Examining the Role of Prediction in Infants' Physical Knowledge

Matthew Schlesinger (MATTHEWS@Siu.Edu)
Department of Psychology, Southern Illinois University
Carbondale, IL 62901 USA

Michael E. Young (MEYOUNG@Siu.Edu)
Department of Psychology, Southern Illinois University
Carbondale, IL 62901 USA

Abstract

The violation-of-expectation paradigm investigates infants' physical knowledge by exploiting their tendency to look longer at events that are surprising, unexpected, or physically impossible. The current simulation study examines the role of prediction as a fundamental component of infants' expectations in physical-knowledge studies. A recurrent network is presented with a computer-animated version of Baillargeon’s “car study” (1986; Baillargeon & DeVos, 1991), in which a car rolls down a ramp and behind a screen. After learning to predict the outcome of a training event, the model is then tested on possible and impossible events from the same study. During testing, the model successfully predicts only superficial features of the test events. These results are used to argue for the necessity of prior physical knowledge, and perhaps also a built-in capacity for mental representation, in order for a prediction system to work.

Introduction

Over the last 20 years, developmental researchers have mounted a broad and compelling challenge to Piaget's theory of infant cognitive development (e.g., Baillargeon, 1995; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Much of this research has focused on two particular elements of Piaget's theory: first, that infants' physical knowledge (i.e., their concepts of objects, space, time, and causality) depends on sensorimotor experience, and second, that the capacity for mentally representing the world develops gradually over the first two years (Piaget, 1952).

In contrast, the "competent infant" view argues that Piaget underestimated what young infants know and understand about the physical world. This approach is based on three closely-related ideas. The first is that infants' visual expectations are guided by a core set of intuitive or naïve physical principles (e.g., that solid objects move along continuous paths; Spelke et al., 1992). In addition, Spelke and others have argued that this core knowledge may either be innate, or develop too early in infancy to depend on input from sensorimotor experience.

The second idea is that the ability to mentally represent the world is also present early in life, if not innate (e.g., Baillargeon, 1986; Meltzoff & Moore, 1998). This capacity is exploited by infants in a variety of ways, including mentally tracking occluded objects (e.g., Carey & Xu, 2001), and also reasoning about the physical properties of those objects while they are out of sight (e.g., size and location; Baillargeon & DeVos, 1991).

The third idea is built on the first two: during everyday experiences, infants exploit both their prior knowledge and capacity for representation as they generate predictions for the events they observe. This tendency to forecast or predict the outcome of events has helped motivate the predominant methodology for studying infants' physical knowledge, that is, the violation-of-expectation (VOE) paradigm. Specifically, the VOE paradigm proposes that infants will increase their attention toward events that violate their understanding of the physical world, or in other words, events that are surprising, unexpected, or physically impossible (e.g., Baillargeon, 1993; Spelke, 1985).

Learning by Prediction

A number of developmental theorists have highlighted the role of prediction-learning as a developmental mechanism, and in particular, a wide variety of connectionist models have implemented this idea in an artificial neural network (e.g., Elman, 1990; McClelland, 1995). Prediction-learning is typically simulated by training a neural network to predict a sequence of stimuli (e.g., speech segments) as the sequence is presented one element at a time. The success of these models, which have no built-in knowledge, suggests that prediction-learning can function without an *a priori* knowledge base. In addition, mental representation need not play a central role, at least in a strong form (e.g., internal symbols, recall memory, etc.). However, weaker forms of representation may be necessary for supporting a prediction-learning system. For example, in the face of an ambiguous stimulus, a sensory trace can provide a form of implicit memory that facilitates predicting the next experience (e.g., in a recurrent network; Mareschal, Plunkett, & Harris, 1999; Munakata, McClelland, Johnson, & Siegler, 1997).

