Adaptive **Behavior**

The Neural Basis for Visual Selective Attention in Young Infants: A Computational Account

Matthew Schlesinger[1], Dima Amso[2], Scott P. Johnson[3]
[1]*Department of Psychology, Southern Illinois University Carbondale, Illinois, USA*
[2]*Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, Ithaca, New York, USA*
[3]*Department of Psychology, New York University, New York, USA*

Recent work by Amso and Johnson (*Developmental Psychology*, *42*(6), 1236–1245, 2006) implicates the role of visual selective attention in the development of perceptual completion during early infancy. In the current article, we extend this finding by simulating the performance of 3-month-old infants on a visual search task, using a multi-channel, image-filtering model of early visual processing. Model parameters were systematically varied to simulate developmental change in three neural components of visual selective attention: degree of oculomotor noise, growth of horizontal connections in visual cortex, and duration of recurrent processing in parietal cortex. While two of the three components—horizontal connections and recurrent parietal processing—are each able to account for the visual search performance of 3-month-olds, recurrent parietal processing also suggests a coherent pattern of developmental change in visual selective attention during early infancy. We conclude by highlighting plausible neural mechanisms for modulating recurrent parietal activity, including the development of feedback from prefrontal cortex.

Keywords object perception · selective attention · visual search · infant development

1 Introduction

The concept of *active vision* plays a central role in the constructivist theory of cognitive and perceptual development (Piaget, 1955, 1969). According to this view, organisms actively deploy their attention toward information-rich areas of the visual world, while exploiting efficient scanning strategies that shift attention from one surface, object, or location to another.

From a developmental perspective, active vision has direct implications for how human infants acquire the capacity to perceive, recognize, and internally represent objects. In particular, proponents of the active vision approach have argued that as infants gain skill in visual exploration, their knowledge of objects becomes progressively more complex and elaborated (e.g., Cohen, Chaput, & Cashon, 2002; Haith, 1980; Johnson, Slemner, & Amso, 2004; Piaget, 1955). Studies with human infants are complemented by work in machine vision and developmental robotics, which also highlights the role of guided visual exploration as a critical element of perceptual learning in artificial systems (e.g., Ballard, 1991; Sandini, Gandolfo, Grosso, & Tistarelli, 1993).

Correspondence to: Matthew Schlesinger, Department of Psychology, Southern Illinois University Carbondale, Carbondale, IL 62901, USA.
E-mail: matthews@siu.edu
Tel.: +1 618 453 3524; *Fax*: +1 618 453 3563

(2007), Vol 15(2): 135–148.
DOI: 10.1177/1059712307078661
Figure 5 appears in color online: http://adb.sagepub.com

The purpose of the current simulation study was to identify and investigate three neural mechanisms that may serve to link changes in visual-motor skill with parallel changes in object perception. In the next section, we introduce the concept of perceptual completion, and review evidence for the development of this capacity in young infants. In particular, we describe recent work that illustrates a connection between the emergence of perceptual completion and visual search performance in 3-month-old infants (Amso & Johnson, 2006; Johnson, 2004). Next, we describe a multichannel, image-filtering model that is used to simulate infants' visual search performance. Following a description of the model, we present the findings from a series of simulation studies that examine the influence of three neural constraints on the development of visual search in young infants (i.e., oculomotor noise, horizontal connections in visual cortex, and recurrent processing in parietal cortex). In the final section, we highlight the role of recurrent parietal processing as a neural constraint that may subserve the developments of visual selective attention and perceptual completion.

2 The Development of Perceptual Completion

A fundamental step in the development of object perception is the ability to perceive an object as complete when it is only partially visible. This capacity is known as *perceptual completion*. For example, consider the partially-occluded rod in Figure 1A, which moves laterally behind a screen. At birth, infants appear to lack the capacity for perceptual completion, and instead perceive displays like these as two separate surfaces that move at the same time. In contrast, by age 4 months, infants exhibit *unity perception*, that is, they perceive the display as a single rod that is partially occluded.

Evidence for this developmental pattern comes from a unity-perception task in which infants first see a display like Figure 1A. This display is presented repeatedly until infants habituate, that is, until their looking time falls below a predetermined threshold. After habituating, infants then watch two test events in alternation: during one (1B, the *complete-rod* display), a single rod moves laterally, whereas in the other (1C, the *broken-rod* display) two smaller, separate rods move laterally at the same time.

Note that a key assumption underlying the test phase is that after infants habituate to the occluded-rod display, they are expected to look longer at whichever test event appears more novel or dissimilar to the occluded-rod display (i.e., show a novelty preference; see Gilmore & Thomas, 2002; Sirois & Mareschal, 2002). In particular, the relative time that infants spend looking at each of the test events provides a behavioral index of whether they perceive the occluded rod (in the habituation phase) as one solid object (with two exposed segments), or alternatively, as two objects that move in parallel. In particular, 4-month-olds look significantly longer at the broken-rod display, indicating that the complete rod is more similar to the occluded rod. In contrast, newborn infants look significantly longer at the complete-rod display, suggesting instead at this age that the broken-rod display is more similar to the occluded rod (Johnson, 2004).

Amso and Johnson (2006) reasoned that if advances in visual-motor skill contribute to the development of perceptual completion, then the same underlying skill may also influence performance on other visual tasks. In particular, they proposed that *visual selective attention*—the ability to select and attend to specific stimuli while ignoring or inhibiting attention to alternative

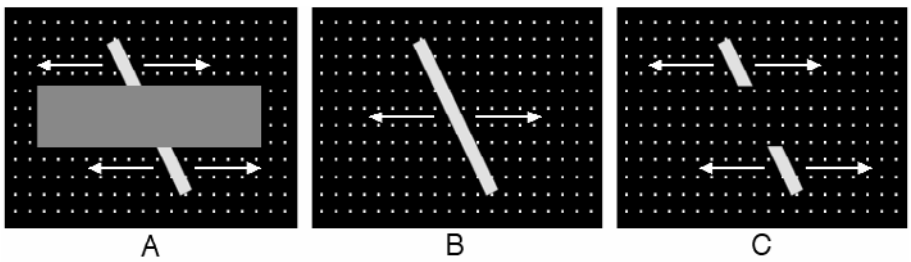

Figure 1 Displays used to assess perceptual completion in infants: (A) habituation display; (B) complete rod; and (C) broken rod test displays.

