Developmental Science

Developmental Science 15:6 (2012), pp 739–752 DOI: 10.1111/j.1467-7687.2012.01177.x

PAPER

Simulating the role of visual selective attention during the development of perceptual completion

Matthew Schlesinger,[1] Dima Amso[2] and Scott P. Johnson[3]

1. *Department of Psychology, Southern Illinois University Carbondale, USA*
2. *Department of Cognitive, Linguistic & Psychological Sciences, Brown University, USA*
3. *Department of Psychology, University of California, Los Angeles, USA*

Abstract

We recently proposed a multi-channel, image-filtering model for simulating the development of visual selective attention in young infants (Schlesinger, Amso & Johnson, 2007). The model not only captures the performance of 3-month-olds on a visual search task, but also implicates two cortical regions that may play a role in the development of visual selective attention. In the current simulation study, we used the same model to simulate 3-month-olds' performance on a second measure, the perceptual unity task. Two parameters in the model – corresponding to areas in the occipital and parietal cortices – were systematically varied while the gaze patterns produced by the model were recorded and subsequently analyzed. Three key findings emerged from the simulation study. First, the model successfully replicated the performance of 3-month-olds on the unity perception task. Second, the model also helps to explain the improved performance of 2-month-olds when the size of the occluder in the unity perception task is reduced. Third, in contrast to our previous simulation results, variation in only one of the two cortical regions simulated (i.e. recurrent activity in posterior parietal cortex) resulted in a performance pattern that matched 3-month-olds. These findings provide additional support for our hypothesis that the development of perceptual completion in early infancy is promoted by progressive improvements in visual selective attention and oculomotor skill.

Introduction

A fundamental step in the development of object perception is the emergence of *perceptual completion*, the ability to integrate multiple surfaces or regions of an object, which are separated due to occlusion, into a unified percept. Experiments that examined perceptual completion in infants at birth have led to conflicting findings, with some studies showing that neonates perceive visible portions of a center-occluded moving rod as disconnected (Slater, Johnson, Brown & Badenoch, 1996; Slater, Morison, Somers, Mattock, Brown & Taylor, 1990), and others showing that neonates perceive the rod parts as connected if they undergo apparent (rather than smooth) motion (Valenza & Bulf, 2011; Valenza, Leo, Gava & Simion, 2006). Studies with older infants have suggested that perceptual completion is relatively well established by age 4 months (e.g. Kellman & Spelke, 1983; Johnson, 2004; Johnson & Aslin, 1995; Slater *et al*., 1996).

A variety of visual cues are often available to support the process of perceptual completion. Figure 1A illustrates a moving rod that is partially occluded by a large screen. Potential cues in this display include: (a) synchronous, lateral motion of the upper and lower rod segments (i.e. 'common fate'), as well as common (b) orientation, (c) color, and (d) alignment of the two segments. One theory of perceptual completion development proposed that learning to detect and exploit such cues underlies age differences in performance observed in experiments with human infants (Mareschal & Johnson, 2002). This possibility was tested with connectionist models that learned to associate the presence of specific cues with object unity as the model was exposed to different perceptual environments. The model was tested for generalization of learning with novel object displays. Results of the models implied a strong role for association learning and perceptual skills (detecting visual cues) in emerging perceptual completion. More recently, we proposed that infants discover cues for perceptual completion as a function of improvements in oculomotor skill, that is, as they develop strategies for attending to and scanning visual stimuli (Johnson, Slemmer & Amso, 2004; Amso & Johnson, 2006). In particular, we are investigating the hypothesis that *visual selective attention* – the ability to focus or deploy attention while ignoring irrelevant stimuli – is critical for infants' perceptual completion.

As we highlight below, there are several lines of evidence that support our hypothesis. For example, at age 3 months, infants' performance on a perceptual

Address for correspondence: Matthew Schlesinger, Psychology Department, Southern Illinois University Carbondale, Carbondale, IL 62903, USA; e-mail: matthews@siu.edu

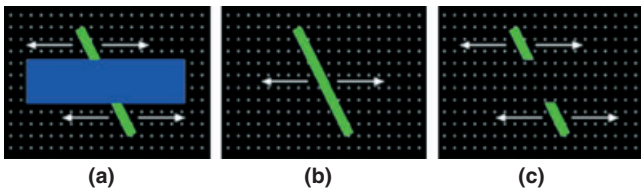

Figure 1 *Displays used to assess perceptual completion in infants: (A) occluded-rod (habituation) display, and (B) complete-rod and (C) broken-rod test displays.*

completion task is correlated with their performance on a visual search task (Amso & Johnson, 2006). This pattern of findings is consistent with the idea that both tasks may be served by the same underlying attentional mechanism. In addition, we have used a *salience-based* computational model to identify and evaluate potential neural circuits that may help explain developmental changes in infants' visual selective attention (Schlesinger *et al.*, 2007). To date, the model has produced two major findings. First, it successfully captures the visual search performance data reported by Amso and Johnson (2006). Second, this performance pattern is achieved through simulated changes in either of two specific brain regions (visual cortex and posterior parietal cortex), which may help account for systematic changes in infants' visual selective attention during early infancy.

An important issue left unaddressed by our modeling work thus far is whether changes in either one or both of these two brain areas can also account for developmental changes in the perceptual completion task studied by Amso and Johnson (2006). Therefore, the goal of the current study is to extend our model toward the simulation of perceptual completion, in order to help demonstrate how changes in the network supporting visual attention may result in more effective perceptual completion.

The rest of the paper is organized as follows. We first briefly review the method and findings from Amso and Johnson's (2006) study of perceptual completion and visual search in 3-month-olds, and then describe how we have used our model to simulate infants' performance during the visual search task. Next, we provide a detailed overview of the model, including a comprehensive description of the major components and processing stages. In this section, we not only highlight important features of the model that make it a valuable tool for studying development, but also identify some of the model's key assumptions. We then describe the process of simulating the unity perception task, and present our simulation findings. In the final section, we conclude by discussing the implications of our simulation findings for the study of perceptual completion in infants, as well as future questions that our model can be used to address.

The role of attention in perceptual completion

During the *unity perception task*, perceptual completion is assessed in young infants by presenting stimuli like those in Figure 1. Infants first view the occluded-rod display (Figure 1A) until they habituate, and then during the posthabituation test phase, they view the broken- and complete-rod displays (Figures 1B and 1C, respectively) on alternating trials. The tendency to look longer at one of the two test displays is assumed to reflect a novelty preference (e.g. Gilmore & Thomas, 2002), and provides a basis for inferring or interpreting how infants perceive the occluded rod. Following this rationale, infants who perceive the occluded rod as a coherent, unified object (i.e. *unity perception*) should experience the complete rod as a familiar display, and therefore show a preference for the broken-rod display. We refer to infants who demonstrate this behavior pattern as *perceivers*. Alternatively, infants who perceive the occluded rod as two disjoint or disconnected surfaces should look longer at the complete-rod display. In this case, we refer to these infants as *nonpercievers*.

By 2 months of age, infants begin to show evidence of unity perception in moving rod-and-box displays when the stimuli are presented in a manner that facilitates detection of the relevant features (e.g. a narrow occluder), and by 4 months, infants provide evidence of unity perception even when the occluder is substantially larger (Johnson, 2004). The time period between ages 2 and 4 months, therefore, appears to represent a rapid transitional phase in the development of perceptual unity. In particular, Johnson *et al.* (2004) predicted that a sample of infants selected near the midpoint of this period would include a mixture of both perceivers and nonperceivers. Johnson *et al.* also predicted that if unity perception is supported by scanning of the relevant regions of the occluded-rod display (e.g. the moving, exposed rod segments), then perceivers should attend to these features more frequently than nonperceivers. This prediction was tested by tracking the eye movements produced by 3-month-olds as they viewed the occluded-rod display. Infants then viewed the posthabituation test displays, and were classified as either perceivers or nonperceivers. As expected, perceivers produced significantly more fixations toward the rod than nonperceivers, and also scanned across the rod's path more frequently. This result was replicated with a sample of 2-month-olds by Johnson, Davidow, Hall-Haro and Frank (2008) using a narrow-occluder stimulus.

