Connection Science, Vol. 15, No. 4, December 2003, 271–280

Research Note

Where infants look when impossible things happen: simulating and testing a gaze-direction model

Matthew Schlesinger and Patrick Casey
Department of Psychology, Southern Illinois University, Carbondale, IL 62901, USA
email: matthews@siu.edu

Abstract. Schlesinger (2003, *Adaptive Behavior*, **11**: 97–107) recently proposed a model of eye movements as a tool for investigating infants' visual expectations. In the present study, this gaze-direction model was evaluated by: (a) generating a set of predictions concerning how infants distribute their attention during possible and impossible events; and (b) testing these predictions in a replication of Baillargeon's 'car study' (Baillargeon, 1986, *Cognition*, **23**: 21–41, Baillargeon and DeVos, *Child Development*, **62**: 1227–1246). We found that the model successfully predicts general features of infants' gaze direction, but not specific differences obtained during the possible and impossible events. The implications of these results for infant cognition research and theory are discussed.

Keywords: object permanence, infant cognition.

1. Introduction

One of the most important debates in the field of infant cognition focuses on the origin of naïve physics, that is, infants' implicit or intuitive knowledge of the physical world (e.g. objects, space, time and causality). At the centre of this debate is the *violation-of-expectation* (VOE) paradigm, the primary method for studying infants' physical knowledge (Spelke 1985, Baillargeon 1993).

A key assumption of the VOE paradigm is 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. Spelke 1985, Baillargeon 1993). For example, Baillargeon (1995) found that infants as young as 3.5 months looked significantly longer at a drawbridge that appeared to rotate through a solid box.

One way to interpret these results, consistent with the assumptions of the VOE paradigm, is that young infants have a precocious or possibly innate understanding of the physical world (e.g. Spelke 1998, Baillargeon 1999). As a consequence, this physical knowledge guides infants' perception in a primarily top-down manner: where and when infants look is determined by their *a priori* expectations of what they anticipate, predict, or expect to see.

Connection Science ISSN 0954-0091 print/ISSN 1360-0494 online
http://www.tandf.co.uk/journals
DOI: 10.1080/09540090310001655093

A growing number of researchers, however, have questioned the assumptions of the VOE paradigm (e.g. Haith 1998, Smith 1999). A common concern focuses on the role of perceptual differences between the possible and impossible events, which may influence infants' attention and visual processing. According to this bottom-up approach, infants may look longer at an impossible event, not because it violates their understanding of the physical world, but because one or more features of the event are perceptually salient and more interesting (e.g. Bogartz *et al.* 2000, Schlesinger 2003).

In order to investigate this bottom-up or perceptual-processing approach, we designed and tested a model that simulates infants' eye movements during possible and impossible events (Schlesinger 2003). The rationale of this gaze-direction model is that by presenting it with events like those seen by infants in a VOE study, we then obtain a simulated behavioural profile of infants' visual activity based on the premise that this activity is guided solely by perceptual processes (Schlesinger 2003). Next, careful replication and reanalysis of infants' visual activity (e.g. visual scanning patterns) during comparable events, followed by a comparison of these patterns with the profile generated by the model, allow us to estimate the role of perceptual processing 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 not only provides a platform for applying the gaze-direction model, but also allows us to test the predictions of the model in a sample of 6-month-old infants. In section 3, we provide an overview of the model, including the behavioural predictions it generates. Section 4 outlines the methods of the current study, while section 5 highlights the major findings. In the final section, we consider various interpretations of our findings, and discuss implications for infant cognition theory and research.

2. The car study

Figure 1 presents a schematic display of Baillargeon's car study (Baillargeon 1986, Baillargeon and DeVos 1991), which is designed to investigate whether young infants understand that: (a) occluded objects continue to exist while out of sight; and (b) two objects cannot be in the same place at the same time.