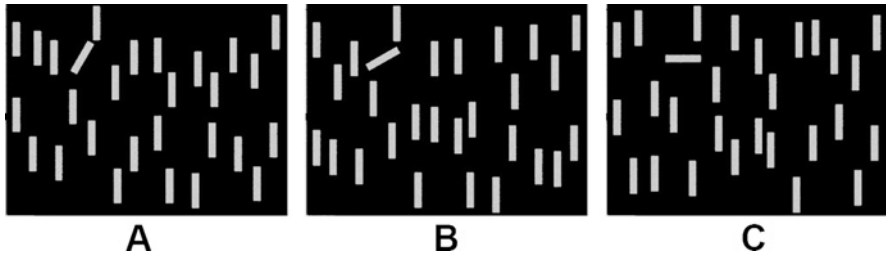

Figure 2 Three of the displays (from the orientation condition) used to assess visual search in infants: (A) 30° target display; (B) 60° target display; and (C) 90° target display.

or competing stimuli—is not only a critical component of visual-motor skill, but also should be reflected in infants' performance on both unity-perception and visual search tasks. As a result, they predicted that infants who show a more advanced level of perceptual completion (i.e., unity perception) should also perform more successfully during visual search.

In order to test this hypothesis, Amso and Johnson (2006) assessed both (a) perceptual completion and (b) visual search in a sample of 3-month-old infants. Three-month-olds were specifically chosen for study, not only because infants near this age appear to undergo a shift from reflexive to controlled visual attention (e.g., Johnson, 1990), but also because this age range represents a transitional period that precedes the development of unity perception.

Perceptual completion was assessed by presenting infants with the standard unity-perception task (see Figure 1): infants were first habituated to the occluded-rod display, and then presented with the broken-rod and complete-rod displays (in alternation) during the test phase. Looking time during the unity-perception task was determined by a trained observer, who viewed the infant's face on a remote television monitor and signaled a computer whenever the infant looked at any part of the display. In addition, the same infants were also presented with a visual search task, composed of two conditions: during the *motion condition*, a single moving bar appeared in a field of identical stationary bars, whereas during the *orientation condition*, a stationary tilted bar (oriented away from the vertical at either 30°, 60°, or 90°; see Figure 2) appeared in a field of stationary vertical bars. Infants were presented with 24 trials within each condition. During the visual search task, a trained observer operated a corneal reflection eye-tracking system, which provided a real-time estimate of the infant's point of gaze. Each visual search trial ended when either (a) the infant detected the target, or (b) 4 seconds had elapsed.

After testing, Amso and Johnson (2006) divided their sample of twenty-two 3-month-old infants into two groups, as a function of each infant's performance on the unity-perception task. In particular, 11 infants—hereafter, the "perceivers"—looked significantly longer at the broken-rod test display, indicating that they had perceived the occluded rod as one single object (i.e., unity perception). The remaining 11 infants—hereafter, the non-perceivers—meanwhile, did not look longer at either the complete or broken-rod test displays, suggesting that they did not perceive the occluded rod as a single object.

Next, the performance of the perceivers and non-perceivers on the visual search task was compared. Figure 3 presents the primary findings from this analysis. First, as the top panel indicates, perceivers and non-perceivers did not differ in their success at detecting moving targets. In the orientation condition, however, perceivers successfully detected a larger proportion of the tilted target bars. Second, the bottom panel of Figure 3 presents the mean latency (i.e., time to detect the target) in the orientation and motion conditions (note that only successful trials are included in this analysis). Interestingly, while perceivers and non-perceivers did not differ in their time to detect moving targets, perceivers were significantly *slower* than non-perceivers at detecting targets in the orientation condition.

To summarize, a moving target appears to be a highly salient stimulus, and consequently may not require controlled visual search (or inhibition of distractor targets) to be detected by 3-month-old infants. In contrast, tilted targets were detected less successfully by 3-month-olds. More importantly, however, as

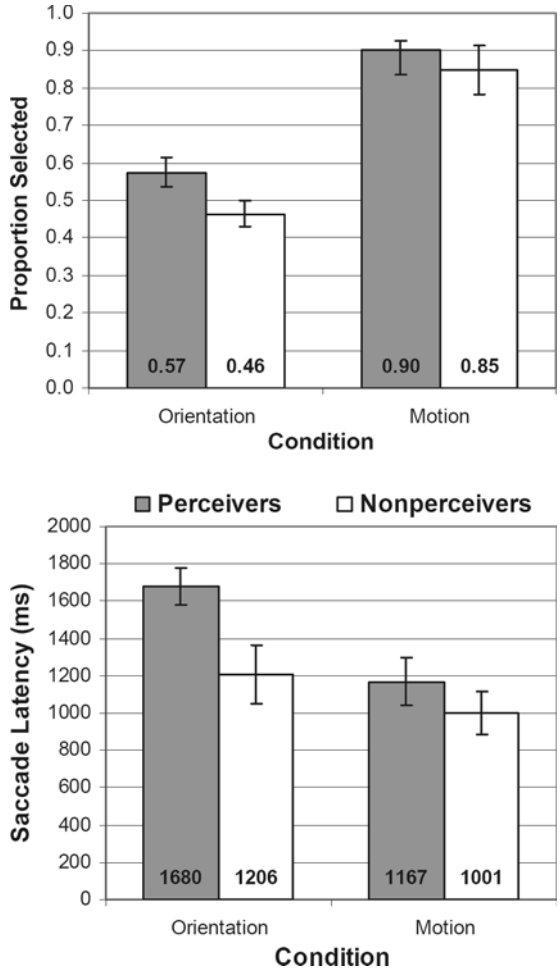

Figure 3 Top panel: Mean detection rates (proportion detected) for the perceivers and non-perceivers as a function of target condition (orientation vs. motion). Bottom panel: mean latency (on successful trials) for the perceivers and non-perceivers as a function of target condition (orientation vs. motion).

Amso and Johnson (2006) predicted, infants who provided evidence of unity perception were also more successful at detecting targets in the orientation condition.

Taken together, the findings from Amso and Johnson (2006) not only demonstrate that unity perception and visual search are related in 3-month-old infants, but more specifically, that infants who have begun to develop unity perception also appear to engage in more exhaustive or deliberate visual search strategies. These data are consistent with the role of visual selective attention as a common feature in the development of both unity perception and visual search. In the next section, we highlight three neural components of visual selective attention, and describe a multi-channel, image-filtering model that simulates the effects of these components on the development of visual search.