These results demonstrate that perceivers are more effective than nonperceivers at deploying their attention toward the relevant features of the occluded-rod display. However, they also raise the question of whether the differences in scanning strategies are specific to the unity perception task, or if instead they are due to a general difference between perceivers and nonperceivers in the ability to deploy attention (i.e. visual selective attention).

In a subsequent study, Amso and Johnson (2006) reasoned that if the different scanning strategies employed by perceivers and nonperceivers were due to systematic differences in visual selective attention, then the two groups of infants should also differ in a

comparable manner on a second task that indexes the same underlying skill. Accordingly, they presented 3-month-olds with the unity perception task in addition to a visual search task. The unity perception task was used to differentiate perceivers from nonperceivers, while the visual search task provided an independent measure of visual selective attention in each group.

Twenty-two infants completed both the unity perception and visual search tasks (task order was counterbalanced). Data were collected during the unity perception task following the procedure employed by Johnson *et al*. (2004), including the use of eye-tracking to record infants' gaze patterns. During the visual search task, infants viewed a display that included a target bar embedded within a field of stationary, vertical bars. Two types of visual search trials were presented: in the *motion condition*, a vertical target bar moved horizontally, while in the *orientation condition*, the target bar was tilted at an angle, but remained stationary. On each trial of the visual search task, infants were credited with detecting the target (i.e. the moving or tilted bar, respectively) if it was fixated within 4 seconds.

Three sets of analyses were performed. First, Amso and Johnson (2006) compared infants' looking times during the test phase of the unity perception task. As expected, 11 of the 22 infants looked significantly longer at the broken-rod display and were therefore categorized as perceivers, while the other 11 infants looked longer at the complete rod and were categorized as nonperceivers. Second, when infants' gaze patterns during the occluded-rod display were analyzed, Amso and Johnson (2006) also found that perceivers directed a significantly higher proportion of their fixations toward the rod segments than nonperceivers ($M = 0.19$ versus 0.13, respectively). Both the looking time and gaze pattern findings replicated the results reported by Johnson *et al*. (2004).

For the third analysis, Amso and Johnson (2006) compared the performance of perceivers and nonperceivers during the visual search task. In particular, they predicted no differences between the two groups in the motion condition – which includes a highly salient target – while higher performance was predicted for the perceivers in the orientation condition – which is more challenging and presumably places an increased demand on visual selective attention. The results were consistent with both predictions. In the motion condition, there were no significant differences between perceivers and nonperceivers: the mean proportion of targets detected was 0.9 and 0.85, respectively. Meanwhile, during the orientation condition, perceivers succeeded in detecting the target more often than nonperceivers: 0.57 and 0.46, respectively.

To summarize, these behavioral results demonstrate that: (a) while some 3-month-olds have acquired the capacity for perceptual completion as measured by the canonical task (i.e. perceivers), others have not yet reached the same milestone, (b) perceivers direct their attention more often than nonperceivers to the relevant features of the occluded-rod display, and most importantly, (c) this ability to deploy attention appears to be a general skill that is also manifested in other measures of visual attention, such as visual search. Taken together, these findings support the hypothesis that visual selective attention underlies performance not only on tasks such as visual search, which require infants to deploy their attention systematically, but also on tasks such as perceptual completion, which require the detection and subsequent integration of multiple visual cues (e.g. common motion, alignment, etc.).

Neural substrates for the development of visual selective attention

The work described thus far is constrained in two important ways. First, because the data provided by Amso and Johnson (2006) are correlational, it may be premature to conclude that selective attention is a *causal influence* on the development of perceptual completion (e.g. other influences or relations cannot be ruled out). Second, even after a causal link between the two capacities is identified in human infants, it may not be possible to study perturbations in the developmental process (e.g. sensory deprivation), or directly modify the neural structures or pathways that support visual selective attention and perceptual completion.

In order to address these issues, we have designed and investigated a computational model that simulates the development of visual perception and oculomotor control in young infants (Schlesinger *et al*., 2007). The model provides an important complement to our behavioral research with infants. The model was not designed to learn; rather, our goals were to use the model to identify neural structures that may serve as a substrate for visual selective attention, and to systematically examine how changes in those structures influence the production of real-time behavior (e.g. eye-movement patterns).

A central element of our modeling approach is the concept of a salience map, a two-dimensional structure that receives input from a number of lower-level feature-detection systems, and then combines these inputs into a retinotopic representation that encodes the relative salience of objects at their respective locations in the visual field (see Figure 2). The salience map is not only inspired by the anatomy and physiology of the mammalian visual system (e.g. Fecteau & Munoz, 2006; Kastner & Ungerleider, 2000; Koch & Ullman, 1985), but also by psychological theories of visual attention and computational models that have helped to elucidate these theories (e.g. Itti & Koch, 2000; Lu & Sperling, 1995; Treisman, 1988; Wolfe, 1994).

The model includes several parameters or components that are designed to be analogs for specific anatomical regions or structures in the mammalian visual system (e.g. occipital cortex, parietal cortex, etc.). Using these

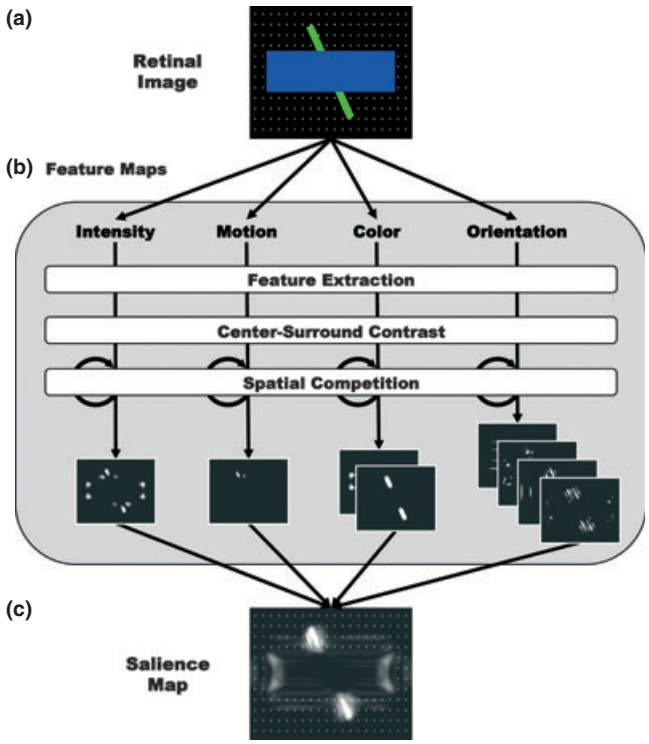

Figure 2 *Schematic diagram of the salience-based model: (A) An input image is projected onto the retina, (B) the retinal image is projected through four feature channels (intensity, motion, color, and orientation), and (C) feature maps produced across the four feature channels are pooled into single, unified salience map.*

structures as a starting point, we then posed three questions:

1. Can the model replicate the *visual search* data reported by Amso and Johnson (2006), including the performance patterns of perceivers and nonperceivers?
2. If so, what neural mechanisms can be used to account for the difference between perceivers and nonperceivers on the visual search task?
3. Can the same model also be used to simulate the performance pattern of perceivers and nonperceivers on Amso and Johnson's (2006) *unity perception* task?

