**4th International Conference on**
**Development and Learning and on Epigenetic Robotics**
**October 13-16, 2014. Palazzo Ducale, Genoa, Italy**

WeIDP.4

# Learnability of Infants' Center-of-Gaze Sequences Predicts Their Habituation and Posthabituation Looking Time

Matthew Schlesinger
Brain and Cognitive Sciences
Department of Psychology
Southern Illinois University
Carbondale, IL 62903
Email: matthews@siu.edu

Scott P. Johnson
Department of Psychology
University of California, Los Angeles
Los Angeles, CA 90095
Email: scott.johnson@ucla.edu

Dima Amso
Cognitive, Linguistic, and Psychological Sciences
Brown University
Providence, RI 02912
Email: Dima_Amso@brown.edu

*Abstract*— **Our recent work, using both computer-animated and natural images, demonstrates that infants' eye movements are sequentially structured. In particular, we employ a simple recurrent network (SRN) to estimate the learnability of infants' gaze sequences. The current study directly compares this learnability metric with infants' global looking time on a trial-by-trial basis, during both an habituation display and two posthabituation test displays. The results confirmed our prediction that the relative learnability of infants' gaze sequences during the latter part of the habituation phase predicts global looking time. We also found that learnability during the later habituation trials predicts looking time to one of the two posthabituation test displays. These findings provide support for our hypothesis that infants' visual exploration is guided implicitly by the goal to generate gaze patterns that are sequentially predictable.**

*Index Terms*— **perceptual development, center-of-gaze sequences, visual prediction learning, oculomotor skill.**

## I. INTRODUCTION

The *visual-processing bottleneck* refers to the fact that the retina collects more information than can be effectively processed by downstream brain areas. As a result, a number of evolutionary and developmental adaptations have emerged in mammals to address this bottleneck, including specialization not only at the retinal level (e.g., foveal and peripheral vision) but also within the cortical pathways that process visual information (e.g., dorsal and ventral visual streams) [1], [2]. In particular, the foveal vision system exploits a strategy that *trades space for time*: rather than taking in large amounts of data in a single fixation, the fovea acts like a small window over the visual scene, which moves rapidly from location to location.

Like other complex motor skills (e.g., reaching, crawling, etc.), eye movements are not fully organized and coordinated at birth. Instead, infants spend several months learning to control their gaze, and in particular, to generate gaze sequences that efficiently target informative regions of the visual scene [3], [4]. A fundamental question that we are investigating is: how do improvements in young infants' oculomotor skill influence the ability to perceive and recognize objects?

Over the last decade, we have explored this question by studying the *perceptual-completion task*, which is designed to assess infants' perception of a moving, partially occluded object. Figure 1A illustrates this occluded-rod display, which is presented first to infants, and repeated until they habituate to the display. Two subsequent displays (i.e., the complete rod and broken rod test displays; see Figures 1B and 1C) are then presented to infants and used to probe their perception and memory of the occluded-rod display.

One of the key findings from this paradigm is that *how* infants distribute their gaze spatially during the occluded-rod display is strongly associated with *whether they perceive* the rod as one single object, or instead as two disjoint objects. In particular, infants who direct their gaze more frequently to the visible rod segments of the occluded rod (Figure 1A) are more likely to perceive it as a single, unitary object [5], [6].

The behavioral data we have gathered thus far have focused on how infants' gaze patterns are structured over space. This work is complemented by a series of computational models that we have developed, which also highlight the presence of temporal structure that is embedded within infants' gaze sequences [7]-[9]. In particular, we are investigating the idea that infants' gaze sequences are *sequentially predictable*, that is, that the stream of recent
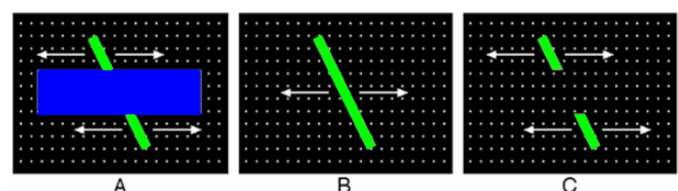


Fig. 1 Displays used to assess perceptual completion in infants: (A) habituation display, and (B) complete rod and (c) broken rod test displays.

fixations toward an object or scene generated by an infant provide sufficient information to predict the content of upcoming fixations.