3 A Model of Visual Selective Attention

By implicating the role of visual selective attention in the development of unity perception, the findings from Amso and Johnson (2006) raise an important question: What are the developmental mechanisms that drive changes in visual selective attention (i.e., attention deployment)? Are these changes specifically due to the emergence of unity perception? In other words, does emerging knowledge of a predictable world of objects lead to a coherent strategy for gathering visual information? Alternatively, does visual selective attention emerge as the result of a more general, underlying capacity?

To investigate this issue, a multi-channel, image-filtering model of visual processing in the occipital and parietal cortex was used to simulate the visual-search performance data obtained by Amso and Johnson (2006). The model was originally developed by Itti and Koch (2000), and incorporates several basic principles of neural processing in the mammalian visual system, including: (a) decomposition or filtering of the visual stream into multiple parallel feature channels, (b) retinotopic feature maps, (c) center-surround organization in early visual areas, and (d) competition for attention (between multiple salient locations) in a retinotopic "salience map" (e.g., Kastner & Ungerleider, 2000). From a computational perspective, the model is also consistent with theoretical accounts of visual attention that emphasize the extraction and combination of multiple visual features in parallel, such as *feature integration theory* (Treisman & Gelade, 1980).

3.1 Model Overview: Structure and Function

The intuition behind the image-filtering model is that patterned projections of light (i.e., the optic array) that enter the eye experience a number of well-defined transformations that are analogous to a series of optical filtering systems. These neural "filters" include the retina, lateral geniculate nucleus (LGN), and occipital cortex, as well as the extrastriate areas (e.g., parietal and temporal cortices). The model is not a detailed

representation of striate and extrastriate anatomy and physiology, but is instead designed to capture five general stages of visual processing in the mammalian visual system: (a) feature detection or extraction (across four visual channels), (b) center–surround contrast enhancement, (c) within-feature competition, (d) integration of within-feature maps into a unified salience map, and (e) selection of a salient location for fixation.

Input to the model is presented as a three-dimensional array of pixel values (i.e., three two-dimensional arrays that specify the amount of red, green, blue for each pixel in the input image), taken from the same stimuli used by Amso and Johnson (2006). As Figure 4 illustrates, the initial array of pixel values (analogous to a retinotopic map) is transformed across several dimensions (i.e., filter channels) in parallel. At the final stage of filtering, the separate maps are combined into a single, composite retinotopic (saliency) map.

In Section 4, we illustrate how the resulting saliency map is used to simulate infants' visual search.

We provide here a brief description of the sequence of processing stages in the image-filtering model. It is important to note that while this description highlights the process of optical transformation in the model (e.g., feature extraction, contrast enhancement, etc.), the mathematical operations that underlie each of the filtering stages are general enough that they can be computed through a number of different numerical techniques, including both linear filtering methods and artificial neural networks.

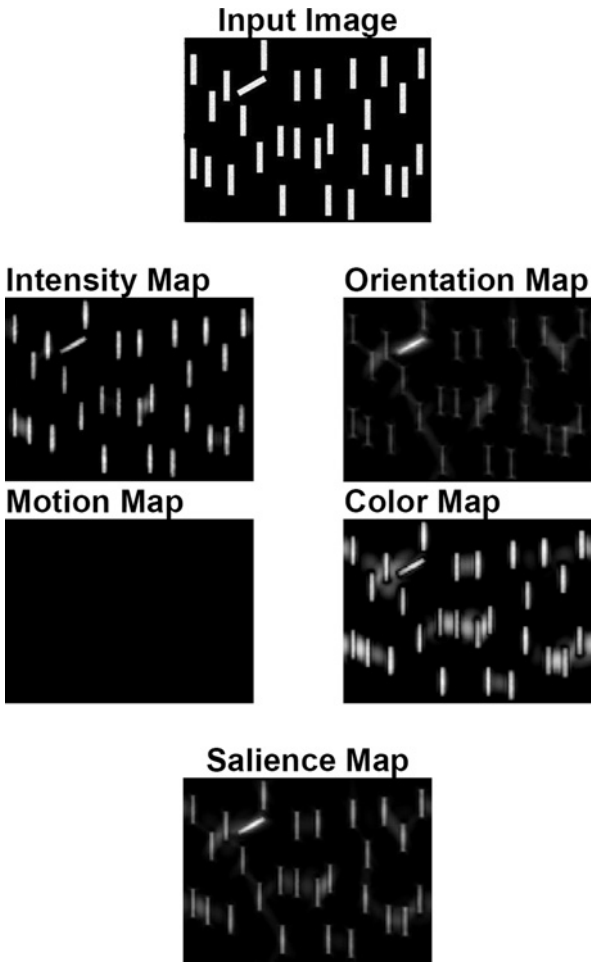

Figure 4 Illustration of two of the filtering stages from the image-filtering model, including the end of the within-feature orientation stage (middle four maps) and the salience map stage (bottom map).

3.1.1 Feature Extraction The first stage of the image-filtering model corresponds to feature extraction in early visual processing (e.g., retinal ganglion cells and LGN). Accordingly, four parallel image filters extract image intensity, oriented edges, motion, and opponent-color contrast from a raw input image (see Figure 4): (1) intensity maps are obtained by converting color input images to grayscale; (2) oriented edges are abstracted by processing the input image with oriented gabor wavelets (see Mermillod, Chauvin, & Guyader, 2004, for a comparable approach); (3) motion is computed as the (absolute) difference in corresponding pixels at each location in the input image, between consecutive images; and (4) opponent-color processing is determined by weighting and summing the red–green and blue–yellow color channels (for a detailed description of how each feature is extracted, see Itti, Koch, & Neibur, 1998). Eight separate feature dimensions are computed, including one intensity, four orientation (i.e., 0°, 45°, 90°, and 135°), one motion, and two color (i.e., red–green and blue–yellow) dimensions. A total of 24 feature maps are computed, as the eight dimensions are extracted across three spatial scales or frequencies (i.e., fine, medium, and coarse).

3.1.2 Contrast Enhancement During the second stage of filtering, each of the feature maps is processed through a filter analogous to center–surround excitation–inhibition. This stage not only replicates a key functional property of early visual processing, but also serves to enhance contrast in the feature map

while reducing background noise present in the input image.