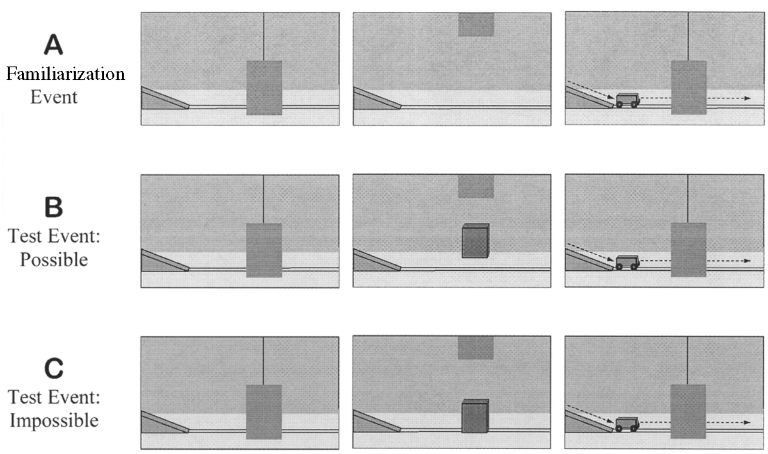

Figure 1. Schematic display of the (A) familiarization, (B) possible and (C) impossible events studied by Baillargeon (1986, Baillargeon and DeVos 1991).

In the car study, infants watch a simple mechanical display, in which a car rolls down a ramp, behind a screen, and out the other side. Figure 1(A) presents a schematic display of this *familiarization* event. Note that at the start of the familiarization event, the screen is raised in order to show the infant that nothing is behind it.

After watching several repetitions of the familiarization event, infants then see two test events in alternation (see figure 1(B, C)). During both the *possible* and *impossible* test events, a box is revealed behind the screen. 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 even earlier, infants look significantly longer at the impossible event than the possible event. How did she interpret these findings? First, she suggested that infants mentally represent both the occluded box and the car as it passes behind the screen. Second, she proposed that infants use these representations to 'compute' when the car should reappear, and are consequently surprised to see the car reappear during the impossible event, when its path is obstructed by the box. Thus, because the impossible event is surprising or anomalous to infants, they spend more time looking at it.

3. The gaze-direction model

As we noted earlier, the gaze-direction model was designed to simulate infants as they learn to watch simple mechanical events, such as a moving object that passes briefly behind a screen and reappears on the other side. There are three key elements of the model. First, it is completely naïve at the start of training (i.e. there is no prior or 'built-in' knowledge). Second, the model occupies a 'snapshot' world, and has no structures or mechanisms for representing the past or future (i.e. either memory or prediction systems). Finally, the model learns to track salient, moving objects by reinforcement (i.e. trial-and-error learning processes).

A related feature of the model is that it employs an *agent-based approach*, in which its input is determined in part by its previous 'actions' (i.e. outputs; Schlesinger and Barto 1999, Schlesinger and Parisi 2001, Schlesinger 2003). Specifically, the model has a simplified 'fovea' that changes its input over time by movement in two dimensions.

We present here a brief outline of the gaze-direction model, and refer the interested reader to Schlesinger (2003) for a more detailed description.

3.1. *Model overview*

The gaze-direction model is divided into three major components: the input, the architecture and the learning algorithm.

3.1.1. *The input.* Three computer-animated events were designed, corresponding to the familiarization and test events from the car study. These events were two-dimensional (i.e. 80-by-20 pixels), and rendered in greyscale. In order to abstract the key spatiotemporal relations from the car study, extraneous features were not included in the animated versions (e.g. the ramp and track). Similarly, many perceptual details from the real-world version were simplified (e.g. the car was replaced by a moving square).

3.1.2. *The architecture.* We used a fully-connected, three-layer neural network to simulate the model's oculomotor control system. On each timestep, one frame of the animation was projected on to the input layer, which was divided into two processing streams. One of the systems functioned like a fovea, and received high-resolution input (144 input units); the other system functioned like a periphery, and received lowresolution input (e.g. pixel values that were averaged over large patches of the animation frame; 33 input units).