Our analytical strategy is as follows: first, infants' eye movements are recorded as they view a display. We then convert the sequence of fixations produced by each infant into a stream of "center-of-gaze" (or COG) image samples, where each sample approximates the portion of the image visible to the fovea of a human observer while fixating the given location on the image. Next, we present these COG sequences to a simple recurrent network (SRN), which is trained to predict the sequence. The learning trajectory of the SRN then provides a metric for estimating the "learnability" of the COG sequence.

We have successfully applied this approach to infants' gaze data not only from the perceptual-completion task, but also from an image free-viewing task. In both cases, a robust finding is that infants' COG sequences are learned by the SRN more accurately and rapidly than the sequences produced by either a saliency-based or random-gaze model [8], [9]. In the case of the image free-viewing task, we have also found that infants' COG sequences are more learnable than the COG sequences produced by a sample of adult observers [8].

In the current paper, we expand our analytical framework beyond the moment-to-moment timescale of discrete gaze samples, to also include the longer timescale on which *global looking time* is measured. Specifically, we define global looking time as the cumulative number of seconds that an infant views a stimulus or event, during an experimentally-defined trial. This longer timescale is a crucial dimension, as it is the conventional method by which infants' visual attention is measured in habituation-dishabituation and preferential-looking studies [10].

We then hypothesize that individual differences between infants in the learnability of their COG sequences will be correlated with their global looking time, during both habituation trials and subsequent posthabituation test trials. This prediction is motivated by the "comparator theory," which proposes that looking time reflects the process of encoding the visual stimulus, and that longer looking times correspond to incomplete or partially-encoded displays [11], [12]. Specifically, we reason that infants' COG sequences and their global looking time should be correlated if gaze behavior on both timescales is driven by a common, underlying cognitive process.

Therefore, our first prediction was that during habituation to the occluded-rod display, infants who produce more learnable COG sequences (i.e., generate lower residual SRN errors) would also tend to have shorter looking times.

Second, we also predicted that the correlation between COG learnability (i.e., residual SRN errors) and looking time would vary as a function of habituation level. Thus, we expected that residual SRN errors and looking time would be uncorrelated at the start of habituation (i.e., the first three

trials; hereafter referred to as "early habituation"), due to the fact that infants were initially unfamiliar with the occluded-rod display, and therefore, their early gaze patterns were largely exploratory. In contrast, a positive correlation was predicted between residual SRN errors and infants' corresponding looking times at the end of habituation (i.e., the final three trials; hereafter referred to as "late habituation").

Third and lastly, if COG learnability during habituation accurately estimates infants' encoding of the occluded-rod display, it should also predict their looking time to the two test displays (i.e., the complete rod and broken rod; see Figures 1B and 1C). In particular, we followed the reasoning that if infants perceived the occluded rod as a unitary object, then the broken rod test display should be comparatively novel relative to the complete rod test display, thus leading to longer looking to the broken rod over the complete rod. Given this logic, we predicted that residual SRN errors during late habituation would be *negatively correlated* with looking time to the broken rod test display, and *positively correlated* with looking time to the solid rod test display.

## II. METHOD

### A. Stimuli

The habituation display (Figure 1A) was 640 x 480 pixels. The display began with the partially-occluded rod on the right side of the screen. The rod translated laterally to the left and then returned to its starting position, moving at a constant speed and completing one cycle of movement in 3.5 s. During both the complete-rod and broken-rod test displays, the center occluding box was removed, revealing either a solid rod or two rod segments (Figure 1B and 1C, respectively. In all other respects, the test displays were identical to the habituation display.

### B. Participants

Twelve 3-month-old infants (age, $M = 87.7$ days, $SD = 12$ days; 5 females) participated in the study. Infants sat on their parents' laps approximately 60 cm away from a 76 cm monitor in a darkened room. Eye movements were recorded using the Tobii 1750 remote eye tracker. Before the beginning of each trial, an attention-getter (an expanding and contracting children's toy) was used to attract infants' gaze to the center of the screen. As soon as infants fixated the screen, the attention-getter was replaced with the experimental stimulus and timing of trials began. The habituation event repeated while infants continued to view it. Each trial ended when the infant looked away for 2 s or when 60 s had elapsed.