3.1.3 Within-Feature Competition During the third stage, each feature map is processed through an image filter that represents short- and long-range connections in V1; short-range connections are excitatory, whereas long-range connections are inhibitory. As a result of this filtering process, the presence of a feature at a given location (e.g., a vertical edge) is self-stimulating within a local neighborhood, but also inhibitory to the same type of feature (e.g., other vertical edges) at greater distances. In the current model, it is important to note that within-feature competition is implemented as a discrete iterative process that can occur one or more times. As a result, the number of iterations or loops in the within-feature competition process is parameterized in the model, and is intended to correspond to the temporal duration of recurrent activity in posterior parietal cortex.

At the end of this stage, the feature maps are summed across the three spatial scales and within each of the four feature channels. Figure 4 illustrates these four "conspicuity" maps (i.e., intensity, orientation, motion, and color channels, respectively). The net effect of filtering and combining the feature maps at the third stage is that, within each dimension, similar features are inhibited while distinct or spatially isolated features are enhanced. In particular, note in the orientation map that vertical lines are suppressed, while the diagonal line "pops out."

3.1.4 Salience Map No additional filtering occurs during the fourth stage. Rather, the four conspicuity maps are summed into a single retinotopic map, which represents the integration of separate visual channels into a unified salience map. Note that the salience map does not encode the presence of a particular feature, but instead the relative salience of one or more features at each location in the visual field.

3.1.5 Location Selection The final stage of the model converts activity over the salience map into an array of candidate or potential targets for fixation (i.e., locations to which the fixation point is likely to be shifted). Rather than sorting locations by their activation (i.e., salience) level, and then simply assuming that fixations deterministically shift from the highest to the next highest salience-point in the map, we instead employ a stochastic process, in which a probability of fixation is associated with each location in the salience map.

In particular, a *softmax* equation was implemented, including a parameter (*tau*) that modulates the probability of selecting different locations on the salience map (*tau* is analogous to the temperature parameter in the algorithm for simulated annealing). By varying the value of *tau*, we can flexibly shift from an optimal, deterministic pattern of location selection (i.e., values of *tau* near 0) to a predominantly stochastic pattern, in which both highly-salient and less-salient locations have a chance to be selected.

3.2 Neural Constraints on Visual Selective Attention

The image-filtering model was used to examine the hypothesis that visual selective attention is constrained by the development of one or more underlying neural subsystems. In particular, three specific subsystems were identified, parameterized in the model, and independently varied in order to generate corresponding developmental trajectories.

3.2.1 Oculomotor "Noise" Several lines of research suggest that variability is not only an inherent feature of behavior, but that it also provides an important source of experience during early motor development (e.g., Piek, 2002). According to this view, "noisy" or variable behavior during early development is not necessarily due to immature control systems, but instead may be an adaptive strategy for exploring sensorimotor contingencies. There are also several analogous examples of this approach in the machine learning literature, including the concepts of "motor babbling" and the "exploration-exploitation" tradeoff (e.g., Kuperstein, 1988; Sutton & Barto, 1998).

A general theme that emerges from this work is a developmental pattern that begins with relatively high amounts of behavioral variability, followed by a gradual decrease in variability over time. This pattern was simulated in the image-filtering model by systematically tuning the value of *tau*, which modulates the probability of fixating a salient location (i.e., oculomotor noise).

In particular, at low levels of *tau*, there is a high probability of fixating a salient location (i.e., "greedy" action selection), while at high levels of *tau* this probability is decreased (i.e., an increase in sub-optimal or exploratory actions).

3.2.2 Horizontal Connections in V1

A second potential constraint on the development of visual selective attention is the growth of lateral or horizontal connections in V1 (e.g., Albright & Stoner, 2002; Hess & Field, 1999). Although the precise function of these connections is currently the subject of debate, they appear to play an important role in the perception of contours, and in particular, the integration or perceptual "fill-in" of contours that are interrupted by spatial gaps (e.g., due to occlusion; Albright & Stoner, 2002).

A related computational function was examined in the present model by varying the size of the filter used to compute within-feature competition. In particular, the size of the filter (i.e., the percentage of the input image that the filter spans) enables the distance of "long-range" interactions (i.e., inhibition and excitation) in the retinotopic feature maps to be directly modulated. Analogous to the growth of horizontal connections in V1 during infancy, an increase in this distance corresponds to a larger range of alternative salient locations that simultaneously compete for attention. In effect, as the distance parameter increases, the likelihood diminishes that a spurious, less-salient location will succeed in attracting the model's attention.

3.2.3 Recurrent Processing in Parietal Cortex

A third neural constraint focuses on the role of recurrent or sustained activity within the posterior parietal cortex. Neural activity in this region has direct implications for visual attention, insofar as parietal retinotopic maps (e.g., in the intraparietal sulcus) appear to encode for the salience of visual stimuli (e.g., Gottlieb, Kusunoki, & Goldberg, 1998; Shafritz, Gore, & Marois, 2002).

Itti and Koch (2000) propose a recurrent model of parietal activation, in which recurrent feedback (in the parietal "salience map") is associated with sustained competition between salient locations. This was achieved in the model by varying the number of discrete iterations or loops that are completed during the within-feature competition stage. In computational terms, extending the duration of this feedback increases the likelihood that the salience map will "settle" into a stable activation pattern. Similarly, in functional terms, varying the duration of recurrent feedback is analogous to modulating the time spent covertly "comparing" two or more putative targets. As we highlight in the final section, there are a number of neural mechanisms that have been proposed for modulating recurrent parietal activity, including signals originating within the parietal and frontal cortex (e.g., Kastner & Ungerleider, 2000).

4 Simulation Results

Two sets of simulations were conducted. In the first set, we simulated the development of visual search by systematically varying each of three parameters in the model, corresponding to oculomotor noise, horizontal connections in V1, and recurrent parietal processing, respectively. A key finding from this first set of simulations was that modulation of recurrent parietal processing results in a developmental profile that corresponds to the performance of perceivers and non-perceivers observed by Amso and Johnson (2006). Accordingly, we pursued this finding in a second set of simulations by using the model to generate real-time visual search patterns.

