concentrate much more on interaction between individuals and environment.

As a first step to constructing a complete computational theory of cognitive development, we address the question of how infants relate to their spatial environments, and how this changes as the infant matures. To explore these issues, we focus on the change of *mental tracking*: the ability to update spatial relations between self and object without real (visual) tracking during the locomotion. We modeled the development of mental tracking as a learning task for a simulated robot, and conducted experiments simulating an infant's experience of locomotion.

The following sections describe our first results of modeling infant's spatial development. In Section 2, psychological evidence for spatial development is introduced. In Section 3, we elaborate our model for the development of mental tracking. Section 4 describes an empirical experiment with a simulated robot. In Section 5, we discuss the implications of our approach and future work.

2 Psychological Evidence for Spatial Development

2.1 Egocentrism in early infants

Piaget suggested that before infants are 1 year old, they exhibit a kind of sensorimotor *egocentrism* [Piaget, 1971]. Although the term egocentrism refers to young children's general tendency to view the world solely from their own perspective[2], we focus on infant's egocentric behavior in the *spatial environments* and how egocentric behavior changes into the *allocentric behavior* that normal adults exhibit. In other words, we address the question of how infants relate to their spatial environment, and how this changes as the infant matures.

Figure 1 shows an experiment designed to investigate infant's spatial searching [Bower, 1979]. A doll (prize) was put inside one of three cups, in this case the middle one, and then the infant moved around the table. The doll's relative position from the infant's view was changed from 'middle' to 'right'. Thus, the infant should look for the doll under the right cup. However, early infants frequently fail to

[2]Piaget used the word egocentrism referring not only to spatial behavior but also to more general aspects of young children such as *egocentic communication*.

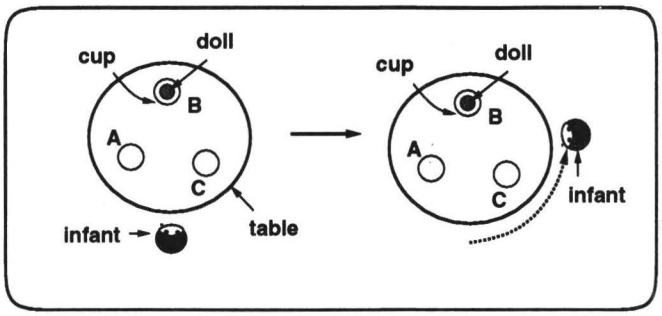

Figure 1: Self centered representation of space.

compensate for changes in their own spatial position. They continue to turn in the direction that previously led them toward the target (middle cup). The egocentric behavior in the searching task can be observed up to the age of about 18-months, but older children consistently show non-egocentric behavior on spatial tasks that involve searching for objects from different view points.

In the searching task the containers (cups) used to hide the prize were the same size and color. However, the actual environment provides much more information than position. In fact, infants can search the prize more correctly when distinctive cues (landmarks) are provided [Acredelo, 1979].

Phenomena concerning egocentrism are quite controversial and there are still many on-going studies. However, it's important to note that egocentic encoding of the target's position is not so irrational for infants who cannot move around. Pre-walking infants don't have to take into account changes in their own position in order to search for the target.

2.2 Effect of Locomotor Experience

So far, we have seen an interesting developmental change in infants spatial behavior. 12-year-olds behave egocentrically and 18-year-olds do not. What is the difference between 12-year-olds and 18-year-olds? How does egocentric behavior change into allocentric behavior? Experience in moving around environments seems to be one of the factors that effects the difference in the searching task. Kermoian and Campos suggested the importance of locomotor experience [Kermoian and Campos, 1988]. Infants who have experience in a walker or who can crawl are superior in spatial tasks to infants who have no such experiences.