Two recent models explore the role of prediction by simulating the development of object-oriented behaviors in infants (i.e., visual tracking and reaching; Mareschal et al., 1999; Munakata et al., 1997). In particular, these models simulate the ability to track the movement of an object while it is briefly occluded. Although there are important differences between the architectures and learning algorithms employed by Mareschal and Munakata, both models rely on a comparable learning rationale.

Specifically, a recurrent network (i.e., a feed-forward network that also includes an additional input loop from the hidden layer back to the input layer; see Elman, 1990) is presented with the event sequence, one "frame" at a time, and the task of the network is to learn to predict the next step in the sequence (using backpropagation-of-error as a learning algorithm). Both models demonstrate that recurrent feedback can function like an internal sensory trace, helping the model to predict the reappearance of the target while it is occluded.

The current investigation extends the work of the Mareschal and Munakata models, by asking whether a recurrent network that learns by prediction--but that has no prior knowledge--can generalize what it learns to either possible or impossible events. By analogy, to what extent do infants' reactions in VOE studies depend on prediction-learning mechanisms versus prior knowledge of the physical world? Therefore, the goal of this paper is to decouple these two processes, and to focus on the role of prediction during possible and impossible events.

The rest of the paper is organized as follows: In the next section, we briefly describe Baillargeon's "car study", which provides a platform for investigating the role of prediction in VOE studies. We then provide an overview of the prediction model, which first learns to predict the outcome of a computer-animated training event, and is then tested on possible and impossible events from the same study. Next, Simulations 1 and 2 examine the model's ability to generalize to the novel test events. Finally, we discuss the performance of the model, and relate the findings to current debates in early infant cognition.

The "Car Study"

Baillargeon (1986; Baillargeon & DeVos, 1991) studied infants' knowledge of the permanence and solidity of objects by presenting young infants with a simple mechanical display, in which a screen is raised then lowered, and then a car rolls down a ramp, passing behind the screen and reappearing on the other side (see Figure 1A, *Habituation event*). Infants watched this event repeat several times until they habituated (i.e., grew disinterested and began to look away). After habituating, infants then saw two test events in alternation (see Figures 1B and 1C). During both the *Possible* and *Impossible* test events, a box is revealed behind the screen. During the Possible event, the box appears behind the track; during the Impossible event, however, the box is placed on the track, in the path of the car. Nevertheless, during both test events the car reappears after passing behind the screen.

Baillargeon found that by at least age 6 months, and perhaps earlier, infants look significantly longer at the Impossible event than the Possible event. These findings were replicated in a follow-up study, in which infants saw the car placed on (Impossible) versus in front of the track (Possible). She interpreted these results to suggest (a) that infants mentally represent both the car and the box while they are occluded, (b) that they do not expect the car to reappear during the Impossible event, and (c) they consequently look longer at the Impossible event because it violates their expectations.

The Prediction Model

Note that the events in Baillargeon's car study pose at least two challenges for a prediction model. First, like the events simulated by Mareschal and Munakata, there is a moving object that is briefly occluded. Second, there is also a potential causal interaction between the car and box (i.e., obstruction or collision), which is also occluded. While the occluded movement may be predictable, it is not clear what experiences may be necessary for correctly learning to predict an occluded obstruction or collision event.