4.1 Simulating the Development of Visual Search

Development of visual search was simulated by presenting the image-filtering model with the same stimuli used by Amso and Johnson (2006). In particular, recall that infants were tested in two conditions (i.e., 24 motion trials and 24 orientation trials). The motion condition was simulated by sampling the first two frames from each of the 24 motion stimuli used by Amso and Johnson (2006). In particular, because the target bar moved at a fixed velocity during each motion trial, only the first two frames were needed to activate the motion map (i.e., with the use of frame-differencing to detect motion). Similarly, the orientation condition was simulated by sampling the first frame from each of the 24 orientation trials (i.e., only one frame was needed, since there was no motion in the orientation condition). Input images were downscaled to 133 by 100 pixels.

Target detection rates were estimated by passing the appropriate image frame (or frames) through the image-filtering model, and then assigning a probability of fixation to the 100 most active locations on the corresponding salience map. Fixations that contacted the target were pooled, and the corresponding probabilities for these fixations were summed, resulting in a cumulative (estimated) probability of fixating the target for each particular stimulus.

Preliminary simulations of the visual search task were used to identify optimal values for the three developmental parameters (i.e., oculomotor noise, horizontal connections, and recurrent parietal processing). In particular, the mean probabilities of fixating the target were 99.42% and 90.75% during the motion and orientation conditions, respectively, with the following parameter values: *tau* = 0.5, range of horizontal connections = 99 pixels, and number of recurrent parietal loops = 10. Development of visual search was then simulated by varying each of these parameters systematically (while holding the other two parameters fixed at their optimal values), and computing the mean target detection rates generated by the model during the motion and orientation conditions.

As noted earlier, Amso and Johnson (2006) reported that while both perceivers and non-perceivers were near-optimal at detecting the moving target in the motion condition, perceivers succeeded at detecting the tilted target in the orientation condition roughly 55% of the time, whereas non-perceivers succeeded roughly 45% of the time (see Figure 3). Recall that perceivers also had significantly longer search times. As a consequence, the simulation analyses were guided by a generalized "goodness of fit" strategy, in which the developmental trajectory produced while modulating each of the three model parameters was compared with the performance profile generated by 3-month-old infants. In particular, an acceptable degree of fit was defined a priori by a shift in the model from 45% to 55% target detection rate in the orientation condition, accompanied by near-optimal detection rates in the motion condition for the corresponding parameter values.

4.1.1 Developmental Decline of Oculomotor Noise

The developmental decline of oculomotor noise was simulated by testing the image-filtering model under a regime in which *tau* was varied from 30 to 0.5. As

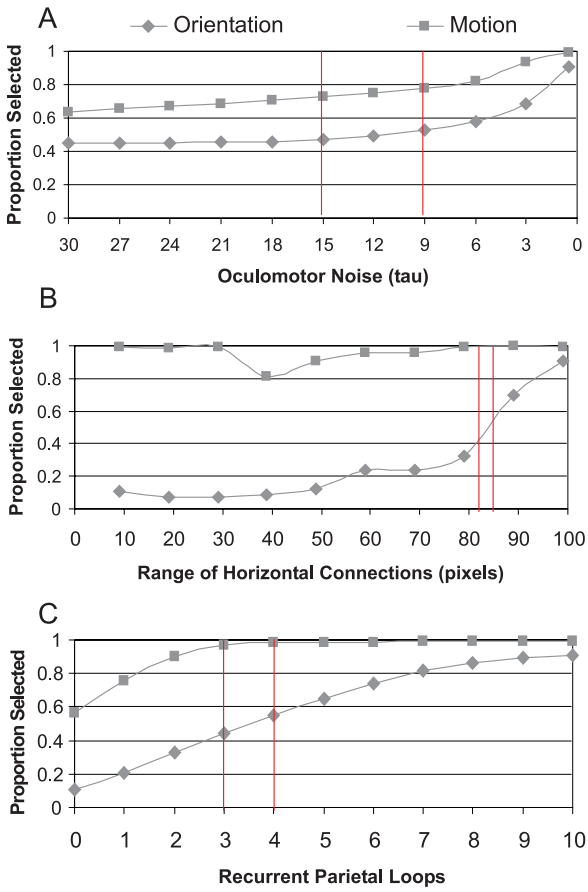

Figure 5 Developmental trajectories of the model on the visual search task as three parameters were varied: (A) oculomotor noise; (B) range of horizontal connections; and (C) number of recurrent parietal loops. Vertical lines denote points in the model's performance that correspond to non-perceivers and perceivers (see text for details).

Figure 5A illustrates, detection rates during both the motion and orientation conditions increased as *tau* fell (note that the direction of the *x*-axis is reversed).

Two relevant findings emerged as *tau* was modulated. First, as for 3-month-old infants, overall the model was more successful at detecting the target during the motion condition ($t(10) = 14.50$, $p < 0.001$); the mean proportion of detected targets was 0.76 and 0.54 for the motion and orientation conditions (SD = 0.11 and 0.14, respectively).

Second and more importantly, the developmental trajectory that was generated as *tau* declined is not consistent with the search performance of 3-month-old infants. In particular, the pair of vertical lines in Figure 5A highlight the values of *tau* where the model

transitions from 45% to 55% detection in the orientation condition (i.e., approximately between 15 and 9). Thus, at values of *tau* where the model's performance on the orientation condition captures the difference between 3-month-old non-perceivers and perceivers, its performance in the motion condition lags well behind that of both groups of 3-month-old infants.

4.1.2 Development of Horizontal Connections

Next, the development of horizontal connections in V1 was simulated by testing the image-filtering model as the range of these connections was increased from 9 to 99 pixels. As before, the model was more successful overall at detecting the target during the motion condition ($t(9) = 7.83$, $p < 0.001$); the mean proportion of detected targets was 0.96 and 0.29 for the motion and orientation conditions ($SD = 0.06$ and 0.28, respectively).

The pair of vertical lines in Figure 5B highlight the transition in the model from 45% to 55% detection in the orientation condition, as horizontal connections increased in size (i.e., from approximately 82 to 85 pixels). Between these parameter values, the model is near 100% accurate at detecting the target in the motion condition. Thus, although the developmental shift in the model that corresponds from non-perceivers to perceivers is comparatively rapid, the qualitative pattern of this trajectory is comparable to the visual search performance patterns observed in 3-month-old perceivers and non-perceivers.