Acredelo, Asams and Goodwyn conducted experiments to test the role of self-locomotion as opposed to passive transport concerning infant's spatial cognition [Acredelo *et al.*, 1984]. Their results sug-

gested the importance of active movement with visual tracking. When 12-month-olds walked to the other side of a layout and have the opportunity to continually looked in the direction of a hidden prize, they looked in that direction more often and subsequently did better at turning toward the object from the new location than children who were carried. In contrast, when they could not see the prize as they walk from one position to the other, they were subsequently no better in turning toward it than children who were carried. Based on these results they hypothesized self-produced motion leads to more effective deployment of visual attention.

3 Learning to Mentally Track

The psychological experiments mentioned previously suggest two important factors for the spatial development:

- self-locomotor experience; and
- visual tracking.

Yet, it is still unclear what information is central in promoting the change from egocentric behavior to allocentric behavior. In Acredelo's experiments, 18-month-olds can behave correctly without visual tracking of the target object. This leads us to the necessity of modeling the mechanism of the change, taking into account of the effect of locomotion experience.

In this section, we focus on the change of the ability of updating the spatial relation between self and target object without visual tracking during locomotion. We call this ability *mental tracking*, and propose a learning architecture for mental tracking by which robots can learn it analogously to infants. Firstly, we present our assumptions and identify the information that is available during the experience of self-locomotion.

3.1 Formalization as a Robot Learning Task

The Robot

Figure 2 shows a robot that was used for modeling infant spatial development. The robot is based on Nomad 200 (Nomadic Technologies,Inc.). It can control two wheels and trunk orientation. The robot is equipped with a movable stereo-camera (Sony EVID30 x 2) that is connected with a vision processing unit that uses a Fujitsu tracking module and DSP board(TMS320C40 x 2) for accelerating image processing. Using these facilities, the robot can detect relative distance and orientation to the target.

Figure 2: A robot for modeling cognitive development.

Locomotion Experience

Suppose that the above robot is the infant who has just started toddling. What types of information can the robot receive from walking? We assume that the self-locomotion experience of robots can be characterized by applying a next-state function f and an output function g successively. At time step $n-1$ the robot produces motor command $u[n-1]$. In conjunction with the state of the environment $x[n-1]$, the motor command determines the next state:

$$x[n] = f(x[n-1], u[n-1]). \quad (1)$$

Corresponding to each state $x[n]$ there is also a *sensation* $y[n]$:

$$y([n]) = g(x[n]). \quad (2)$$

We assume that the robot has access to the state of the environment: $y([n])$ can be seen as visual information directly obtained from the camera. The formalism is analogous to a standard state-action loop of mobile robot.

The Learning Task for Visual Tracking

Now we model the experience of locomotion with visual tracking. Locomotion experience with visual tracking can be modeled with a robot that moves while tracking the target with its movable camera and trunk. In other words, the task of visual tracking can be seen as generation of motor commands to the camera and the trunk to keep the target object on the center of the visual image. Note that we should consider two types of motor commands; one for moving and the other for visual tracking.

Let $y^*([n])$ be a desired sensation, and in this case the target object in the center of the visual image.

Let $u_m[n]$ be a motor command for moving[3], and $u_v[n]$ be a command for visual tracking. Given the state $x[n-1]$ (representing the current posture of the robot), $y^*([n])$ and $u_m[n]$, the robot produces an action $u_v[n-1]$:

$$u_v[n-1] = h(x[n-1], u_m[n-1], y^*([n])). \quad (3)$$

The learning task for visual tracking with locomotion is to make appropriate adjustments to the input-to-action mapping h based on data obtained from interaction with the environment. Note that we assume $u_m[n-1]$ is also given. This is because the robot should know how to move to the next position. The robot produces $u_m[n-1]$ independently of visual tracking.

3.2 The Learning Architecture for Mental Tracking

So far, we have defined the learning task for visual tracking with locomotion: Nonetheless, what we need is a model of mental tracking. It must be noted here that *mental tracking can be accomplished by mentally simulating visual tracking.* In other words, if the robot can learn to track the target while in motion, the robot can mentally track the target by applying input-to-action mapping h successively as an internal process. First, we present a learning architecture for visual tracking with locomotion, and then describe how to use the acquired knowledge for mental tracking.