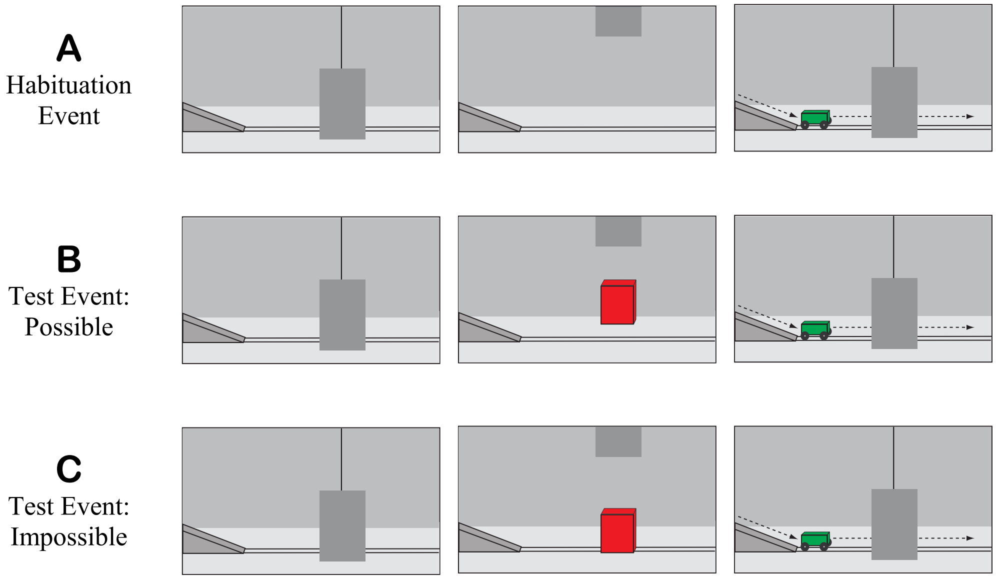

Figure 1: Schematic display of the Habituation (A), Possible (B), and Impossible (C) events studied by Baillargeon (1986; Baillargeon & DeVos, 1991).

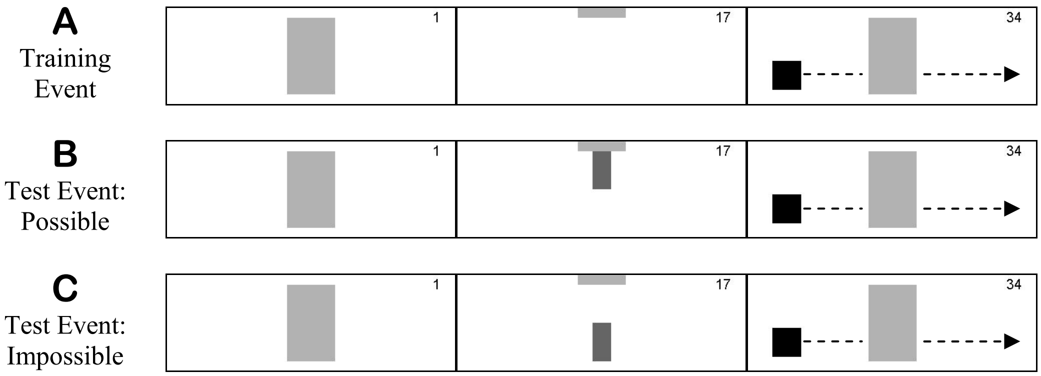

Figure 2: Selected frames from the animation events used in Simulation 1 to train (A) and test (B-C) the prediction model (frame number displayed in upper right corner).

From a design standpoint, the current prediction model shares a number of features with both the Mareschal and Munakata models. First, like the Mareschal model, the prediction model receives as input a 2-dimensional array of pixel values from an animation event, which is projected onto a simplified retina. Second, like the Munakata model, the input is propagated through a hidden layer, and then to an output layer with the same number of units as the input layer. On each timestep, the task of the network is to produce as output the pattern of pixel values that correspond to the next animation frame. Third, like both of the prior models, the observed values on the next frame provide the basis for a "teaching signal", which is used to adjust the connection weights in the network.

Stimuli

Three animation events were designed as 2-dimensional analogs to those in Baillargeon's car study. Figure 2 presents selected frames from these events, which were used to train and test the prediction model. Each event is 82 frames in duration. Note that unlike Baillargeon's study, the model is trained rather than habituated. Consequently, the habituation event is renamed the Training event in the prediction model.

The Training and Test events were rendered in grayscale, with the "car" pixel values represented as 1.0, the "box" values as 0.6, the "screen" values as 0.3, and the background as 0. The entire event display is 10 pixels tall by 30 pixels wide, for a total of 300 pixels.