The habituation display was presented until looking times declined across three consecutive trials that summed to less than half the total during the first three trials. Across the sample of 12 infants, the minimum number of habituation trials was 5, and the maximum was 11. Immediately after

habituating, infants viewed the test displays twice each in random order.

### C. COG image sequences

The COG image samples were produced by first converting the habituation display into a series of still-frame images. Next, each infant's gaze points were projected on to the corresponding frames of the habituation display. A 41 x 41 pixel image was then sampled from the resulting still frame, centered at the gaze point. The dimensions of the COG sample were derived from the display size and infants' viewing distance, and correspond to a visual angle of 1.8°, which falls within the estimated range of the angle subtended by the human fovea [13]. In order to facilitate the training process, note that each of the COG samples was converted from color (RGB) to grayscale.

### D. Model architecture and learning algorithm

Learnability estimates for the COG sequences were generated by training a 3-layer Elman network architecture, with recurrent connections from the hidden layer back to the input layer (i.e., SRN; [14]).

The input layer of the SRN was composed of 2083 units, including 1681 units that encoded the grayscale pixel values of current COG sample, 400 context units (which copied back the activity of the hidden layer from the previous time step), and 2 input units that encoded the x- and y-coordinates of the upcoming COG sample (normalized between 0 and 1). The input layer was fully connected to the hidden layer (400 hidden units, i.e., approximately 75% compression of the COG sample), which in turn was fully connected to the output layer (1681 units).

An individual training trial proceeded as follows: a COG sequence was selected at random (see *Training Regime*, below), and the first COG sample in the sequence was presented to the SRN. For this first sample, the activation of the context units was set to 0.5. These pixel values then propagated forward through the network, resulting in the predicted next COG sample. This output was compared to the second COG sample in the sequence, and the root mean-squared error (RMSE) was calculated. Next, the standard backpropogation-of-error (i.e., backprop) learning algorithm was used to adjust the SRN's connection weights (i.e., training was pattern-wise). The activation values from the hidden layer were then copied back to the input layer, and the second COG sample was presented to the SRN. This process continued until the second-to-last COG sample in the sequence was presented.

### E. Training regime

In order to generate a unique learnability estimate for each infant, we initialized one SRN per infant, and then trained each network on the COG sequences that were produced by the corresponding infant. While our analyses below contrast the network's performance on the first 3 and last 3 habituation trials, it should be noted that each SRN was trained on *all of the habituation trials* from the respective infant. In addition, within each epoch, the order of trials was randomized so that the learning of the COG sequences was independent of the sequence in which the habituation trials occurred.

In addition, we also noted that the number of fixations generated by each infant varied during habituation. As a result, the COG sequences for some habituation trials were relatively short while others were comparatively long (e.g., 5 vs. 100 COG samples). Because learning was pattern-wise (i.e., the SRN weights were updated after each COG sample), we therefore adjusted the learning rate to compensate for this effect. In particular, the base learning rate was set to 0.10, and then for each trial, the base rate was divided by the number of COG samples (i.e., fixations produced by the infant) for that trial. This step ensured that the overall magnitude of weight change for each COG sequence was not confounded with the number of training patterns presented during the corresponding trial.

Ten training runs were performed. At the start of each run, a single SRN was randomly initialized with connection weights between -1 and 1, and the SRN was then "cloned" 12 times (i.e., the same initial network was used for all infants). Next, each of the 12 SRNs was trained on the COG sequences from the corresponding infant for 200 epochs. For all of the analyses reported below, the performance data for the SRNs were averaged across the 10 runs.

## III. RESULTS

The presentation of the results is divided into two sets of analyses. In the first set, we compare changes in eight measures (including looking-time, eye-movement, and SRN performance data), between early and late habituation trials. In the second set of analyses, we assess the predicted pattern of correlations between the SRN performance data and looking times during the habituation and test trials.