Learning Inverse Model

As mentioned above, the learning task for visual tracking is to determine a proper command $u_v[n-1]$ given $x[n-1], u_m[n-1], y^*([n])$. This is analogous to the so called *inverse model* in control system design. A controller receives the desired sensation $y^*([n])$ as input and must find actions that cause actual sensations to be as close as possible to the desired sensation. The controller must invert the transformation from actions to sensations.

We developed this mechanism based on the neural network architecture proposed by Jordan and Rumelhart [Jordan and Rumelhart, 1992]. There are several reasons for using their approach. One of the advantages of this architecture is that we don't need an explicit teacher. The robot can use the difference between predicted position of the target and the next input of the target position as training data. Another reason is that the architecture is capable of addressing the *many-to-one mapping* problem from actions to sensations. The robot shown in Figure 2 must control at least two parameters, one for the camera and the other for the trunk to track the target. So there are infinite number of possible inverse models[4].

Using these features, the mechanism learns to produce appropriate motor command $u_v[n-1]$ to keep the target in the center of the visual image, given the current state $x[n-1]$ and a motor command to move to the next position $u_m[n-1]$.

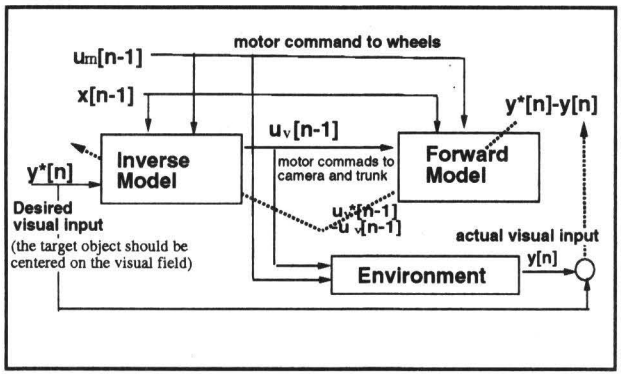

Figure 3: The learning mechanism for visual tracking with locomotion.

Figure 3 shows the learning architecture for visual tracking. $y^*([n])$ denotes a desired position of the target in the visual image, $u_v^*[n-1]$ denotes a proper command to the movable camera. $y([n])$ denotes the actual position of the target in the visual image, and $u_v[n-1]$ denotes the actual motor command for the movable camera.

In order to learn the inverse model to keep the target in the center of the field-of-view, we need the difference between the proper command $u_v^*[n-1]$ and actual command $u_v[n-1]$ to adjust the motor command:

$$u_v^*[n-1] - u_v[n-1] \quad (4)$$

We used the method described in [Jordan and Rumelhart, 1992]. Firstly the robot learns a forward model based on the difference between $y^*[n]$ (the output forward model) and $y[n]$. Here the difference (4) can be acquired by backpropagating the difference between $y^*[n]$ and $y[n]$ through the forward model. Then, the robot learns the inverse model based on the difference (4).

The Network Architecture

We implemented the above learning architecture using a feedforward network based on the block diagram shown in Figure 3. The network is composed

[3]The motor command can be seen as the command to move the robot's two wheels.

[4]See [Jordan and Rumelhart, 1992] for more details.

of two subnetworks: one for the inverse model; and the other for the forward model. The inverse model consists of 5 input, 15 hidden and 1 output units. The forward model consists of 6 input, 15 hidden and 3 output units. The output of the inverse model is taken as input to the forward model.

Mental Tracking via Acquired Knowledge

Figure 4 illustrates the way to mentally track the target using the learned visual tracking. The shaded portion denotes the acquired knowledge for visual tracking. The command for moving around in the environment is denoted as $u_m[n-1]$.

Note that the robot uses the output of the forward model as the current state $x^+[n-1]$, instead of actual input from the camera (environment). In order to mentally track the target with locomotion, the robot produces virtual command $u_v^+[n-1]$ to the forward model as an internal process.