Architecture

The prediction model is implemented with a simple recurrent network (SRN). The SRN has three layers that are fully connected. First, the input layer (300 units) operates like a simple retina; each input unit is activated by a single corresponding pixel in the 10-by-30 animation display. Second, the input layer feeds forward to a hidden layer (20 units), which not only feeds forward to the output layer, but also sends a set of activations back to the input layer (i.e., via recurrent connections). Consequently, on each timestep the hidden layer of the SRN receives signals from both the input layer as well as from itself (i.e., from the hidden layer activation values during the previous timestep). Finally, the output layer is the same size as the input layer (300 units).

Learning Algorithm

The backpropagation-of-error learning algorithm was used to train the SRN. Specifically, on each timestep the SRN received as input one frame from the animation sequence. The corresponding output for that timestep was then compared to the input frame for next timestep, using the mean-squared difference between expected and observed pixel values as the error metric (i.e., mean-squared error or MSE).

Simulation Overview

Simulations 1 and 2 employed the same training and testing regime. In each case, the SRN was first presented with the Training event. The results of pilot simulations suggested that 300 training trials were sufficient to reduce the MSE per pixel to approximately 0.01 (recall that pixel values ranged from 0 to 1). Therefore, training continued for 300 trials (i.e., repetitions of the training event). After completing training, each SRN was tested on the Possible and Impossible events. Note that for each simulation, 50 replications of the SRN were randomly initialized, trained, and tested.

Performance Measures

Recall that the VOE paradigm relies on *looking time* (i.e., the amount of time spent fixating a stimulus or event) as an index for infants' expectations. However, the prediction model does not produce overt eye movements or fixations (cf. Schlesinger, in press). Nevertheless, there are a variety of ways in which both the model's internal activity and output can be viewed as computations that would precede and possibly modulate an attentional signal (e.g., gaze control in the superior colliculus, tracking of motion in area MT, etc.). Two such performance metrics are employed in the current model.

First, prediction-errors (i.e., MSE) in the model can be interpreted as an influence on looking behavior (e.g., Mareschal et al., 1999; Munakata, 1997). That is, when discrepancies occur between predicted and observed inputs, we should expect infants to continue monitoring an event, until their predictions agree with their observations. This is, of course, the rationale of the VOE paradigm.

Second, we can also compare the model's hidden-layer activations across events. Specifically, the internal activation pattern during the training event can be interpreted as a template or sensory encoding, against which the test events are compared (e.g., Mareschal et al, 1999; for a discussion of template-matching as a developmental mechanism, see Charlesworth, 1969). Much like prediction errors, when differences occur between the encoding of the training event and a test event, that test event is assumed to be novel, and therefore, should increase attention.

Simulation 1

Simulation 1 follows the general procedure of Baillargeon's (1986) Experiment 1, in which, during the test phase, infants see the car placed either on or behind the track, respectively (i.e., Impossible or Possible event).

Method

Fifty replications of the model were trained and tested. During each replication, an individual SRN was first initialized with random connection weights (in the range -1 to 1). Next, the SRN was presented with the Training event (see Figure 2A), one frame at a time. For each input frame, the model produced as output a corresponding set of pixel values that were a prediction for the next input frame.

After each output was generated, it was then compared to the next input frame. MSE was computed by comparing the difference between predicted and observed pixel values, and was then minimized by adjusting the connection weights of the network with the backpropagation-of-error learning algorithm. Learning was terminated after 300 repetitions of the Training event. Connections weights in the SRN were then "frozen" (i.e., learning was turned off).

During the test phase, three events were presented. First, in order to establish a prediction-error baseline, the Training event was re-presented; in this case, to distinguish between the SRN's reactions during the training and test phases, this event was called the Control event. Next, the Possible and Impossible events were presented (see Figures 2B and 2C), corresponding to Baillargeon's study in which the box was placed on or behind the tracks, respectively.