The input layer projected from a hidden layer to an output layer. Activity on the output layer drove eye movements, which changed the position of the fovea with respect to the next animation frame. Thus, by generating an appropriate sequence of output values, the model could either fixate a stationary object, or track moving objects.

3.1.3. *The learning algorithm.* The model was first presented with the animated version of the familiarization event, and trained by reinforcement to track the movements of the car. Specifically, we used the SARSA learning algorithm, which belongs to a class of reinforcement-learning algorithms called *temporal-difference methods*. We briefly describe the SARSA algorithm here, and refer the interested reader to Sutton and Barto (1998) for a detailed introduction.

SARSA is an acronym for state-action-reward-state-action. The algorithm is based on the notion of a rational agent that interacts with its environment (i.e. state) by selecting actions that optimize rewards over time. In the current implementation, the model's state was defined as the set of activation values over the input units. Similarly, the model's possible actions were represented by the set of output units. (Note that the neural network therefore provides a *policy* or mapping from the present state to possible actions in that state.) Finally, the reward was defined as the proportion of the car visible within the fovea, after each eye movement. In general terms, the SARSA algorithm works by using the reward signal to strengthen pathways in the oculomotor control system (i.e. the neural network) that link a given state with a desired action (i.e. fixation of the car).

After learning to track the 'car' in the familiarization event, the model was then tested on the impossible and possible test events.

3.2. *Gaze-direction predictions*

Recall that our primary goal was to use the gaze-direction model as a testbed for simulating infants' visual activity in the car study. Consequently, a key constraint on our analysis of the model was that whatever performance measure we chose, we had to be capable of accurately measuring the same behaviour in human infants.

Ultimately, we decided to divide the animation events vertically into three equal-sized 'regions of interest', and then to compute the proportion of time that the model spent fixating each of these regions during the possible and impossible events. This analysis is presented in figure 2, which plots the proportion of fixations in the model toward the left, centre and right during the test events.

Figure 2 suggests two major predictions of infant gaze direction. First, if we average the impossible and possible events together, we find that the model spends most of its time fixating the centre of the display. Not surprisingly, this is where most of the 'action' is (i.e. both the screen and box are in the centre of the display, the screen moves up and down, etc.). The model fixates less often to the right, and least to the left. *Therefore, our first prediction is that infants should look most toward the centre, followed by the right and then the left.*

Second, we also note two differences in how the model fixates the impossible and possible test events. When processing the test events, the model spends more time fixating the centre in the impossible than the possible event. This pattern is reversed on the left, where the model spends more time fixating during the possible event. Meanwhile, there are no significant differences in fixations toward the right. *Consequently, our second prediction is that infants will discriminate between the possible and impossible test events, like the model, by their time spent fixating the left and centre of each event.*

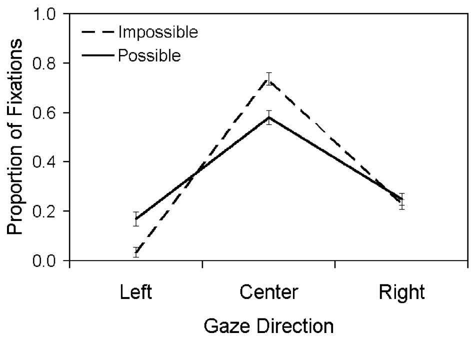

Figure 2. Mean proportion of fixations in the model toward the left, centre and right during the impossible and possible test events (error bars are ±1 SE).

4. Method

Except for a few minor differences, we followed closely the design and procedure of the original car study.

4.1. *Participants*

Twenty 6-month-old infants (mean age = 6 months, 6 days) participated. Prospective families were recruited from Southern Illinois birth records, and sent a letter inviting them to participate in a study of infant development.