### A. Analysis of early versus late habituation trials

Table 1 presents comparisons of changes in eight measures between early and late habituation trials. In the Table, the first and second rows present the mean values obtained during early and late habituation, respectively (i.e., the first three and last three habituation trials). The third row presents the difference between the two phases (i.e., late habituation − early habituation) for each corresponding measure.

1) *Global looking time:* global looking time for each habituation trial was determined by summing the durations of all fixations during a trial. Looking time decreased from early to late habituation by 11.89 s (M = 17.43 and 5.54, SD = 11.87 and 3.68, respectively). This change was statistically significant ($t(11) = 3.86$, $p = .003$).

2) *Fixations:* on average, infants generated 48.38 fixations during early habituation (SD = 28.72). This

Table 1

*Comparison of Looking Times, Gaze Data, and SRN Performance as a Function of Habituation Phase*

| Early vs. Late Habituation | Looking Time (s) | Fixations | Looking Time per Fixation (s) | Salience | Revisit Rate | Dispersion (pixels) | Gaze Shift Distance (pixels) | Residual Error |
|---|---|---|---|---|---|---|---|---|
| First 3 Trials | 17.43 | 48.38 | 0.36 | 0.65 | 0.29 | 70.37 | 59.63 | 0.13 |
| Last 3 Trials | 5.54 | 16.86 | 0.32 | 0.64 | 0.39 | 64.69 | 66.80 | 0.11 |
| Last 3 – First 3 | -11.89** | -31.53** | -0.04* | -0.01 | 0.09** | 5.67 | 7.18* | -0.02** |

** $p < .01$, * $p < .05$

value fell to 16.86 fixations by the last three habituation trials (SD = 10.03). Parallel with the decrease in looking time, the decrease in the number of fixations per trial was also statistically significant ($t(11) = 3.65$, $p = .004$).

3) *Looking time per fixation:* the duration of infants' fixations also decreased from early to late habituation (M = .36 and .32 s, SD = .12 and .13, respectively). Although the magnitude of the change was small (i.e., 40 ms), this change reached statistical significance ($t(11) = 2.29$, $p = .043$).

4) *Salience:* in order to estimate the degree to which infants fixated salient regions of the occluded-rod display, each still-frame image was converted into a saliency map (see [8] and [9] for a complete description of this process), and each fixation was projected onto the corresponding map, resulting in a saliency value (scaled between 0 and 1). Mean saliency values for each trial were then computed by taking the average saliency over the set of fixations during the trial. As Table 1 indicates, on average infants oriented to regions of the occluded-rod display that were roughly 65% of the maximal saliency present in the display (SD = .09 and .07, respectively). There was no significant change in this pattern from early to late habituation ($t(11) = .31$, $p = .76$).

5) *Revisit rate:* the revisit-rate measure estimated the tendency for infants to return to previously-fixated locations in the occluded-rod display. This value was estimated for each trial by computing a 2D histogram of the display, and then incrementing a 41 x 41 pixel window (centered at the fixation point) by 1 for each fixation during the trial. The histogram was then normalized by the number of fixations for the given trial, and the maximum value was recorded, reflecting the proportion of total fixations for which the most frequently-fixated location was visited. Using this procedure, the revisit rate increased significantly from early to late habituation (M = .29 and .39, SD = .09 and .07, respectively; $t(11) = 3.34$, $p < .007$).

6) *Dispersion:* the "spread" of fixations was computed for each trial by calculating the centroid of the gaze points (i.e., the mean fixation location), then calculating the mean distance of the fixations (in pixels) from the centroid for the trial, and then averaging the resulting dispersion values. As a result, mean dispersion of fixations was 70.37 pixels during early habituation, and 64.69 pixels during late habituation (SD = 9.34 and 9.52, respectively). This change was not statistically significant ($t(11) = 1.67$, $p < .12$).

7) *Gaze shift distance:* during early habituation, the average gaze shift traveled 59.63 pixels; this distance increased to 66.80 pixels by late habituation (SD = 17.39 and 20.79, respectively). Though a relatively small change, it was statistically significant ($t(11) = 2.93$, $p = .014$).