In order to investigate these questions, Schlesinger *et al.* (2007) identified three specific parameters of interest in the model. The first parameter is designed to represent the size of horizontal connections in primary visual cortex (V1). These connections play an important role in the perception of contours that span multiple receptive fields (e.g. Albright & Stoner, 2002; Hess & Field, 1999), and also provide a neural mechanism through which stimuli at different locations in the visual field compete for attention (i.e. *surround inhibition*, see Kastner, De Weerd, Pinsk, Elizondo, Desimone & Ungerleider, 2001). The second parameter is an analog for the duration of recurrent activity in posterior parietal cortex, a region of the brain that is associated with the encoding of visual salience and the modulation of visual attention (e.g. Gottlieb, Kusunoki & Goldberg, 1998; Shafritz, Gore & Marois, 2002). Finally, the third parameter represents the presence of endogenous noise or variability in oculomotor behavior, which plays an important role in early motor skill development (Kuperstein, 1988; Piek, 2002).

We then evaluated the model's performance on the visual search task used by Amso and Johnson (2006). In particular, the model was tested while systematically varying each of the three parameters. For each parameter, the range of values was chosen to simulate development or growth of the corresponding neural substrate during infancy. Three important results emerged from the simulations. First, search performance (i.e. the proportion of targets fixated) increased as each of the three parameters increased. This finding provides support for the idea that increasing each parameter value provides a proxy for growth of the underlying neural substrate. Second, like 3-month-old infants, the model was more successful at detecting the target in the motion condition than in the orientation condition. Third and most importantly, as two of three parameters were increased, the model reproduced the visual search performance of nonperceivers and then perceivers. Tuning of the third parameter (i.e. oculomotor noise), however, did not result in a performance pattern that matched either nonperceivers or perceivers.

These simulation findings provide answers to the first two questions raised above. First, the model successfully captures the visual search performance data reported by Amso and Johnson (2006). It not only simulates the visual search behavior of 3-month-olds, but more specifically the model also matches the level of performance found in perceivers and nonperceivers in two search conditions. Second, the model suggests that growth in two specific cortical areas may support the development of visual search.

To conclude this section, we note that our simulation findings thus far are consistent with the hypothesis that visual selective attention plays an important role in the development of both perceptual completion and visual search. First, our model demonstrates that performance on a visual search task varies as a function of development in two specific cortical areas. And second, while we did not investigate perceptual completion directly, the model also accounts for differences in performance on the visual search task that correlate with performance on the unity perception task (i.e. perceivers vs. nonperceivers).

It remains an open question, however, whether the same model can also capture the performance of perceivers and nonperceivers during the unity perception task. Indeed, this is a particularly strong test of the model. A successful simulation will provide additional support for our hypothesis, by demonstrating that the same cortical areas that influence visual search in 3-month-olds also influence perceptual completion in a comparable manner. Lack of positive findings,

meanwhile, may suggest that visual selective attention is not a direct or causal influence on the development of perceptual completion, but rather, that the correlation between the two measures observed by Amso and Johnson (2006) is mediated by changes in other cortical areas or networks not included within the model.

Simulating the development of unity perception

In this section, we first offer a detailed overview of how the model is designed. Next, we present the method and findings from two simulation studies of the unity perception task.

Model design

The salience-based model is designed to simulate three key aspects of infants' experience during a perceptual experiment. First, the model is presented with the same animation events that infants view in paradigms such as the visual search and unity perception tasks. Second, the internal structure and function of the model roughly corresponds to that of the mammalian visual system: (a) projection of the visual field onto the retina, (b) detection of basic visual features (e.g. edges, motion, etc.), and (c) transformation and pooling of the features into an integrated retinotopic map. Third, the model generates a series of virtual eye movements (i.e. overt shifts of attention) in response to the visual input.

As noted earlier, processing within the model is divided into four stages. We provide here an overview of the functional processes that occur within each stage. The Appendix presents the implementation details. We also note that our model is based on previous work by Itti and colleagues (e.g. Itti & Koch, 2000; Itti, Koch & Niebur, 1998).

Retinal image

During the first processing stage, a still-frame image (extracted from the animation event) is projected onto the simulated retina (see Figure 2A). Note that the model employs a monocular vision system, and that the simulated visual receptors are uniform in size and evenly distributed on the retina (i.e. the retina is not divided into a fovea and periphery). In addition, the number, size, and arrangement of the receptors is assumed to be in 1-to-1 correspondence with the input image (i.e. 480 × 360 pixels).

Feature maps

As Figure 2B illustrates, the retinal image is decomposed into eight feature maps, which are distributed over four feature channels: one intensity map (i.e. luminance), one motion map, two color maps (i.e. blue-yellow and red-green opponent pairs) and four oriented-edge maps (i.e. 0°, 45°, 90°, and 135°). During this stage, the feature maps are created through a three-step process. First, each feature is extracted from the input image at three spatial scales (i.e. fine, medium, and coarse), resulting in a total of 24 feature maps (note that Figure 2B illustrates only the medium spatial scale for each feature type). Second, a center–surround receptive-field contrast filter is then applied to each feature map, which mimics the inhibitory-excitatory organization found in early visual processing (i.e. retinal ganglion cells and LGN). This filter also enhances feature contrast within each of the feature maps.

During the final step of the feature map process, each map passes through a *spatial-competition filter*. This filter has four important properties. First, the size of the filter (or more precisely, the filter kernel, which determines the number of pixels that 'compete' at each location in the feature map) can be varied, ranging from the width of a single pixel, up to the size of the entire input image. Second, the spatial-competition filter can be applied an arbitrary number of times (including 0). Third, as indicated by the circular arrows, it is an iterative (or recurrent) process: this means that the feature map produced at the end of an iteration becomes the input into the process during the next iteration. Finally, the spatial-competition filter uses a combination of excitation and inhibition to reshape the pattern of activity on the feature map. In particular, the filter increases activity at each location on the map in proportion to nearby activity (i.e. local excitation) while it simultaneously decreases activity at the same locations in proportion to activity over the entire map (i.e. global inhibition). The net effect of the spatial-competition filter is that it suppresses activity on maps that have many similar features, while it enhances activity on maps with features that are sparsely distributed.

Salience map

During the third stage of processing, the feature maps are pooled into a unified salience map. While the salience map resembles the retinal image, activity on this map does not represent the presence of specific visual objects, but rather, it indicates the location and relative strength of the particular features detected during the second stage. Thus, multiple features at the same location that differ from their local neighborhood (e.g. an oriented edge that is moving) produce higher activity on the salience map than a homogeneous region that has low contrast and only activates one feature channel (e.g. the center of the blue occluding screen).

Target selection

A stochastic selection procedure is used to select a target on the salience map, which enables the model to shift its virtual fixation point from one location on the map to another. First, the 100 most active locations on the salience map are identified. Second a fixation probability is assigned to each of these locations, proportional to the

activation level at the corresponding location. This weighting strategy biases selection toward highly salient locations, while also allowing other, less-salient locations to occasionally be fixated. Finally, a location is randomly selected as a target for the next fixation from this weighted distribution.

The description of the model provided thus far illustrates how a single image is processed. This process can be generalized in a straightforward manner – with one minor modification – to the case of a series of input images that are sampled from an animation sequence. In particular, note that the animated events used by Amso and Johnson (2006) were presented at 30 frames per second. While the model could, in principle, produce a gaze shift in response to each frame, this would result in a saccade frequency (i.e. 30 per second) that would far exceed the rate produced by infants. Indeed, Amso and Johnson (2006) observed a rate of 4.57 saccades per second (i.e. approximately once per 220 ms) as infants viewed the occluded-rod display. Accordingly, as we highlight below, this issue is addressed by allowing activity on the salience map to aggregate between gaze shifts, resulting in a saccade frequency that is adjusted to match the rate produced by infants.