4.2. *Apparatus and stimuli*

The apparatus was designed and constructed according to the specifications provided by Baillargeon and DeVos (1991). We therefore refer the interested reader to the original paper for a detailed description, including all dimensions, measurements, etc.

As illustrated in figure 1, the apparatus consisted of a stage or platform, which included: (a) a ramp on the left side; (b) a track located along the ramp and floor of the stage; (c) a car that ran along the track; (d) a box; (e) a small screen that occluded the centre of the stage, and that could be moved up and down; and (f) a curtain (not shown) that hid the entire stage when lowered. We also mounted a closed-circuit video camera just beneath the front and centre of the stage, in order to videotape each infant's fixations during the session.

Two experimenters worked in parallel to produce three types of events. First, during the *pretest event* the screen was in the raised position and no movement occurred. There were two versions of this event, in which the box was located in the centre of the stage, either on or behind the track (i.e. possible and impossible pretest events). Second, during the *familiarization event* the screen was raised then lowered, revealing an empty stage (i.e. there was no box present). After a brief delay, the car appeared on the left of the stage, rolled down the ramp and along the track, and exited on the right side of the stage.

Third, the *test event* was comparable to the familiarization event, with one exception. Specifically, the box was present, and positioned either on or behind the track (i.e. possible or impossible test events). Note that when the box was on the track, an experimenter briefly moved it out of the path of the car while the car passed behind the screen. This brief movement of the box was not visible from the infants' perspective.

4.3. *Procedure*

After arriving at the laboratory, each infant was allowed to manipulate the box for a few minutes as their parent completed a set of consent forms. The infant and parent were then brought to a darkened testing room, in which the only source of light was the illuminated stage area of the apparatus. Infants were seated approximately 60 cm from the stage.

All infants participated in a sequence of three phases. First, during the *pretest phase* infants viewed the static pretest events. The purpose of the pretest phase was to determine whether infants had a preference for seeing the box either on or behind the track (which might also influence their looking time during the test events). Accordingly, in the possible pretest event, the box was located behind the track; in the impossible pretest event, the box was located on the track (see figure 1). Each event continued either for a maximum of 20 s, or until the infant looked away for two consecutive seconds. The order of the two pretest events was counterbalanced across infants, with each infant randomly assigned to one of the two order conditions (i.e. impossible-first versus possible-first).

Next, the *familiarization phase* began. During this phase, infants watched the familiarization event, which continued to cycle either for a maximum of 45 s or, as before, until the infant looked away for two consecutive seconds. Infants were presented with three familiarization trials. Note that the use of three trials differs from the procedure used by Baillargeon (1986, Baillargeon and DeVos 1991), who habituated infants with a minimum of six and a maximum of nine trials.

The third and final phase was the *test phase*. Infants were presented with six test trials, that is, three possible and three impossible trials, in alternation. Note that during each trial, the test event repeated according to the same criteria as the familiarization trials. Impossible and possible test trials alternated according to the same order that each infant saw during the pretest phase.

4.4. *Data collection and coding*

During the experimental session, a trained observer watched the infant via closed-circuit video from an adjacent, soundproof room. Whenever the observer judged the infant to be looking toward the stage, they activated a switch that signalled a nearby computer. This signal was then used to control the onset and duration of the pretest, familiarization and test trials.

As noted above, sessions were recorded on video. Each videotape session was then transferred to digital format at the rate of 10 fps and analysed frame-by-frame. For the purpose of analysis, we defined two dependent variables. First, *looking time* was defined as the sum of fixations in seconds during a trial, toward any part of the stage. Second, as in the gaze-direction model, *gaze direction* was defined as the sum of fixations in seconds during a trial toward the left, centre, or right of the stage, respectively. Note that fixations away from the stage were excluded from both looking time and gaze direction.