4.1.3 Development of Recurrent Parietal Processing

Finally, the development of recurrent parietal processing was simulated by testing the model while the number of recurrent parietal filter iterations or loops increased from 0 to 10. Again, the model detected the target more often overall during the motion condition ($t(10) = 5.80$, $p < 0.001$); the mean proportion of detected targets was 0.92 and 0.59 for the motion and orientation conditions ($SD = 0.14$ and 0.29, respectively).

Figure 5C illustrates the improvement in visual search performance, during the motion and orientation conditions, as the number of recurrent parietal loops increases. Interestingly, as the pair of vertical lines indicate, the transition from 3 to 4 loops corresponds to the transition from 45% to 55% detection of the target in the orientation condition (i.e., analogous to the difference between non-perceivers and perceivers). During this increase in the number of parietal loops, meanwhile, the model did not improve significantly in the motion condition ($t(23) = 1.61$, $p = 0.12$). In particular, the mean proportion of detected targets was near-ceiling at 0.97 and 0.99 for three and four loops, respectively ($SD = 0.14$ and 0.29).

4.2 Simulating Real-Time Visual Search Performance

The findings from the first set of simulations provide two potential accounts for the difference in visual search performance between perceivers and non-perceivers: *either* (1) the growth of horizontal connections in V1, *and/or* (2) an increase in recurrent parietal processing. However, it is important to note that whereas changes in horizontal connections occur on the spatial dimension, increases in the number of recurrent loops or iterations occur on the temporal dimension. In other words, *more loops take more time*. This is a key distinction, because an increase in the number of loops can be interpreted as a developmental prediction that as infants become *more successful* at detecting stationary, tilted targets in a field of vertical distracters, they should also *take longer* to detect those targets. Indeed, this is exactly the performance difference that distinguishes non-perceivers from perceivers (Amso & Johnson, 2006).

Therefore, we chose to focus our next analysis on the role of recurrent parietal processing as a constraint on the development of visual selective attention, and of visual search in particular. Unfortunately, the findings from the previous simulations are limited in part by the fact that search performance in the model was defined as an aggregate, probabilistic estimate based on a set of potential—but not actually produced—eye movements. In particular, note that while infants were given a maximum of 4 seconds per trial to detect the target in the motion and orientation conditions, the estimated detection rates generated by the model are produced under the assumption of unlimited search time, or more precisely, an unrealistically long series of saccades or gaze shifts (e.g., 100 saccades).

In order to address this issue, we conducted a second set of simulations, in which three additional constraints were incorporated into the model that enable it to produce a sequence of eye movements in simulated

real-time. Before describing these constraints, it is important to stress that our goal in simulating real-time performance was to illustrate how modulating the amount of recurrent parietal processing leads to a corresponding shift in the performance of the model from the non-perceiver profile to the perceiver profile. As we highlight below, the choice of these constraints was guided not only by existing empirical data, but also by heuristics from motor control research.

In order to translate the output of the model (i.e., the salience map) into a series of eye-movements, three specific constraints were introduced: (1) saccade frequency, (2) inhibition of return, and (3) a salience–distance tradeoff.

4.2.1 Saccade Frequency While Amso and Johnson (2006) did not directly measure gaze-shift or saccade frequency during the visual search task, they do have a direct measure from the same infants during the unity-perception task. In particular, infants shifted their gaze approximately once every 220 ms (with a standard deviation of 20 ms). Accordingly, the duration between successive fixations in the model was simulated by assuming that the build-up of a motor signal prior to each eye movement follows a normal distribution (i.e., $\mu = 220$ ms, $\sigma = 20$ ms). In addition, it was also assumed that each iteration of the recurrent parietal processing loop follows the same distribution.

4.2.2 Inhibition of Return A second constraint added to the model was an inhibition-of-return mechanism. Specifically, after each fixation, the next gaze shift was subject to a minimum-distance constraint. In other words, subsequent fixations were required to be at least 24 pixels apart (i.e., 10% or more of the input image width). This constraint is analogous to the inhibition-of-return phenomenon observed in infants, children, and adults, in which attention to a particular location is followed by a tendency to avoid returning to that location (e.g., Hood, 1995; Johnson, 1994). In effect, it ensures that the model does not "lock" onto one salient location and then only generate small saccades to neighboring locations.

4.2.3 Salience–Distance Tradeoff The third constraint was motivated by the reasoning that if the model simply followed the gradient of salience over the input image (i.e., subsequent fixations are to successively less-salient locations), occasional saccades might require traversing the entire distance of the input image (e.g., require both head and eye movements). While it is not clear whether infants' saccade patterns are subject to an energetic or metabolic constraint, it seems both unlikely and inefficient that gaze shifts are determined by salience alone. As a result, a normalization function was implemented, which weighted the salience of each potential saccade target by the distance required to shift the fixation point to that location (i.e., salience divided by saccade distance). In effect, this normalization created a tradeoff between (a) the stimulus salience for each location on the input image versus (b) the "work" required to fixate that location.

Subject to these three constraints, the model was then presented with the same two sets of input images that were used in the previous simulations (i.e., 24 motion trials and 24 orientation trials). The performance of non-perceivers was simulated by allowing the parietal loop to iterate a maximum of three times during each trial, while perceivers were simulated by allowing the loop to iterate a maximum of four times per trial. As the salience map was updated at the end of each parietal processing loop, both simulated perceivers and non-perceivers generated a saccade each time the salience map was updated. Note that while the salience map was eligible to be updated, the delay between saccades was a function of *both* the build-up of the motor signal *and* the time allotted to the parietal loop (i.e., roughly 220 ms for each process); once the parietal loop reached its maximum number of iterations (either three or four, respectively), the salience map was no longer updated, and the delay between subsequent saccades was due to build-up of the motor signal alone.

At the start of each trial, the model's simulated fixation point was positioned at a random location on the input image. The model then produced a series of saccades, following the constraints described above. As with infants, each trial concluded when either (a) the model succeeded in fixating the target (i.e., generating a saccade that made contact with the target bar), or (b) the model failed to fixate the target within 4 s (i.e., 4,000 ms). Similarly, 10 simulated perceivers and 10 simulated non-perceivers were tested by initializing and testing the model ten times in each condition.