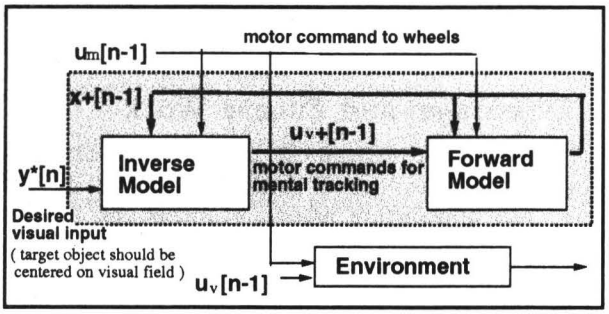

Figure 4: The architecture for the mental tracing.

4 Empirical Experiments with a Simulated Robot

4.1 The Setup

In the following experiments, we simulated three stages of a child's development of motor skills with the robot by varying its permitted actions. In stage 1, the robot is only permitted head rotation. In stage 2, the robot can rotate both head and body. Finally, in stage 3, the robot is also permitted self-locomotion, whereas in stages 1 and 2, locomotion was performed by an external agent.

For each stage, the forward and inverse models of the network were trained until:

Forward model Prediction error was less than 0.0005 for 50 training steps, or 50000 training steps were completed.

Inverse model The difference between the proper command $u_v^*[n-1]$ and actual command $u_v[n-1]$ was less than 0.0005 for 50 training steps, or 50000 training steps were completed.

4.2 The Three Cups Task

Following the "three cups" paradigm [Bower, 1979] discussed previously, the robot is placed in front of three cups and shown which cup hides the target object[5]. The robot moves (or is moved) to a new position from which it must predict which cup hides the target object.

4.3 Results

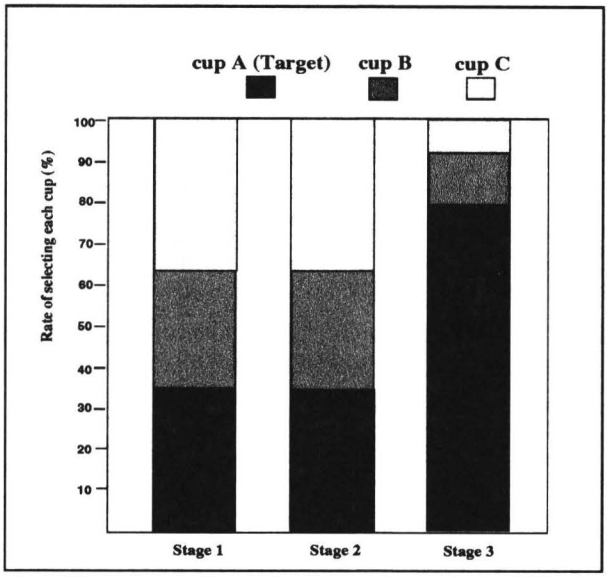

Figure 5: Prediction rate after each stage (average over 10 runs).

Figure 5 shows the rate of selection one of three cup after training for each of the three stages. The target cup (cup A) is black. The robot's performance at stages 1 and 2 was at chance level (35%). However, analysis showed that responses were consistent with an egocentric based prediction, not random choice. For example, from an allocentric perspective, a cup on the left-hand side of one's field of view would appear on the right-hand side if one views the cups from behind. From an egocentric perspective, however, one would predict the target as being on the left-hand side, which was the robot's prediction for stages 1 and 2. Since the cups were arranged in the shape of an equilateral triangle, only 35% of positions will yield a correct prediction based on egocentric knowledge. Under random choice there is no correlation between the relative

[5]This was done by labeling the target cup at the initial location.

positions of the selected cups before and after movement. In stage 3, when the robot also had control of translational movement, its predictive accuracy was above chance and egocentric levels, and more consistent with an allocentric based choice. Thus, locomotion experience was important for learning to predict the target's position.