A second observer coded 30% of the sample, selected at random. Interobserver reliability for the looking time measure, using the intraclass correlation, was 0.99 ($F(65, 65) = 139.12$, $p < 0.001$). Similarly, interobserver reliability for the gaze-direction measure was 0.95 ($F(197, 197) = 21.88$, $p < 0.001$).

5. Results

Two sets of analyses were conducted. First, we examined infants' looking times to determine whether infants looked significantly longer at the impossible test event (*looking time analysis*). Second, we decomposed looking time into three gaze directions (left, centre and right), in order to compare infants' distribution of fixations during the impossible and possible test events (*gaze-direction analysis*).

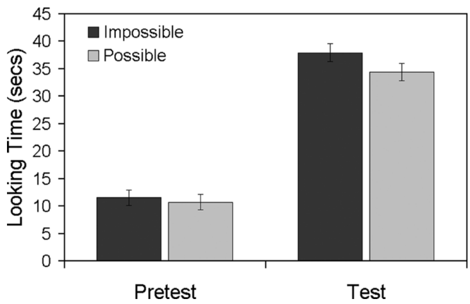

Figure 3. Mean looking time during the pretest and test phases to the impossible and possible events (error bars are ±1 SE).

5.1. *Looking time analysis*

Figure 3 presents mean looking times to the impossible and possible events during the pretest and test phases (error bars are ±1 SE). Our preliminary analysis focused on the pretest phase, in order to determine whether infants had a preference for seeing the box either on or behind the track. An analysis of variance on infants' pretest looking times revealed no significant differences between the two pretest events ($F(1, 18) = 0.73$, $p = \text{ns}$). However, there was a significant effect of event order. Specifically, infants in the possible-first group looked significantly longer to both pretest events than infants in the impossible-first group ($M = 13.9$ and 8.3 s, respectively; $F(1, 18) = 5.78$, $p < 0.05$; compare pretest phase of figure 4(A, B)).

We next analysed infants' looking times during the test phase. First, as figure 3 illustrates, infants looked significantly longer at the impossible test event ($F(1, 18) = 6.03$, $p < 0.05$). Note that this result is a replication of Baillargeon's main finding. Second, like Baillargeon, we also found a significant effect of trial (recall that there were six test trials in total). Infants tended to look less as the test phase progressed ($F(2, 36) = 4.17$, $p < 0.05$).

Finally, also in parallel with Baillargeon's results, there was a significant event × order interaction ($F(1, 18) = 24.44$, $p < 0.001$). In other words, infants' looking times to the two types of test events depended on which event was presented first during the test phase (recall that the impossible and possible events alternated in one of two counterbalanced orders).

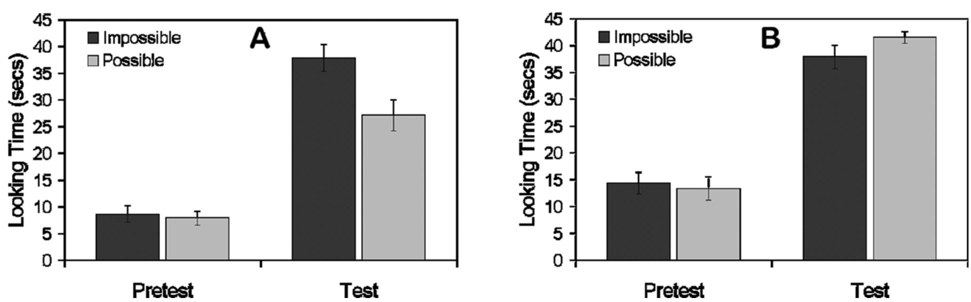

Figure 4. Mean looking time during the pretest and test phases to the impossible and possible events (error bars are ±1 SE), in the (A) impossible-first and (B) possible-first groups.