Testing the model

The rationale that guided testing of the model was informed by two key issues. First, since the model's virtual fixation point is limited to the display, it is not possible for the model to look 'off-display'. In other words, there is no behavioral analog in the model for measuring habituation, and more importantly, no direct method for comparing the model's looking time during the post-habituation test displays (see Franz & Triesch, 2010). However, recall that Amso and Johnson (2006) found that rod scans provided a behavioral metric that differentiated perceivers from nonperceivers. Thus, we focused on the proportion of rod scans produced by the model as the primary performance measure during testing.

Second, while simulating the development of visual search, we identified two parameters in the model that, when varied, captured the pattern of performance on the visual search task produced by nonperceivers and perceivers (see Figure 3; Schlesinger *et al.*, 2007): the size of horizontal connections in V1, and the duration of recurrent processing in posterior parietal cortex. One limitation of this finding, however, is that each parameter was varied independently (i.e. each was varied while the other was held at a fixed value). This testing procedure helps to isolate the effect of each parameter, but precludes the ability to examine the behavior of the model for interactions between parameters. In the current simulation study, we therefore decided to systematically test the model while varying the two parameters in tandem (i.e. exhaustively sweeping through the range of combinations of parameter values).

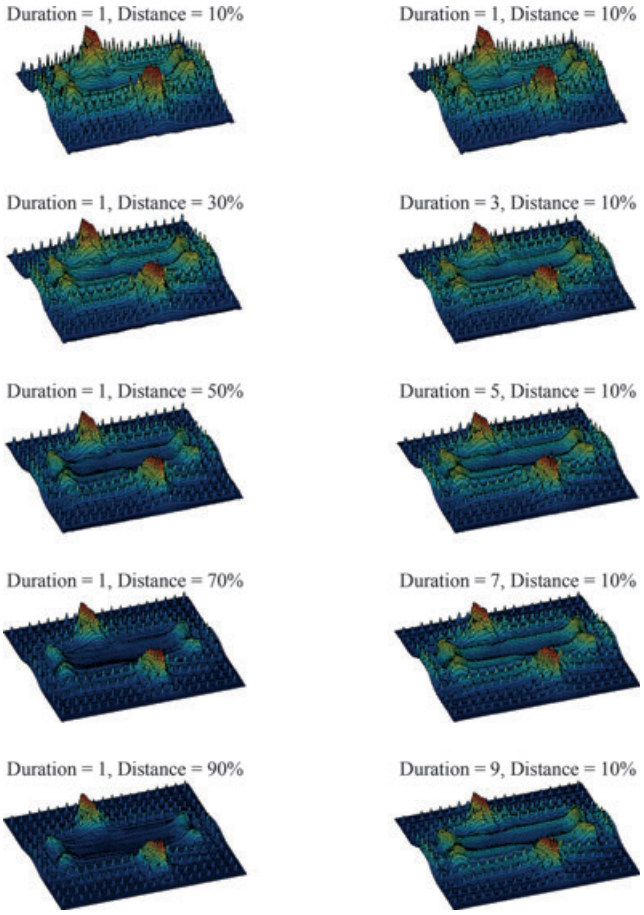

Figure 3 *Illustration of activity on the salience map as the two parameters of interest are systematically varied. In the left panel, the duration of spatial competition is held fixed at 1 iteration while the distance of the spatial-competition filter is increased. In the right panel, the distance of the spatial-competition filters is fixed at 10% of the input image size while the duration of spatial competition is increased.*

Model parameters of interest

Before explaining how the model was tested, it is important to describe how each of the neural mechanisms being investigated was translated into a parameter that can be directly manipulated within the model. First, the parameter that corresponds to size or distance of horizontal connections in V1 is implemented in the model as a value that determines the size of the spatial-competition filter (see Figure 2B). Recall that this filter is a square region that can range in size from one pixel to 100% of the input image height (i.e. 320 × 320 pixels), and that it distributes a mixture of excitation (locally) and inhibition (globally). Increasing the size of the filter extends the reach of the inhibitory component, which increases the amount of 'competition' between active locations on each filter map (i.e. features within the same feature channel). The left side of Figure 3 illustrates how this parameter influences the resulting salience map: each image represents the state of the model's salience map, across a range of possible filter sizes (i.e. 10%, 30%, 50%,

70%, and 90% of the input image), after presenting an image selected from the occluded-rod display (i.e. frame 75 of 150). (Note that each salience map is presented here as a 3D surface, where the height of the surface corresponds to the amount of salience.)

The right side of Figure 3 illustrates how an increase in the second parameter – the number of loops or iterations in the spatial-competition process – affects the resulting salience map. As Figure 2B indicates, this process occurs near the end of the second processing stage, and can be applied an arbitrary number of times. This parameter has both a computational and a physiological interpretation. At the computational level, because the spatial-competition process is recurrent, activity on the feature maps gradually reaches a stable state after several iterations (typically 10 or fewer). Thus, increasing the duration of spatial competition tends to drive activity on the map toward a stable overall pattern. Similarly, at the physiological level, changes in the duration of recurrent processing are functionally equivalent to raising or lowering a neural threshold for firing (e.g. Lo & Wang, 2006; Wong & Wang, 2006). For example, increasing the duration of recurrent processing – like raising the neural threshold – results in a longer delay or latency between neural pulses as input into the network continues to accumulate.

Interestingly, both parameters have a qualitatively similar effect on the salience map. In particular, note that for low values of either parameter (e.g. the top two maps), the map is characterized by numerous peaks, including not only the upper and lower portions of the occluded bar, but also the edges of the occluding box and the background texture elements. However, as each parameter is increased, smaller activation peaks on the map are gradually inhibited, ultimately resulting in a few, distinct locations on the map with large activation peaks (e.g. the bottom two maps).

Simulation 1: the 'standard' screen

In the first simulation study, the model was presented with the canonical occluded-rod display used by Amso and Johnson (2006) in their assessment of perceptual completion in 3-month-olds. In this display, approximately one-third of the moving rod is occluded by a rectangular, blue screen (see Figure 1A). A total of 110 testing runs were performed, corresponding to 10×11 combinations or permutations of the two parameters of interest. Each run represented a unique combination of the spatial-competition distance and duration parameters: (a) the size of the spatial-competition filter varied in 10% increments from 10% to 100% of the input image, while (b) the number of spatial-competition processing iterations varied from 0 to 10. Within each run, all model parameters were held fixed and 10 infants were simulated, in order to equate sample sizes between the simulation and infant studies. Data for each simulated infant were generated by presenting the model with 20 repetitions of the occluded-rod display. The duration of a single display was 5 seconds (i.e. 150 frames at 30 frames per second), resulting in a total of 100 seconds (i.e. 5 seconds per display $\times$ 20 repetitions) per simulated infant.

In order to equate gaze shift frequency in the model with the frequency of saccades produced by 3-month-old infants, a pseudo-random value was selected prior to each gaze shift from a normal distribution corresponding to the parameters reported by Amso and Johnson (2006; i.e. $M = 4.57$ saccades per second, $SD = 0.52$). While latency (i.e. time since the last gaze shift) remained below this value, the model's virtual fixation point was held in place, and activation on the salience map was summed over consecutive input images. After latency reached or exceeded the value, the aggregate salience map was used to select a new location on the display (following the procedure described above), and the model's virtual fixation point was shifted to this new location.

The primary dependent measure was the proportion of rod scans produced by the model. The same coding scheme employed by Amso and Johnson (2006) was used to evaluate the performance of the model. In particular, Figure 4 illustrates the six areas of interest (AOIs) used to code the model's fixations. Areas 1 and 2 correspond to the upper and lower segments of the rod, respectively, while the remaining four areas correspond to the four quadrants. Note that a mutual-exclusivity constraint applied, such that each fixation could only be coded as contacting a single AOI. In particular, fixations to either of the rod AOIs took precedence over the four quadrants. Using this scheme, *rod scans* were defined as a gaze shift in which the start and end point of the shift either (a) remained within one of the rod AOIs (i.e. a lateral saccade, following the movement of a rod segment), or traveled from one rod AOI to the other.