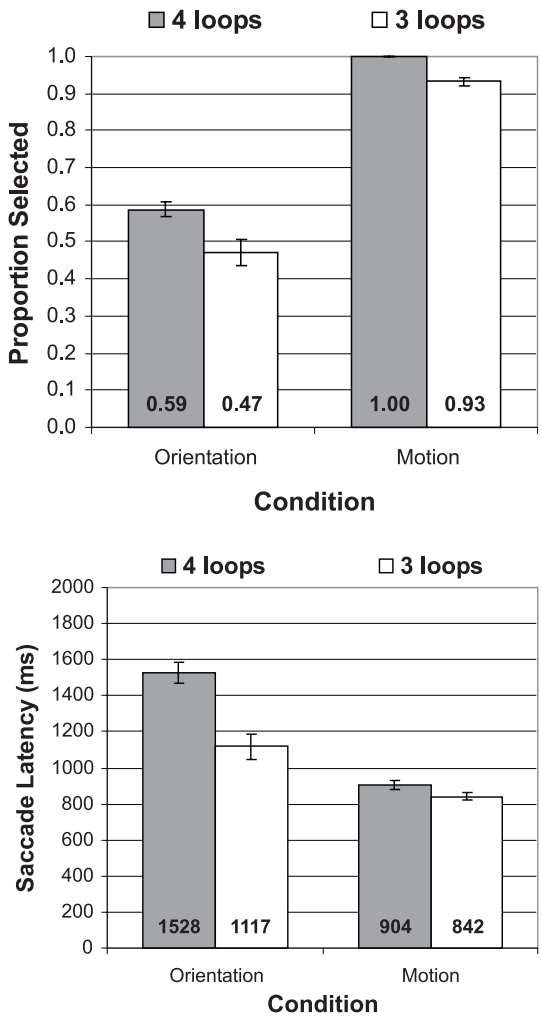

Figure 6 Top panel: Mean simulated detection rates (proportion detected) generated by the model with a maximum of either three or four parietal loops as a function of target condition (orientation vs. motion). Bottom panel: mean simulated latency (on successful trials) with a maximum of either three or four parietal loops as a function of target condition (orientation vs. motion).

The results from the simulated real-time performance of the model are summarized in Figure 6. First, the upper panel presents the proportion of trials in which the model was successful in detecting the target. Like infants, the model not only detected more targets during the motion condition than during the orientation condition, but more importantly, increasing the number of parietal processing loops from three to four resulted in an increase in the proportion of targets detected from 0.47 to 0.59 in the orientation condition (SD = 0.11 and 0.07, respectively; $t(18) = 2.88$, $p < 0.01$). While there was also a significant difference in the proportion detected during the motion condition (unlike infants), this was likely due in part to the fact that the model detected *all* moving targets (i.e., variance was 0) in the four-loop condition ($t(18) = 6.02$, $p < 0.001$). On a related note, the model also appears to be slightly more successful than infants in the motion condition. Notably, however, the overall pattern of performance generated by the model closely mirrors that produced by infants (see Figure 3).

Second, the bottom panel in Figure 6 presents mean detection latency for the model, as a function of target condition (recall that only successful trials are included in this analysis). Again, the overall pattern of performance generated by the model is remarkably close to the pattern seen in infants. As the model succeeded in detecting moving targets relatively quickly (i.e., typically between the second and third iteration of the parietal loop), the difference in mean latency when the number of parietal loops was increased from three to four did not reach significance ($t(18) = 1.76$, $p = 0.09$). During the orientation condition, however, there was a significant difference in mean latency between three and four loops ($t(18) = 4.39$, $p < 0.001$). In particular, the model was significantly faster at detecting the tilted bar when parietal processing was limited to three loops versus four (M = 1117 and 1528 ms, respectively; SD = 224 and 192).

5 Discussion

The aim of the present simulation study was to identify and investigate the influence of three neural constraints on the development of visual search. Accordingly, a multi-channel, image-filtering model was used to simulate the development of visual search in young infants. Oculomotor noise, growth of horizontal connections in visual cortex, and recurrent parietal processing were each varied independently, in order to generate three sets of developmental trajectories. Although changes in oculomotor noise did not produce a developmental pattern that corresponded to observed performance differences in 3-month-old infants, changes in both horizontal connections and recurrent parietal processing did.

These findings highlight two potential neural mechanisms that may account for developmental changes

in visual search, and more generally, the development of visual selective attention. On the one hand, growth of connections between neurons in V1 may promote the capacity to pre-attentively or subconsciously "compare" multiple, salient stimuli. Alternatively, an increase in the duration of recurrent activation in the parietal cortex may allow salient stimuli more time to compete for attention.

A second set of analyses was also conducted, focusing on the latter mechanism. In particular, three additional constraints were incorporated into the model, which allowed real-time visual search performance to be simulated while recurrent parietal activity was systematically modulated. Two key findings emerged from this second set of simulations. First, as predicted by the first set of simulations, increasing the amount of recurrent parietal processing resulted in a more successful visual search in the orientation condition. More specifically, the model's detection rates closely matched the performance of perceivers and non-perceivers. Second, the same increase in recurrent parietal processing also accounted for the difference in search time exhibited by perceivers and non-perceivers. In other words, like human infants, greater parietal processing resulted in more accurate visual search, as well as longer search times.

Taken together, these findings highlight the role of the posterior parietal cortex in the development of visual selective attention. In addition, they may also implicate a specific neural mechanism to account for developmental changes in oculomotor skill. In particular, the image-filtering model not only provides a coherent computational framework for illustrating how retinotopically-organized visual maps are successively formed and transformed, but it also suggests a processing pathway that integrates sensory input and motor action through the concept of a salience map.

The current findings also raise three important questions. First, why does increasing the amount of recurrent parietal activity result in slower, but also more accurate visual search? A tentative answer to this question can be gleaned by observing changes in the salience map over successive iterations of the recurrent parietal processing loop: specifically, the within-feature competition algorithm is designed to inhibit or suppress similarly-activated regions on the salience map (i.e., homogeneous regions), while enhancing activity in portions of the salience map where unique features are located (i.e., non-homogeneous regions). In topological terms, recurrent parietal processing serves to isolate one or a few peaks on the salience map, while lowering the activity of shallow, uniform areas.

By this account, it is possible that a shallow map may result in larger or more spatially-distributed movements of the fixation point over the visual field, but with no corresponding increase in the probability of detecting a particular target. In other words, a shallow salience map is likely to provoke relatively indiscriminate overt scanning (e.g., a type of "shotgun" strategy). In contrast, if we assume that a given target has not yet been detected, and the salience map continues to evolve in real-time, the emergence of one or a few activation peaks on the map may serve to orient or direct subsequent fixations to those peaks, resulting in more focused scanning. The cost for this success, though, is longer search time as the peaks on the salience map gradually take shape.