We pursued this interaction further by analysing the simple effect of event for each order group. Accordingly, figure 4 represents the results from figure 3, divided into the two order groups. As figure 4(A) illustrates, infants in the impossible-first group looked significantly longer at the impossible event ($F(1, 9) = 25.67$, $p < 0.01$). Meanwhile, infants in the possible-first group did not look significantly longer at either event (see figure 4(B); $F(1, 9) = 3.32$, $p = \text{ns}$). A comparable pattern of results was reported by Baillargeon (1986).

5.2. *Gaze-direction analysis*

Given that we successfully replicated Baillargeon's original study, we then focused our second set of analyses on the question of how infants distributed their fixations during the impossible and possible events.

The first analysis addressed the prediction that during both types of test events, infants would look most toward the centre, followed by the right and then the left of the stage. As figure 5 illustrates, infants' pattern of fixations were consistent with this prediction. Like the gaze-direction model, infants directed their gaze most toward the centre, with less time spent fixating the right, and the least time toward the left ($F(2, 38) = 46.67$, $p < 0.001$).

Next, we analysed the effect of event type on infants' gaze direction. Specifically, we determined whether there were any differences in infants' gaze direction during the possible and impossible events. We analysed the two order groups separately (i.e. impossible-first versus possible-first), since only the impossible-first group looked significantly longer at the impossible event (see section 5.1).

Figure 6 represents the results from figure 5, divided into the two order groups. As figure 6(A) illustrates, infants in the impossible-first group distributed their gazes differently during the impossible and possible events ($F(1, 9) = 16.30$, $p < 0.01$). Specifically, they spent more time during the impossible event looking toward the centre and right ($t(9) = 3.28$ and 4.96, respectively, and $p < 0.01$). However, in the possible-first group there were no significant differences in how infants distributed their fixations during the impossible and possible events ($F(1, 9) = 3.41$, $p = \text{ns}$; see figure 6(B)).

6. Discussion

The replication of Baillargeon's car study provides three major results. First, as in the earlier studies (Baillargeon 1986, Baillargeon and DeVos 1991), 6-month-old infants looked significantly longer at the impossible event. Like Baillargeon's original results,

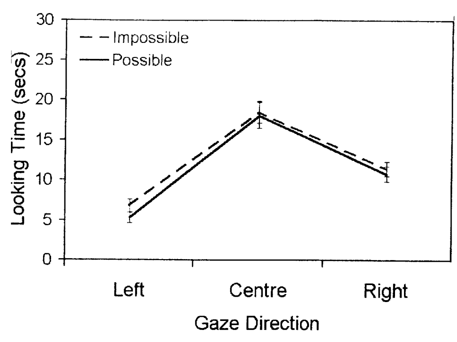

Figure 5. Mean looking time toward the left, centre, and right of the stage during the impossible and possible test events (error bars are ± 1 SE).

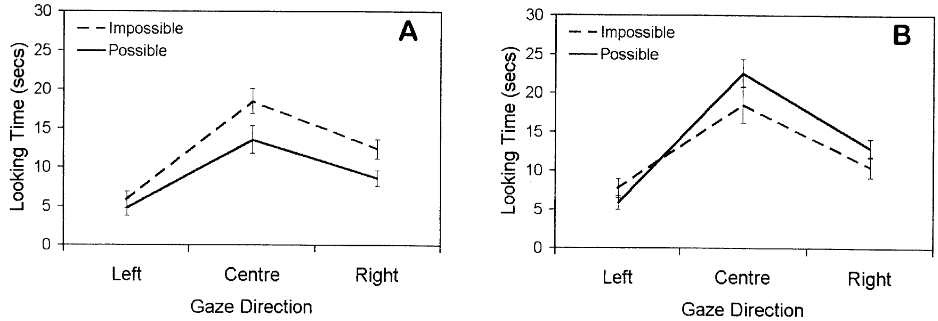

Figure 6. Mean looking time toward the left, centre and right of the stage during the impossible and possible test events (error bars are ±1 SE), in the (A) impossible-first and (B) possible-first groups.

we also found that presenting the impossible event first during the test phase enhanced this effect, while presenting the possible event first reduced or eliminated it.