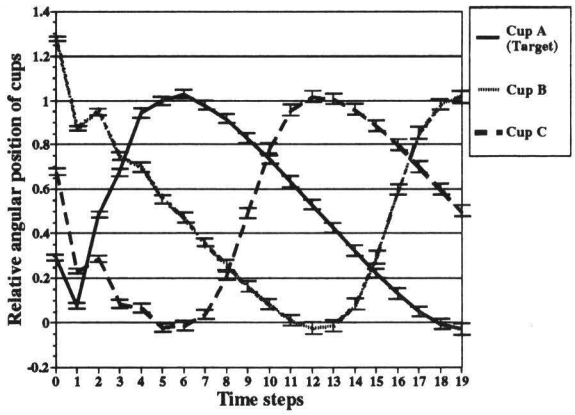

Figure 6: Relative angular position of cups as a function of robot location at the end of stage 1.

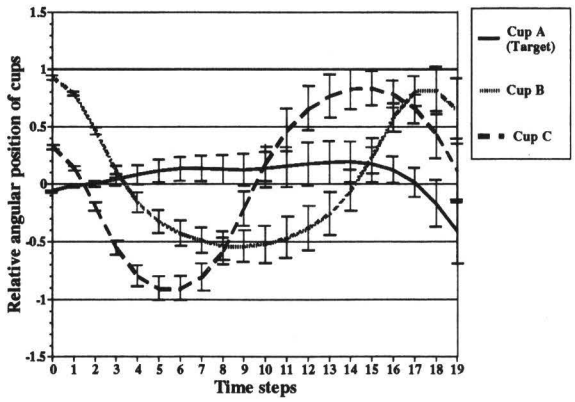

Figure 7: Relative angular position of cups as a function of robot location at the end of stage 3.

The contrast between egocentric behaviour (at stage 1) and allocentric behaviour (at stage 3) is made clearer by plotting cup position in the robot's field-of-view as a function of the robot's location.

Figure 6 shows the angular position of each cup relative to the center of the robot's field of view at various locations around the cups. The robot was moved along the circumference of a circle enclosing the cups (see Figure 1), and cup positions were recorded at 1/20th intervals. For example, 0 on the x-axis corresponds to the robot's initial position. For the y-axis, negative and positive values correspond to the left and right halves of the field-of-view (respectively). For each location, the cup with the smallest angular position (in magnitude) is the selected cup. For example, at position 5, cup C was selected. As evident from Figure 6, the robot always selected the leftmost cup as the target. In other words, the robot adjusted its head so that the leftmost cup (relative to the robot) was positioned at the center of the field-of-view (0 angular position). Consequently, the other two cups appeared in the right half of the field-of-view. This is to say that the robot behaved egocentrically. Crossovers on the graph (e.g., position 8) occurred because there are 6 locations on the circumference of the enclosing circle for which one cup is occluded by another.

Figure 7 shows mental tracking at the end of stage 3. As can be seen from the graph the target cup (cup A) remains closest to the center of the robot's field-of-view for most locations (i.e., it behaved allocentrically). Again, the exceptional cases (i.e., crossovers) are due to occlusions.

5 Discussion and Future Work

So far we have described mental tracking as a computational model of infant's spatial development. The simulation results support the evidence found in developmental psychology: the importance of self-locomotion. Furthermore, the results of our simulation suggest that experience of self-locomotion *with visual tracking* can accelerate spatial development. This offers the key to an understanding of *how* egocentric behavior changes into allocentric behavior.

As for mental tracking, Tani proposed a similar idea in the context of robot navigation [Tani, 1995]. He developed a robot that is capable of *mentally simulating* action plans based on a forward modeling scheme using recurrent network learning. Although he was not concerned with cognitive development, he did suggest its relevance to cognition.

Perhaps the closest approach to ours is the Cog Project of Brooks and colleagues [Brooks and Lynn, 1993]. They have been developing human-body like robots, *humanoid.* The idea of creating humanoids to investigate human cognition is very attractive in a sense that *intelligence cannot come without body.* We believe that the key concept of using robot for modeling cognitive development can be achieved even with simple mobile robots.