Results

As proxies for looking time, analyses focused on prediction errors and similarity between hidden-layer patterns. First, as noted above, *mean prediction error* was computed as the MSE per pixel in the output layer (averaged over the 82 frames of animation during each test event).

Figure 3A presents the MSE per pixel as a function of test event. Mean prediction errors were 0.014, 0.015, and 0.019 during the Control, Possible, and Impossible events, respectively. All three events were significantly different. Specifically, prediction error was higher during the Possible than the Control event ($t(49) = 36.60$, $p < .001$), but lower during the Possible than the Impossible event ($t(49) = 37.57$, $p < .001$).

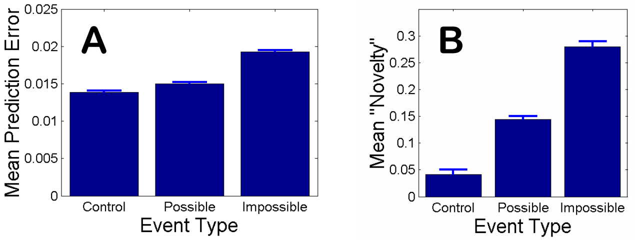

Figure 3: Mean prediction error (A) and "novelty" (B) in Simulation 1, during the Control, Possible, and Impossible events. Error bars indicate 95% confidence intervals.

Next, we computed Euclidean distance between Training and the Control, Possible, and Impossible events. First, as a baseline, hidden-layer activations during the last five training trials were pooled and averaged, resulting in an 82x20 (i.e., frames by hidden-layer units) matrix. Second, Euclidean distance was then computed, using comparable activation values during the Control, Possible, and Impossible events.

Figure 3B presents the mean "novelty" (i.e., Euclidean distance) during the test phase (averaged over the 82 frames). Note that higher novelty corresponds to greater dissimilarity or distance between the Training and test event. Mean novelty was 0.04, 0.14, and 0.28 for the Control, Possible, and Impossible events, respectively. As before, all events were significantly different. In particular, the Impossible event was significantly more novel than the Possible event ($t(49) = 20.63$, $p < .001$).

Discussion

Both sets of analyses provide convergent results. In particular, the prediction model produces (a) higher prediction errors and (b) a more novel or dissimilar pattern of internal activity, during the Impossible event.

A preliminary conclusion based on these findings is that prediction-based learning that occurs during the Habituation event may be sufficient to explain infants' greater attention to the Impossible event during the test phase. A related conclusion is that prior physical knowledge (i.e., naïve physics) does not seem necessary to explain why infants look longer at the Impossible event in the car study.

However, a potential qualification to these results is that the box appears at different times, and for different durations, during the two test events. Specifically, it is revealed sooner and for more time during the Impossible event. Recall that Baillargeon addressed this confound by testing infants in a second condition, in which the car was placed either on (Impossible) or in front of the tracks (Possible). Therefore, in order to replicate and extend the current results, Simulation 2 modifies the trajectory of the car, so that now (in contrast to Simulation 1) the box appears first, and for more time, during the Possible event.

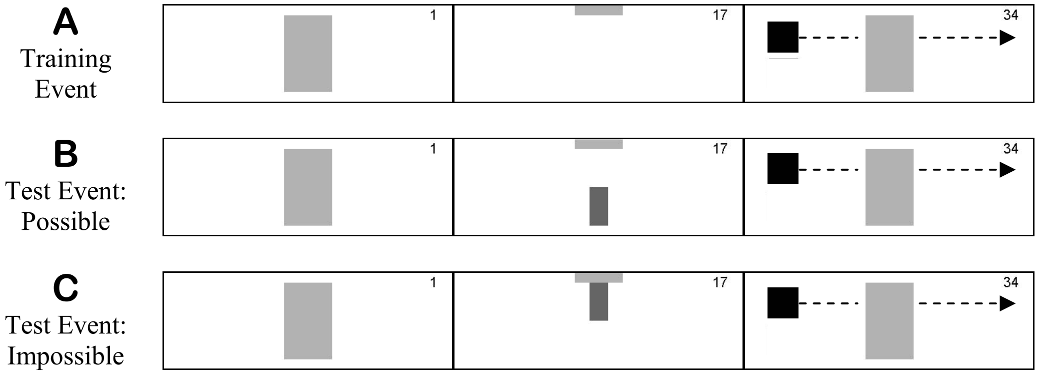

Figure 4: Selected frames from the animation events used in Simulation 2 to train (A) and test (B-C) the prediction model (frame number displayed in upper right corner).