A second question concerns the specific neural mechanisms that are responsible for modulating recurrent parietal processing. The current results demonstrate that "manually" varying recurrent feedback—akin to a hardwired or maturational constraint—impacts directly on visual search. However, rather than assuming this change is due purely to maturational factors, the modulatory role may instead be accomplished by the internal dynamics of activity within the parietal cortex, or alternatively, from feedback connections from developing areas in the prefrontal cortex (e.g., Canfield & Kirkham, 2001; Kastner & Ungerleider, 2000; Spratling & Johnson, 2004). As a result, two goals of future modeling work are (1) to identify and compare neurally-plausible optimization techniques that are available to simulate adaptive changes in recurrent feedback (e.g., reinforcement learning), and (2) to use these techniques to examine both environmental and neurobiological mechanisms that may drive the development of visual selective attention.

Third, and perhaps most importantly, the current findings provide a suggestive link between unity perception and visual search. However, recall that we began by hypothesizing that visual selective attention is the fundamental skill or capacity that distinguishes perceivers from non-perceivers on both the unity perception and visual search tasks. In order to provide more substantial support for this hypothesis, a long-term goal is to continue refining our approach, so that the same core model (i.e., architecture, learning algo-

rithm, free parameters, etc.) will be able to simulate infants' performance on both tasks.

Finally, we return to a central theme raised at the beginning of the paper, that is, the role of active vision in the development of object perception. In particular, it may appear that the model's behavior is largely reactive or passive (i.e., stimulus-driven), insofar as attentional shifts are produced in response to the salience of external features. However, it is important to note that the salience map is not simply the product of external input, but it is also constrained or influenced by three endogenous processes: (1) feature extraction, (2) feature competition, and (3) the selection of potential targets for fixation. While each of these processes has one or more parameters that were held fixed in the current model, it is a reasonable next step to allow each parameter to vary adaptively, as a function of either the model's momentary state (e.g., habituation level) or its long-term experience (e.g., prior encounters with other objects). As a consequence, the resulting model more transparently illustrates the real-time interaction between stimulus-specific and endogenous or organism-specific factors that we believe occur during active vision.

To conclude, recent work provides evidence that visual search and perceptual completion rely on a common, underlying perceptual-processing system (Amso & Johnson, 2006). The concept of visual selective attention was proposed in the current study as a capacity that may underlie performance on both of these tasks, and provide a fundamental constraint on the development of object perception and cognition. This approach is consistent with the theory of active vision, and in particular, with the idea that human infants construct a progressively more complex world of objects as their visual-motor skill improves. Future work will extend the current simulation findings, to investigate whether the neural constraints examined here also influence the development of perceptual completion.

References

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience*, *25*, 339–379.

Amso, D., & Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *42*, 1236–1245.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, *48*, 57–86.

Canfield, R. L., & Kirkham, N. Z. (2001). Infant cortical development and the prospective control of saccadic eye movements. *Infancy*, *2*, 197–211.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognitive development. *Cognitive Development*, *17*, 1323–1343.

Gilmore, R. O., & Thomas, H. (2002). Examining individual differences in infants' habituation patterns using objective quantitative techniques. *Infant Behavior & Development*, *25*, 399–412.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481–484.

Haith, M. M. (1980). *Rules that babies look by*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Hess, R., & Field, D. (1999). Integration of contours: New insights. *Trends in Cognitive Sciences*, *3*, 480–486.

Hood, B. M. (1995). Shifts of visual attention in the human infant: A neuroscientific approach. In C. Rover-Collier, & L. Lipsett (Eds.), *Advances in infancy research* (pp. 163–216). New Jersey: Ablex.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual-attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience*, *2*, 81–95.

Johnson, M. H. (1994). Visual attention and the control of eye movements in early infancy. In C. Umilta, & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 291–310). Cambridge, MA: MIT Press.

Johnson, S. P. (2004). Development of perceptual completion in infancy. *Psychological Science*, *15*, 769–775.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*, 185–201.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.

Kuperstein, M. (1988). Neural model of adaptive hand-eye coordination for single postures. *Science*, *239*, 1308–1311.

Mermillod, M., Chauvin, A., & Guyader, N. (2004). Efficiency of orientation channels in the striate cortex for distributed categorization process. *Brain and Cognition*, *55*, 352–354.

Piaget, J. (1955). *The child's conception of reality*. London: Routledge and Kegan Paul.

Piaget, J. (1969). *Mechanisms of perception*. London: Routledge and Kegan Paul.

Piek, J. (2002). The role of variability in early motor development. *Infant Behavior and Development*, *25*, 453–465.

Sandini, G., Gandolfo, F., Grosso E., & Tistarelli, M. (1993). Vision during action. In Y. Aloimonos (Ed.), *Active perception* (pp. 151–190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences, USA*, *99*, 10917–10922.

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*, 293–298.

Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience*, *16*, 219–237.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

About the Authors

Matthew Schlesinger is an Associate Professor of Psychology at Southern Illinois University. He received his Ph.D. in developmental psychology from the University of California at Berkeley in 1995, and subsequently trained as a postdoctoral researcher in computer science and robotics at the Italian National Research Council in Rome, and the University of Massachusetts at Amherst. His research employs a range of multi-disciplinary techniques including behavioral methods, computational modeling, and neural imaging, and focuses on the development of object perception and representation in both natural and artificial systems.

Dima Amso is Assistant Professor of Psychology in Psychiatry, Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University. She received her Ph.D. from New York University in 2005. She studies selection mechanisms that support perception and learning and pursues two lines of research to address these issues. The first is a set of eye tracking investigations into the mechanisms underlying selective attention in infancy and the role of the development of this mechanism on information gathering. The second involves eye tracking and fMRI investigations of the genetic, cognitive, and neural processes underlying frequency- and association-based learning across development.

Scott P. Johnson is an Associate Professor of Psychology and Neural Science at New York University. He received a Ph.D. in developmental psychology in 1992 at Arizona State University and postdoctoral training at the Center for Visual Science at the University of Rochester. His principal research interests center on cognitive and perceptual development, with an emphasis on infant development, visual perception, speech perception, eye movements, face perception, object knowledge, learning mechanisms, and cortical development.