Simulation 1 results

Figure 5 presents the results from Simulation 1, organized in a two-dimensional grid. Each cell (i.e. colored

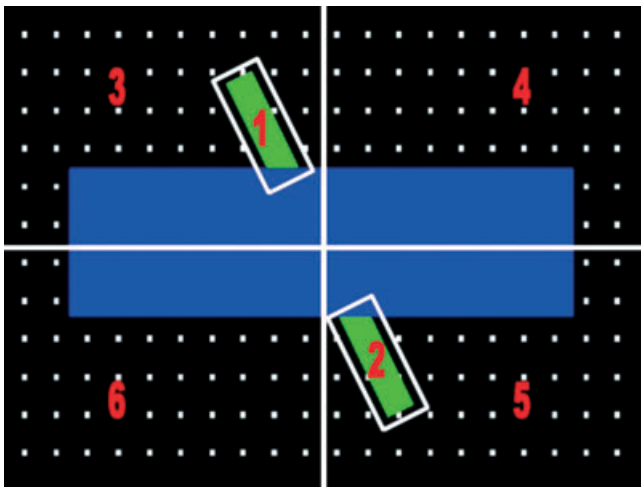

Figure 4 *Six areas of interest (AOIs) used for coding gaze shifts in the model.*

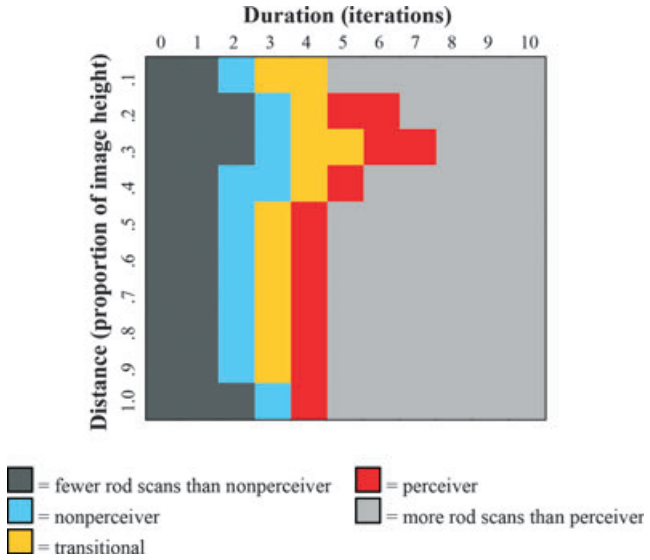

Figure 5 *Proportion of rod scans produced by the model in Simulation 1, as a function of the two parameters of interest. Duration of spatial competition increases toward the right, while distance (or size) of spatial competition filter increases toward the bottom. Each color indicates whether the proportion of rod scans (corresponding to that particular pair of parameter values) differed significantly from non-perceivers or perceivers (see text for details).*

square) in the figure corresponds to a specific combination of the two parameters of interest: (a) spatial-competition distance (increasing from top to bottom), and (b) spatial-competition duration (increasing from left to right). The color in each cell was determined by testing the mean proportion of rod scans produced by the model against the 95% confidence intervals for infant perceivers and nonperceivers, respectively. In particular, the mean proportion of rod scans generated by 3-month-old perceivers was 0.19 (SD = 0.11), resulting in the confidence interval [0.12, 0.25]. Similarly, the confidence interval for 3-month-old nonperceivers was [0.09, 0.17], based on a 0.13 mean proportion of rod scans (SD = 0.07). Using this analysis scheme, five performance groups were identified:

1. Dark gray = significantly fewer rod scans than nonperceivers
2. Light blue = within the confidence interval for nonperceivers
3. Yellow = within the confidence intervals for both nonperceivers and perceivers
4. Red = within the confidence interval for perceivers
5. Light gray = significantly more rod scans than perceivers

Figure 5 suggests several important results. First, yellow cells reflect a cluster of parameter values in the model that result in performance that spans both nonperceivers and perceivers. As this performance level overlaps both groups of 3-month-olds, we refer to cells (i.e. parameter combinations) that fall in this category as 'transitional'. Second, note that increasing the duration of spatial competition – while holding constant the distance of spatial competition – produces a consistent developmental trajectory in which nonperceivers emerge before perceivers (i.e. blue → yellow → red or blue → red). The only exception to this pattern occurs at the smallest filter size (i.e. 10%), where the model passes through the nonperceiver and transitional levels, but bypasses the perceiver level. Thus, across the majority of parameter values tested, increasing the duration of spatial competition is sufficient to reproduce the performance shift from nonperceivers to perceivers.

Third, and in contrast, increasing the distance or size of the spatial-competition filter – while holding the duration constant – does not result in a developmental trajectory in which nonperceivers emerge before perceivers. As Figure 5 illustrates, the only way to achieve this pattern by increasing spatial-competition distance is to also increase the duration of spatial competition. In light of the previous result, then, the current finding suggests that increasing the duration of spatial competition is not only a sufficient, but also a necessary condition for capturing the performance pattern reported by Amso and Johnson (2006).

Finally, Figure 5 also suggests an important interaction between the two parameters. In particular, for moderate to long distances of spatial competition (i.e. between 50% and 100%), the model reaches the perceiver performance level within four iterations of the spatial-competition loop. However, for shorter spatial-competition distances, a longer duration of spatial competition is needed to reach the corresponding performance level. This result is consistent with the prior observation that increasing either spatial-competition parameter has a qualitatively similar effect on the topology of the salience map (see Figure 3). In addition, it also suggests a 'compensatory' relation between the spatial-competition distance and duration. In other words, as the effective distance of spatial-competition filter is increased – which extends the reach of the inhibitory component – shorter durations of the spatial-competition process are needed to achieve the same level of performance.

Simulation 2: the 'narrow' screen

An important question suggested by the findings from Simulation 1 is whether the model can also account for infants' performance on the unity perception task prior to age 3 months. If the two dimensions of the grid illustrated in Figure 5 are viewed from a developmental perspective – that is, as dimensions of neural growth or maturation – then one prediction is that the performance of younger infants on the unity perception task should be represented by the upper-left corner of the figure. In other words, this parameter region corresponds to comparatively short horizontal connections among neighboring neurons in V1, and little or no recurrent activity in posterior parietal cortex. Insofar as the model

predicts relatively infrequent rod scans in this region of the parameter space, as the result of little or no spatial competition, we might consequently expect infants represented by these parameter values to not perceive the occluded rod as a coherent object during the unity perception task. (As it happens, there is no discernible effect of spatial competition at the lowest durations, as seen in Figure 5, implying that performance of younger infants may be represented by the leftmost column in its entirety.) Indeed, this is exactly the behavior pattern that has been reported in infants between ages 0 and 2 months (e.g. Johnson, 2004; Slater *et al*., 1996).

However, recall that 2-month-olds do show evidence of perceptual completion when the stimuli are designed to facilitate detection of the relevant features. For example, Johnson (2004; see also Johnson & Aslin, 1995; Johnson *et al*., 2008) presented 2-month-olds with a comparatively narrow screen (see Figure 6) in which a larger portion of the rod was exposed, and found that infants responded during the test phase like perceivers. In light of the current simulation results, this finding suggests that when infants lack sufficient endogenous resources to guide their attention toward relevant visual stimuli, exogenous (visual-spatial) influences can help guide attention in an equivalent manner.