Simulation 2

While the results of Simulation 1 suggest that the Impossible event may be more difficult to predict, a careful examination of Figure 2 shows that in fact the Possible and Impossible events are nearly identical. In fact, the only visible difference is the location and timing of the box's appearance during the two test events. One possibility is that the Impossible event is more "surprising" because the box is revealed in the car's trajectory. Alternatively, it is because the box, a novel object, is revealed sooner and for more time during the Impossible event.

Simulation 2 investigates this second account by moving the car's trajectory to the upper half of the display (see Figure 4). Thus, the box is now revealed later and for less time during the Impossible event. If in fact the model acquires some kind of expectations or general knowledge about physical objects during the training phase, the Impossible event should still generate greater prediction errors and be more dissimilar to the Training event. Alternatively, if the prediction model is simply reacting to the appearance of a novel object in the display, then the Possible event should now produce greater prediction errors and be more dissimilar to the Training event.

Method

The method of Simulation 2 was identical to that of Simulation 1, with one exception as noted above: specifically, the path of the car was modified so that it moved along the upper half of the display (see Figure 4). As before, 50 SRNs were trained and tested.

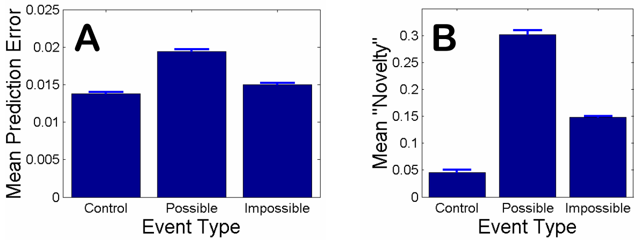

Figure 5: Mean prediction error (A) and "novelty" (B) in Simulation 2, during the Control, Possible, and Impossible events. Error bars indicate 95% confidence intervals.

Results

As Figures 5 indicates, the overall pattern of results in Simulation 2 was the mirror-image of that in Simulation 1. First, Figure 5A presents MSE per pixel during the test phase. Mean errors were 0.014, 0.019, and 0.015 for the Control, Possible, and Impossible events, respectively. In contrast to Simulation 1, prediction errors during the Impossible event were significantly *lower* than during the Possible event ($t(49) = 48.90$, $p < .001$).

Second, mean novelty of the Control, Possible, and Impossible events was 0.05, 0.30, and 0.15, respectively (see Figure 5B). Paralleling the previous analysis, the Impossible event was significantly less novel than the Possible event ($t(49) = 26.46$, $p < .001$).

Discussion

Unlike the results of Simulation 1, those of Simulation 2 suggest that the Possible event should not only be more difficult to predict, but also more dissimilar to the Training event than the Impossible event. Therefore, in this case the prediction model fails to replicate the findings of Baillargeon (1986; Baillargeon & DeVos, 1991), as it implies that infants in this condition should look longer at the Possible than the Impossible event.

General Discussion

Taken together, the results of Simulations 1 and 2 provide at best a partial replication of Baillargeon's car study experiments. However, even the success of Simulation 1 seems to raise more questions than it answers. In particular, why does the performance of the prediction model correspond with infants' looking time patterns, when (a) the model has no prior physical knowledge, and (b) it is presented with computer-animated events (i.e., that are not bound by the laws of physics)?