Second, infants' overall gaze-direction patterns during the test phase were consistent with the model's prediction. Specifically, infants in both the impossible-first and possible-first groups spent most of their time watching the centre of the display, while looking less toward the right and the least toward the left. Therefore, this general correspondence between visual activity patterns in the model and human infants provides support for the perceptual-processing approach. In particular, not only the presence of salient objects at key locations in the display (e.g. the screen and box in the centre), but also the appearance and movement of the car seem to attract and guide infants' visual activity.

Third, however, the gaze-direction model was only partially successful in predicting differences between infants' gaze direction during the impossible and possible events. On the one hand, the model correctly predicted that when fixating the centre, infants would look more at the impossible event than the possible event. Unexpectedly, the same difference also emerged when infants looked toward the right.

It is not entirely clear why the gaze-direction model provided only a partial fit to infants' gaze patterns. There are a number of future research directions that may address this question.

First, additional behavioural features or capacities may be incorporated into the model. For example, there is a temporal dynamic quality to infants' gaze patterns (e.g. looking in one direction may bias looking either toward or away from other directions). One technique for studying this temporal effect in the model may be to include recurrent connections in the oculomotor control system, so that the model's gaze pattern depends on both current and previous perceptual inputs.

A second possibility is to contrast the performance of the gaze-direction model with other architectures that have been studied, including networks that learn to predict future inputs (e.g. Munakata *et al.* 1997, Mareschal *et al.* 1999, Schlesinger and Young 2003). Along these lines, one might also study the performance of a hybrid model that co-ordinates two visual processing streams (e.g. one for tracking the car and one for predicting future inputs).

In summary, our findings suggest three key conclusions. First, infants do not scan possible and impossible events in the same way. In particular, when they look longer at an impossible event, this increase in attention is directed at specific parts of the event, rather than uniformly distributed over the event. Second and more importantly, several qualitative features of these gaze patterns are predicted by a bottom-up gaze-direction model, which has no built-in memory capacities, prior knowledge, or conceptual skills. Taken together, these findings provide compelling support for the conclusion that infants' gaze patterns during surprising or unexpected events can be accounted for, at least in part, on the basis of bottom-up perceptual processing.

Acknowledgement
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* (Norwood NJ: Ablex), pp. 305–371.

Baillargeon, R., 1999, Young infants' expectations about hidden objects: a reply to three challenges. *Developmental Science*, **2**: 115–132.

Baillargeon, R., and DeVos, J., 1991, Object permanence in young infants: further evidence. *Child Development*, **62**: 1227–1246.

Bogartz, R. S., Shinskey, J. L., and Schilling, T. H., 2000, Object permanence in five-and-a half-month-old infants? *Infancy*, **1**: 403–428.

Haith, M. M., 1998, Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development*, **21**: 167–179.

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

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.

Schlesinger, M., 2003, A lesson from robotics: modeling infants as autonomous agents. *Adaptive Behavior*, **11**: 97–107.

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* (Hillsdale NJ: Erlbaum), pp. 625–630.

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

Schlesinger, M., and Young, M. E., 2003, Examining the role of prediction in infants' physical knowledge. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (Boston, MA, USA).

Smith, L. B., 1999, Do infants possess innate knowledge structures? The con side. *Developmental Science*, **2**: 133–144.

Spelke, E. S., 1985, Preferential looking methods as a tool for the study of cognition in infancy. In G. Gottlieb and N. Krasnegor (eds.) *Measurement of Audition and Vision in the First Year of Postnatal Life* (Norwood NJ: Ablex).

Spelke, E. S., 1998, Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, **21**: 181–200.

Sutton, R. S., and Barto, A. G., 1998, *Reinforcement Learning: An Introduction* (Cambridge MA: MIT Press).