Therefore, a prediction that follows from this line of reasoning is that by presenting the narrow-screen display to the model, we should expect an increase in the proportion of rod scans, compared to the proportion produced during the standard occluding screen. However, the increase in rod scans should also be qualified by the presence or degree of spatial competition in the model: specifically, presentation of the narrow screen should increase rod scans *for low values of spatial competition, but as spatial competition increases, rod scans should in fact decrease*. This prediction is based on the fact that for low values of spatial competition, revealing more of the occluded, moving rod should increase its salience. Meanwhile, for moderate to high values of spatial competition, revealing more of the rod should *decrease overall salience of the rod* by allowing a larger portion of the upper and lower segments to compete for attention. In other words, with higher levels of spatial competition (i.e. global inhibition), a larger surface area of exposed rod will result in decreased salience of the rod.

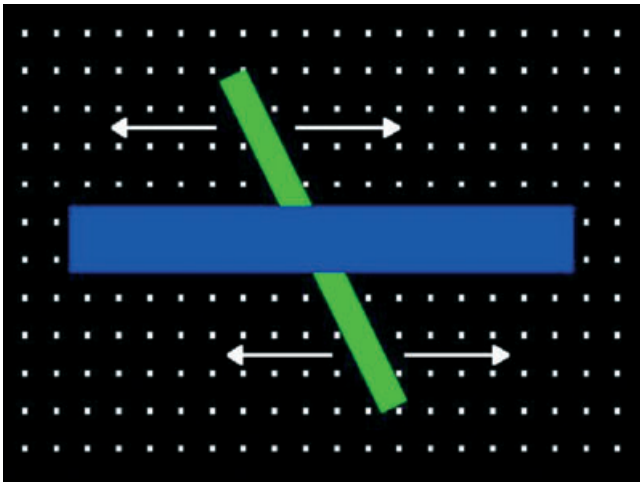

Figure 6 *The 'narrow' screen display used in Simulation 2.*

In order to evaluate these predictions, we repeated the testing procedure used in Simulation 1. As Figure 6 illustrates, however, the standard occluder was replaced with a comparatively narrow screen, which was roughly 50% narrower than the original. As a result, the AOIs corresponding to the upper and lower rod segments (i.e. AOI 1 and 2) were increased in size as a function of the larger segments. Otherwise, Simulation 2 followed the same procedure as before, and in particular, the same 110 combinations of parameter values were tested.

Simulation 2 results

The results from Simulation 2 are presented in Figure 7. Before comparing these findings with those from Simulation 1, we note two important considerations. First, as Johnson (2004) measured looking time but not gaze patterns, the data from that study do not provide an estimate of the proportion of rod scans in 2-month-olds. Second, it is an open question whether the mean proportion of rod scans generated by 3-month-old perceivers and nonperceivers (Amso & Johnson, 2006) should generalize to younger infants (i.e. 2-month-olds) tested with the same display. However, 2-month-olds' scanning patterns have been recorded with a narrow screen stimulus (Johnson *et al*., 2008), and as a group, infants in that study showed a greater proportion of rod scans overall (M = .28) than the 3-month-olds observed by Amso and Johnson (M = .16), evidence that the narrow occluder facilitated a higher baseline rate of rod scans. (For perceivers in the Johnson *et al*. experiment, M rod scans = .40 versus .18 for nonperceivers.)

In order to compare the results of Simulations 1 and 2, Figure 7A uses the same color scheme as Figure 5 to group the model's proportion of rod scans into four levels (note that the lowest level of rod scans, 'significantly fewer rod scans than nonperceivers', was not observed in Simulation 2). There are three major results illustrated in Figure 7A. First, as in Simulation 1, the four performance groups are organized in roughly vertical regions, supporting the previous finding that changes in the duration of spatial competition are the primary influence on the proportion of rod scans. Second, relative to the pattern of results in Figure 5, note that parameter regions corresponding to the nonperceiver and transitional levels (i.e. light blue and yellow, respectively) have expanded and shifted toward the left in Figure 7A (i.e. toward shorter durations of spatial competition). This finding is consistent with the first prediction, and suggests that for low levels of spatial competition, presentation of the narrow occluder increases the proportion of rod scans in the model.

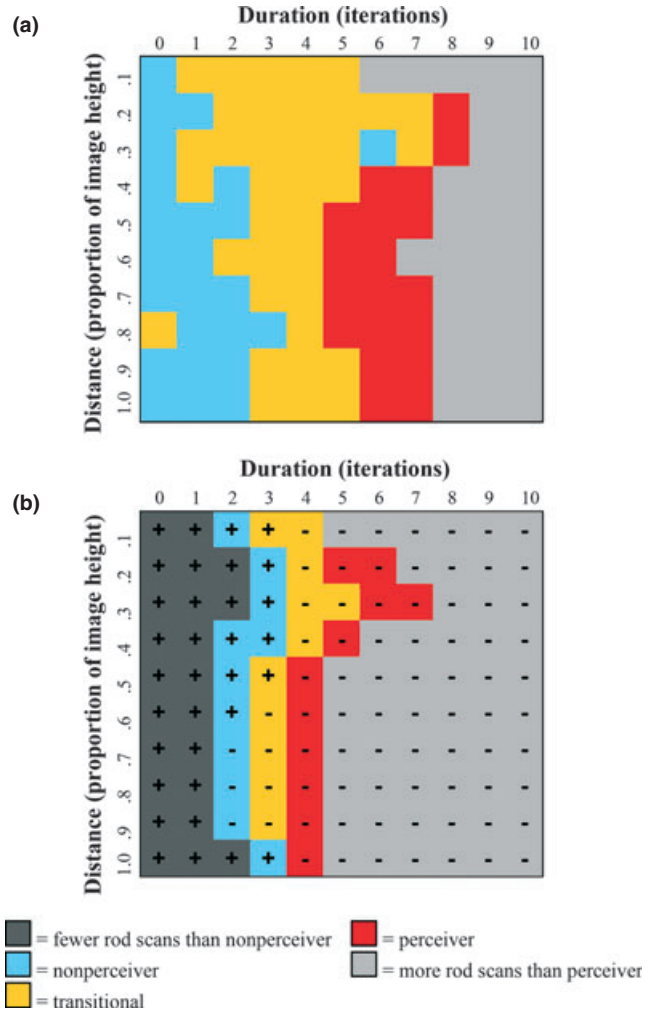

Figure 7 *(A) Proportion of rod scans produced by the model in Simulation 2, in response to the narrow screen. (B) The distribution of rod scans from Simulation 1 is presented, with a change score (i.e. Simulation 2 – Simulation 1) superimposed on each cell. A '+' indicates an increase in rod scans from the standard to the narrow screen, while '-' indicates a decrease in rod scans.*

In order to directly evaluate both this and the second prediction – that is, decreased rod scans for moderate to high levels of spatial competition – Figure 7B presents a change score (i.e. narrow occluder minus standard occluder) superimposed on to the results from Figure 5 (standard occluder). In particular, plus signs (+) indicate parameter combinations in which presentation of the narrow occluder increased rod scans over those produced during the standard occluder, while minus signs (-) indicate parameter combinations in which presentation of the narrow occluder resulted in fewer rod scans. As Figure 7B indicates, the data supported both predictions: for low values of spatial competition (i.e. the left third of Figure 7B), the narrow occluder resulted in an increase in rod scans, while rod scans decreased for moderate to high levels of spatial competition (i.e. the middle and right third of Figure 7B). Importantly, these results help interpret data from the Johnson *et al.* (2008) study for a higher baseline rate of rod scans when infants view a narrow occluder stimulus, in their suggestion that a reduction in spatial competition is responsible for this effect.

Discussion

The current simulation findings both replicate and extend our earlier work. First, the salience-based model captures changes in a key scanning pattern that is produced by infants during the unity perception task, and which distinguishes nonperceivers from perceivers (e.g. Johnson *et al.*, 2004; Amso & Johnson, 2006). In particular, we presented the model with the same occluded-rod display that infants view, and measured the model's gaze patterns using the same behavioral index (i.e. rod scans). As we varied a model parameter that represents the duration of recurrent activation in the posterior parietal cortex, we found that lower values of the parameter resulted in performance that matched nonperceivers, while higher values resulted in performance that matched perceivers.