8) *Residual error:* while the previous measures focused on infants' gaze data, the final comparison presents a key result from training the SRNs on the COG image sequences. In particular, we defined *residual error* as the average RMSE produced by the SRNs during the final 20 epochs of training (recall that RMSE reflects the normalized difference, averaged per pixel, between the observed and predicted sequences of COG samples). An important finding is that residual error was significantly higher for the early habituation trials than it was for the late trials (M = .13 and .11, SD = .01 and .02, respectively; $t(11) = 3.35$, $p = .006$). This is a noteworthy result, first because it suggests that the COG sequences produced by infants during early habituation were more difficult for the SRNs to learn than those produced at the end of habituation. In addition, it also provides support for our hypothesis that infants' COG sequences reflect the same underlying learning process that occurs over habituation time (i.e., gradual decrease in attention). We return to this issue in the Conclusion section.

*B. SRN performance - looking time correlations*

Table 2 presents the pattern of correlations between residual error, during early and late habituation, and looking time during habituation and test trials. First, as predicted, the mean residual errors produced during training on the first three habituation trials were not correlated with mean looking times during either early or late habituation (both $p$s > .16). Similarly, residual errors from early habituation were

Table 2

*Correlations between SRN Performance and Looking Time as a Function of Habituation Phase*

| Early vs. Late Habituation | Habituation Trials | | Test Trials | |
| --- | --- | --- | --- | --- |
| | LT First 3 Trials | LT Last 3 Trials | LT Complete Rod | LT Broken Rod |
| Residual Error First 3 Trials | -0.08 | 0.01 | 0.33 | -0.42 |
| Residual Error Last 3 Trials | -0.30 | 0.42* | 0.08 | -0.63* |

* $p < .05$

also not correlated with looking times to either the complete rod or broken rod test displays (both $ps > .17$).

Second, and also as predicted, mean residual errors from training on the late habituation trials were significantly correlated with looking time during the same trials. In particular, late habituation trials that produced lower residual errors during training of the SRNs tended to have shorter looking times ($p = .012$). In contrast, residual errors from the late habituation trials were not significantly correlated with looking times during early habituation trials ($p > .07$).

Finally, and also as expected, late habituation residual errors were significantly (negatively) correlated with looking time during the broken rod test display ($p = .028$). Specifically, infants who generated COG sequences during late habituation that were easier for the SRNs to learn subsequently went on to look longer at the broken rod test display during the posthabituation phase. Lastly, and contrary to our prediction, late habituation residual errors were not correlated with looking time to the complete rod test display ($p > .80$). These findings are both also novel and potentially informative, and we return to each of them in the next section.

IV. CONCLUSION

Our recent work demonstrates three key properties of infants' gaze patterns [7]-[9]. First, there is sequential structure embedded within the eye movements that infants produce. Second, this structure can be estimated by analyzing the stream of COG samples that infants generate as they scan an event or scene. Third, infants' COG sequences are easier for an SRN to learn than the sequences produced by human adults, as well as those produced by a variety of artificial-observer models.

The primary assumption guiding our approach is that infants' gaze patterns are a form of visual exploration that is intrinsically motivated [15], [16]. In particular, we are hypothesizing that the mechanism underlying infants' gaze sequences is a "drive" to learn to predict the content of upcoming fixations. The current study offers a critical test of our approach, by investigating whether variation across

infants (i.e., individual differences) in the learnability of their COG sequences is correlated with the standard measure of visual attention, that is, looking time.

Our first prediction was that the COG sequences infants generated at the beginning of habituation would not be correlated with their looking time during those trials, due to the fact that such sequences should reflect a relative unfamiliarity with the event being viewed (i.e., the occluded-rod display). In other words, infants' ability to forecast or predict "what their gaze will encounter next" should be poor when they are initially exploring the display. The results confirmed our prediction.