The answer to this question, perhaps obvious in retrospect, is that during training the prediction model learns to base its predictions, not on a set of underlying physical regularities or principles, but instead on superficial perceptual features of the event display. Specifically, the model's performance during the test phase is determined in large part by the appearance of the box. Therefore, in whichever event the box appears sooner and for more time (i.e., the Impossible event in Simulation 1, and the Possible event in Simulation 2), that event leads to greater prediction errors and appears more novel in comparison to the Training event.

Consequently, at least one implication of the prediction model is that in order to correctly predict or anticipate the outcomes of causal events (and consequently, be surprised when those predictions are violated), prior knowledge or experience may be necessary. Given this close and possibly necessary tie between causal expectations and prior knowledge, it is perhaps inevitable that some theorists have taken a strong theoretical stand in favor of innate, or at least very precocious physical knowledge in infants (e.g., Baillargeon, 1986; Spelke et al., 1992).

How might we incorporate prior knowledge into the prediction model, so that the appearance of novel objects such as the box has a negligible effect, while violations of basic physical principles (e.g., two objects in the same place at the same time) cause large prediction errors? Specifically, what would need to be added to the model in order to replicate Baillargeon’s findings? One solution would be to give the prediction model basic knowledge about the behavior of solid objects. This knowledge could be pre-programmed in any of several ways (e.g., via the network architecture, connection weights, etc.), or alternatively, learned through an appropriate series of pre-training experiences. For example, prior to watching the training event, the model could learn to predict the path of a car that approaches a fully visible obstacle (for an example of this training strategy, see Schlesinger & Barto, 1999). This prior knowledge would then provide a basis for correctly predicting when the car should reappear during the Possible and Impossible test events.

A related question concerns the fact that a large number of physical knowledge studies not only use the VOE paradigm, but also use occluded objects. Therefore, an additional implication of the prediction model is that a "strong" form of representation (e.g., counterfactual or hypothetical reasoning) may also be necessary, so that the prediction system can systematically generate predictions about events that are only partially observed.

Acknowledgments

This work was supported by a grant from the National Institute of Child Health and Human Development (1R03 HD40780-02).

References

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, *23*, 21-41.

Baillargeon, R. (1993). The object concept revisited: New directions in the investigation of infants' physical knowledge. In C.E. Granrud (Ed.), *Visual perception and cognition in infancy*. Hillsdale, NJ: Lawrence Erlbaum.

Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier and L.P. Lipsitt (Eds.), *Advances in Infancy Research* (pp. 305-371). Norwood, NJ: Ablex.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development*, *62*, 1227-1246.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, *80*, 179-213.

Charlesworth, W.R. (1969). The role of surprise in development. In D. Elkind & J. Flavell (Eds.), *Studies in cognitive development: Essays in honor of Jean Piaget*. Oxford, UK: Oxford University Press.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

McClelland, J.L. (1995). A connectionist perspective on knowledge and development. In T.J. Simon & G.S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Lawrence Erlbaum.

Mareschal, D., Plunkett, K., and Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviours in infancy. *Developmental Science*, *2*, 306-317.

Meltzoff, A.N., & Moore, M.K. (1998). Object representation, identity, and the paradox of early permanence: Steps toward a new framework. *Infant Behavior and Development*, *21*, 201-235.

Munakata, Y., McClelland, J.L., Johnson, M.H., and Siegler, R.S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, *104*, 686-713.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Schlesinger, M. (in press). A lesson from robotics: Modeling infants as autonomous agents. *Adaptive Behavior*.

Schlesinger, M., and Barto, A. (1999). Optimal control methods for simulating the perception of causality in young infants. In M. Hahn and S.C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. New Jersey: Erlbaum.

Schlesinger, M, and Parisi, D. (2001). The agent-based approach: A new direction for computational models of development. *Developmental Review*, *21*, 121-146.

Spelke, E.S. (1985). Preferential looking methods as a tool for the study of cognition in infancy. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life*. Norwood, NJ: Ablex.

Spelke, E.S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605-632.