Second, the model described here also captures infants' performance on a visual search task (Schlesinger *et al.*, 2007). It is important to note that in both simulation studies, an increase in the same model parameter (i.e. recurrent parietal loops) results in the model transitioning from the performance pattern produced by nonperceivers to that produced by perceivers. Therefore, the model not only provides an account for how perceptual completion and visual search are correlated at age 3 months (Amso & Johnson, 2006), but more importantly, it also suggests that performance on both tasks can be modulated by developmental changes in a single, underlying neural mechanism. Taken together, the findings from the two simulation studies provide additional support for our hypothesis that progressive improvements in visual selective attention promote the development of perceptual completion.

Third, the current simulation findings also suggest that the growth of horizontal connections in V1 is not sufficient to account for the developmental transition on the perceptual completion task from nonperceiver to perceiver. However, it is likely that this substrate contributes to the development of visual selective attention, insofar as variation in the corresponding model parameter accounts for infants' performance on the visual search task. In addition, the same neural mechanism is also implicated in the development of perceptual 'fill-in' (e.g. Albright & Stoner, 2002; Peterhans & von der Heydt, 1989; Ruthazer & Stryker, 1996).

In addition to capturing infants' performance on the canonical version of the perceptual completion task, the model also helps to explain why presenting the narrow screen facilitates unity perception in younger infants (e.g. Johnson, 2004; Johnson & Aslin, 1995). In particular, the findings from Simulation 2 suggest that for infants who

are unable to 'connect' the two rod segments occluded by the standard screen (i.e. non-perceivers), presenting a narrow screen increases the salience of the exposed rod segments, and as result, increases the proportion of rod scans. Thus, the model generates two key predictions. First, non-perceivers – as assessed with the standard screen – should produce an increase in fixations toward the rod segments when presented with the narrow screen display, and consequently show a post-habituation preference for the broken-rod test display. Second, and perhaps counter-intuitively, the model also predicts that infants who perceive unity during the standard-screen display should in fact *decrease* their rod scans when viewing the narrow-screen display.

The current modeling findings also suggest at least two important implications for the development of perceptual completion, and more generally, visual selective attention. First, as Figure 5 illustrates, there are multiple, convergent pathways through the parameter space from nonperceiver to perceiver, which may be manifested as individual differences between infants in developmental rate. An open question is whether such differences are caused by transient fluctuations in the developmental process, which ultimately even out among infants, or if instead they reflect qualitatively distinct attention 'styles' that remain stable over time (e.g. Bornstein & Sigman, 1986; McCall & Carriger, 1993).

A second issue concerns the way in which the current model was used to explore the two-dimensional parameter space. In particular, the perceptual completion and visual search tasks were simulated by explicitly setting the parameters of interest to particular values, and then measuring the model's performance on each task. The strategy of hand-tuning the model is roughly analogous to a maturational process, in which the underlying neural substrate grows independent of environmental influence. However, neurophysiological evidence suggests that while the initial growth of this substrate (i.e. prenatally) is largely due to genetic influence, subsequent developmental changes are primarily driven by visual experience (e.g. Greenough & Black, 1999). In infant ferrets, for example, a coarse pattern of long-range horizontal connections in V1 is established shortly after birth, followed by a more systematic, experience-dependent pattern that takes shape with additional visual input (e.g. Ruthazer & Stryker, 1996).

The strategy we are pursuing to address this issue is to introduce a prediction-learning system into the model, which serves two functions. First, this system learns to detect statistical regularities between stimulus features in the visual input. This statistical-learning mechanism is similar to other models of perceptual completion (e.g. Franz & Triesch, 2010; Mareschal & Johnson, 2002). However, rather than experiencing the input passively, the prediction-learning system is driven by the sequence of eye movements generated by the gaze-control system. Indeed, our latest simulations findings demonstrate that the prediction-learning system improves its performance as spatial competition is increased (Schlesinger, Amso & Johnson, 2011). Second, while learning, the prediction system also generates errors. These prediction errors can then be used as a 'training signal' that tunes the values of parameters within the salience map system (e.g. Balkenius & Johansson, 2007; Weber & Triesch, 2006). Thus, our next goal is to demonstrate that a systematic increase in spatial competition will emerge as a by-product of two-way feedback between the gaze-control and prediction-learning systems.

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.

Balkenius, C., & Johansson, B. (2007). Anticipatory models in gaze control: a developmental model. *Cognitive Processing*, **8**, 167–174.

Bornstein, M.H., & Sigman, M.D. (1986). Continuity in mental development from infancy. *Child Development*, **57**, 251–274.

Fecteau, J.H., & Munoz, D.P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, **10**, 382–390.

Franz, A., & Triesch, J. (2010). A unified computational model of the development of object unity, object permanence, and occluded object trajectory perception. *Infant Behavior and Development*, **33**, 635–653.

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

Greenough, W.T., & Black, J.E. (1999).Experience, neural plasticity, and psychological development. In N.A. Fox, L.A. Leavitt, & J.G. Warhol (Eds.), *The role of early experience in infant development* (pp. 29–40). New York: Johnson & Johnson Pediatric Institute.

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

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

Johnson, S.P., & Aslin, R.N. (1995). Perception of object unity in 2-month-old infants. *Developmental Psychology*, **31**, 739–745.

Johnson, S.P., Davidow, J., Hall-Haro, C., & Frank, M.C. (2008). Development of perceptual completion originates in information acquisition. *Developmental Psychology*, **44**, 1214–1224.

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., De Weerd, P., Pinsk, M.A., Elizondo, M.I., Desimone, R., & Ungerleider, L.G. (2001). Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, **86**, 1398–1411.

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

Kellman, P.J., & Spelke, E.S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, **15**, 483–524.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, **4**, 219–227.

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

Lo, C.C., & Wang, X.J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, **9**, 956–963.

Lu, Z.L., & Sperling, G. (1995). Attention-generated apparent motion. *Nature*, **377**, 237–239.

McCall, R.B., & Carriger, M.S. (1993). A meta-analysis of infant habituation and recognition memory as predictors of later IQ. *Child Development*, **64**, 57–79.

Mareschal, D., & Johnson, S.P. (2002). Learning to perceive object unity: a connectionist account. *Developmental Science*, **5**, 151–185.

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *Journal of Neuroscience*, **9**, 1749–1763.

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

Ruthazer, E.S., & Stryker, M.P. (1996). The role of activity in the development of long-range horizontal connections in Area 17 of the ferret. *Journal of Neuroscience*, **16**, 7253–7269.

Schlesinger, M., Amso, D., & Johnson, S.P. (2007). The neural basis for visual selective attention in young infants: a computational account. *Adaptive Behavior*, **15**, 135–148.

Schlesinger, M., Amso, D., & Johnson, S.P. (2011). Increasing spatial competition enhances visual prediction learning. In A. Cangelosi, J. Triesch, I. Fasel, K. Rohlfing, F. Nori, P.-Y. Oudeyer, M. Schlesinger, & Y. Nagai (Eds.), *Proceedings of the First Joint IEEE Conference on Development and Learning and on Epigenetic Robotics*. New York: IEEE.

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.

Slater, A., Johnson, S.P., Brown, E., & Badenoch, M. (1996). Newborn infants' perception of partly occluded objects. *Infant Behavior and Development*, **19**, 145–148.

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, **13**, 33–49.

Treisman, A. (1988). Features and objects: the Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology A*, **40**, 201–237.

Valenza, E., & Bulf, H. (2011). Early development of object unity: evidence for perceptual completion in newborns. *Developmental Science*, **14**, 799–808.

Valenza, E., Leo, I., Gava, L., & Simion, F. (2006). Perceptual completion in newborn human infants. *Child Development*, **77**, 1810–1821.