Current simulations were limited to one step back in time (i.e., target visible from the previous time step). For more complex environments, the target will be outside the field-of-view for indef-

inite periods. An obvious (and elegant) solution is to incorporate recurrent connections and have the network learn to remember positional information (e.g., [Elman, 1990]). However, learning to maintain information over long periods in the absence of additional input is difficult without special learning techniques (e.g., incremental learning, [Elman, 1993]). The extent to which mental tracking is maintenance of internal representations, or the search for alternative visual cues is an interesting research issue. And one that can best be addressed in real-world active environments such as we have proposed with the use of robots.

In our simulations, we divided child's development of motor skills into three stages. In general, however, motor skills develop more gradually, and interact with spatial development more tightly. More likely is that the development of motor skills and spatial knowledge interact both ways [Thelen and Smith, 1994]. We need to explore this kind of interaction in future work.

6 Conclusion

In this paper, we addressed the question of how infants relate to their spatial environment, and how this changes as the infant matures. To explore these issues, we introduced mental tracking as a key concept, and propose a learning architecture for mental tracking analogous to infants. Although there is much work to be done, we believe that the idea of using autonomous robots as the subject of development will open a new approach to modeling cognitive development. We take inspiration from recent work in robot vision, where the problem of making a robot see generated predictions leading to discoveries in insect vision [Franceschini *et al.*, 1992]. We expect similar results for cognitive development.

Acknowledgement

We thank Hideki Asoh for his comments on initial stage of this work. We also thank Motoi Suwa, Kazuhisa Niki and Hideyuki Nakashima for their support.

References

[Acredelo *et al.*, 1984] L.P. Acredelo, A. Adams, and S.W. Goodwyn. "The role of self-produced movement and visual tracking in infant spatial orientation". *Journal of Experimental Child Psychology*, 38:312–327, 1984.

[Acredelo, 1979] L.P. Acredelo. "Laboratory versus home: The effect of environment on the 9-month-old infant's choice of spatial refence system". *Developmental Psychology*, 15:666–667, 1979.

[Bower, 1979] T.G.R. Bower. *"Human Development"*. W.H. Freeman and Company, 1979.

[Brooks and Lynn, 1993] R. A. Brooks and A. S. Lynn. "Building Brains for Bodies". Technical report, MIT. AI Lab, 1993.

[Dresher, 1991] G.L. Dresher. *"Made-up Minds: A Constructivist Approach to Artificial Ingelligence"*. MIT Press, 1991.

[Elman *et al.*, 1996] J.L. Elman, E.A. Bates, M.H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. *"Rethinking Innateness: A Connectionist Perspective on Development"*. MIT Press, 1996.

[Elman, 1990] J. L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[Elman, 1993] J. L. Elman. Learning and development in neural networks: The importance of starting small. *Cognition*, 48:71–99, 1993.

[Franceschini *et al.*, 1992] N. Franceschini, M. Pichon, and C. Blanes. From insect vision to robot vision. *Phil. Trans. R. Sac. Lond.*, B337:283–294, 1992.

[Jordan and Rumelhart, 1992] M.I. Jordan and D.E. Rumelhart. "Forward Models: Supervised learning with a distal teacher". *Cognitive Science*, 16:307–354, 1992.

[Kermoian and Campos, 1988] R. Kermoian and J.J. Campos. "Locomotor experience: A facilitator of spatial cognitive development". *Child Development*, 59:908–917, 1988.

[Klahr and Wallance, 1976] D. Klahr and J.G. Wallance. *"Cognitive development: An information processing view"*. NJ: Erlbaum, 1976.

[Piaget, 1971] J. Piaget. *"The construction of reality in the child"*. New York: Ballantine., 1971.

[Tani, 1995] J. Tani. "Self-Organization of Symbolic Processes through Interaction with the Physical World". In *Proc. of IJCAI'95*, pages 112–118, 1995.

[Thelen and Smith, 1994] E. Thelen and L.B. Smith. *"A Dynamic Systems Approach to the Development of Cognition and Action"*. MIT Press, 1994.