Weber, C., & Triesch, J. (2006). A possible representation of reward in the learning of saccades. In F. Kaplan, P.Y. Oudeyer, P. Gaussier, J. Nadel, L. Berthouze, H. Kozima, C. Prince, & C. Balkenius (Eds.), *Proceedings of the Sixth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems* (pp. 153–160). Lund, Sweden: Lund University Cognitive Studies.

Wolfe, J.M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, **34**, 1187–1195.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, **26**, 1314–1328.

Received: 16 September 2010
Accepted: 30 March 2012

Appendix

In this section, we provide a detailed description of the computational steps that occur within each of the four major processing stages: retinal image, feature maps, salience map, and target selection.

Retinal image

The size of the occluded-rod animation event used by Amso and Johnson (2006) is 480 × 360 pixels, and is composed of 150 frames displayed at 30 frames per second. In order to use this animation event as input into the model, it is first converted to individual image frames at the same resolution (i.e. 480 × 360). Activity on the model's simulated retina is then produced by copying the pixel values from the input image sequence onto the retina, one frame at a time. These activation values are then propagated forward into the next processing stage.

Feature maps

During the second stage, feature maps are produced by processing the retinal image through four parallel feature-extraction channels: intensity (i.e. luminance), motion, color, and oriented edges. Before feature extraction, retinal images are subjected to two pre-processing steps. First, a grayscale copy of each image is produced (for the intensity, motion, and orientation channels; see below). Second, each pair of color and grayscale images is iteratively blurred seven times (using a Gaussian filter), resulting in a set of 16 pre-processed images (i.e. eight color and eight grayscale images) for each retinal image. As noted below, the blurred image sets are used as input to the feature-extraction process so that the corresponding features can be detected across three spatial scales.

Feature extraction

Features are extracted across all four feature channels. The eight grayscale images are used as input into to the intensity, motion, and orientation channels, while the eight color images are used as input in the color channel.

1. Intensity. The eight grayscale images are used as the raw intensity feature maps. No additional processing occurs to these images at this stage.
2. Motion. Motion is defined as the absolute value of the difference between consecutive intensity maps:

$$M(t+1) = |I(t+1) - I(t)| \quad (1)$$

In order to eliminate spurious image noise, only differences that are greater than 0.05 are maintained on the motion feature map (values below the threshold are set to 0).

3. Color. Following Itti, Koch and Niebur (1998), the three color subchannels in the retinal image are normalized, and then used to compute a fourth, yellow subchannel (all negative values are set to 0):

$$R = R - (G+B)/2 \quad (2)$$

$$G = G - (R+B)/2 \quad (3)$$

$$B = B - (R+G)/2 \quad (4)$$

$$Y = (R+G)/2 - (|R-G|)/2 - B \quad (5)$$

4. Orientation. Oriented edges are extracted from each of the eight grayscale images with a Gabor filter (see Itti, Koch & Niebur, 1998; Gabor filtering combines a cosine grating and 2D Gaussian kernel). Four raw orientation maps are created for each grayscale image: 0°, 45°, 90°, and 135°.

Center–surround contrast

The raw feature maps are next used to compute center-surround contrast feature maps. In particular, recall that each retinal image is blurred iteratively seven times. Therefore, as a proxy for center-surround excitation-inhibition, contrast is defined as the absolute value of the difference between pairs of feature maps: a less-blurred or higher-resolution feature map represents activity in the center, while a more-blurred, lower-resolution map represents activity in the surround:

$$F_C = |F_{Hi} - F_{Lo}| \quad (6)$$

where C indexes the resulting contrast map, F is the corresponding feature map, and Hi and Lo index the level of blur (i.e. 1 = no blur while 8 = strong blur). For each feature channel, contrast maps are computed at three spatial scales (i.e. fine, medium, and coarse) by selecting pairs of feature maps at different levels of blur: 2 versus 4, 4 versus 6, and 6 versus 8, respectively. Note that a slightly different process is used to compute the color contrast maps. In particular, at this stage the center–surround computation is combined with an opponent-color computation, in order to produce RG and BY contrast feature maps:

$$F_{RG} = |R_{Hi} - G_{Hi}| - |R_{Lo} - G_{Lo}| \quad (7)$$

$$F_{BY} = |B_{Hi} - Y_{Hi}| - |B_{Lo} - Y_{Lo}| \quad (8)$$

At the end of the feature-extraction stage, there are a total of 24 feature maps: eight feature channels or subchannels (i.e. one intensity, one motion, two color, and four orientation maps) at each of three spatial scales.

Spatial competition

During the final stage of feature map computation, each of the 24 feature maps is normalized, in order to correct for any biases introduced across the different feature-extraction processes. Next, each map is convolved (i.e. filtered) with a difference-of-Gaussian (DoG) kernel:

$$DoG(x,y) = \frac{c_{ex}^2}{2\pi\sigma_{ex}^2} e^{-(x^2+y^2)/(2\sigma_{ex}^2)} - \frac{c_{inh}^2}{2\pi\sigma_{inh}^2} e^{-(x^2+y^2)/(2\sigma_{inh}^2)} \quad (9)$$

where x and y denote the location on the given feature map, ex and inh represent excitatory and inhibitory components, and c and σ are constants that influence the shape of the excitatory and inhibitory Gaussian functions. For the current simulations $c_{ex} = 0.5$, $\sigma_{ex} = 0.02$, $c_{inh} = 1.5$, and $\sigma_{inh} = 0.25$. Although the DoG kernel must be 'square' (i.e. the same size in the x and y dimensions), the size of the kernel can be arbitrarily chosen. Thus, the smallest possible size is 1×1 pixel, while the largest size is 360×360 pixels (i.e. the height of the retinal image). Note that it is the size of this kernel that was manipulated in the model as a parameter that represents growth in horizontal connections in V1.

Similarly, recall that spatial competition is a recurrent or iterative process, and that the number of iterations was parameterized and also systematically varied in the model. In particular, at each iteration, the product of the filtering process is added to the previous feature map:

$$F = F + (F * DoG) - C_{inh} \quad (10)$$

where C_{inh} is a global constant (0.02) that helps to 'break ties' in regions of the image where inhibition and excitation are approximately equal. Any negative values that are produced by the spatial competition process are set to 0.

Salience map

During the third stage, each of the 24 feature maps is rescaled to 27×20 pixels. This rescaling not only accelerates the modeling process, but also consolidates or pools nearby, active regions on each map active into contiguous areas of activity. The salience map is then produced by summing together the 24 separate feature maps.

Target selection

During a simulated trial, the initial fixation point is determined by randomly selecting a location on the retinal image. In addition, a gaze-shift latency value is also randomly selected from the normal distribution with $\bar{X} = 210$ ms and $\sigma = 20$. The first retinal image is propagated through the model, as described above, resulting in activity on the salience map. Given that the framerate of the original animation is 30 fps, the simulated duration of a retinal image is 33.3 ms. As a result, activity on the salience map aggregates across successive retinal

images, until the gaze-shift latency value is exceeded. At this point, the 100 most active locations on the salience map are selected, and the activity level at each of these locations is converted to a probability p following the softmax function:

$$P(x,y) = e^{S(x,y)/\tau} / \sum_{i=1}^{100} e^{S(x_i,y_i)/\tau} \quad (11)$$

where $S(x,y)$ is the salience at location (x,y), τ is a constant (i.e. 5), and i indexes each of the 100 selected locations. Finally, one of the 100 locations is randomly selected from this distribution, as a function of the probability value assigned to each location.

After a target is selected, the model's virtual fixation point is updated to the selected location. In addition, the activity on the salience map is reset to 0. The process of selecting a new gaze-shift latency value, propagating the retinal image, and updating/aggregating the activity salience map is the repeated. This process continues until the final retinal image is propagated through the model.