Second, and in contrast, we predicted that the COG sequences produced at the end of habituation should accurately estimate or predict infants' looking time during the same trials. As expected, we observed a positive correlation between the COG learnability estimates (i.e., residual errors in the SRNs) and looking time during late habituation, indicating that infants who produced easier-to-learn COG sequences also had shorter looking times. Because we controlled for variation across infants and trials in the number of fixations produced (by adjusting the learning rate), this result is not an artifact of producing more or less fixations. Instead, it supports the idea that infants' COG sequences and their global looking time are either directly linked or have a common influence.

Third, and perhaps most importantly, we predicted that if the COG sequences produced at the end of habituation represented what infants had learned by scanning the occluded-rod display, then the learnability of those sequences should also be correlated with infants' looking time to the two test displays. The results partially supported this prediction. In particular, late habituation COG sequences were negatively correlated with looking time to the broken rod test display, but they were not correlated with the looking time to the complete rod test display.

In a sense, it might be argued that of the two test displays, the broken rod is the more "diagnostic" display. This is because, as we noted in the Introduction, when infants perceive the occluded rod as a coherent whole – that is, they implicitly fill-in or reconstruct the occluded portion of the rod – the broken rod is then a comparatively novel stimulus. From this perspective, then, it is tempting to conclude that infants' late habituation COG sequences in fact provide a meaningful estimate of how they perceive the occluded rod display, as these gaze patterns reliably predicted the number of seconds that infants spent looking at the broken rod test display.

The negative correlation we observed is particularly noteworthy, as it is not only consistent with comparator theory, but also suggests three implications. Specifically, when infants have successfully encoded (i.e., internally represented) the occluded rod display, they then: (1) look less at it because they have habituated, (2) scan it in a more regular, consistent, or stereotyped way, and (3) use that

representation as a reference for encoding the complete rod and broken rod test displays. In other words, shorter looking times and easier-to-learn COG sequences during late habituation may be driven by a common, underlying cognitive process that results in longer looking time to the broken rod test display.

An important question raised by these findings is: how do the specific visual features in the display contribute to changes in COG learnability, both within and across infants? For example, in the case of the occluded rod display, it is hypothetically possible for an infant to only fixate the background texture dots, resulting in a sequence of COG samples that is relatively easy to predict. Alternatively, it is possible for an infant to generate a completely random sequence of eye movements, in which the content of each fixation is independent from all other fixations, and as a result, no correlation between adjacent COG samples.

The data presented in Table 1 suggest some possible clues to answering this question. For instance, while our previous work indicates that saliency plays an important role in guiding or driving infants' gaze patterns, there is no evidence in the current study that saliency alters infants' COG sequences over time [17], [18].

On the other hand, Table 1 indicates a significant increase in the tendency for infants to revisit or return to previously-fixated locations, between early and late habituation. As a result, it is possible that the collective set of COG samples produced by infants are more similar to each other during late habituation than they are during early habituation. If this reasoning is correct, it would help to explain why COG sequence learnability also increases from early to late habituation, that is, increased learnability is due to an increase in the similarity of COG samples over habituation time.

We are currently exploring the statistical relation between revisit rates and COG learnability by generating artificial gaze sequences in which the probability of returning to a specific region of the display is systematically manipulated. In the case of the occluded rod display, however, it is important to note that because the display is dynamic, simply controlling for the location of the fixation (without also incorporating the time point in the event) may not be sufficient. We are therefore concurrently exploring methods for gauging the similarity of COG samples, as a way of measuring the overall "similarness" of infants' sequences.

As our understanding of the mechanisms that drive infants' visual exploration improves, we have begun to consider potential long-term applications of our approach. One ambitious test is to see whether we can develop a computational model that learns online and in real time to predict the next location where an infant will gaze. Such a model could potentially provide a number of unique insights not only for how visual perception and oculomotor control develop during infancy, but also for improving the design and performance of biologically-inspired machine gaze-control systems. An additional direction is to use our approach as a method for identifying deviations from typical development in young infants' gaze behavior, which may provide an early diagnostic tool for detecting risk for subsequent developmental disorders (e.g., autism spectrum disorder or attention deficit/hyperactivity disorder).

## REFERENCES

[1] M.S. Banks and P.J. Bennett, "Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates," *JOSA A*, vol 5, pp. 2059-2079, 1988.

[2] A.D. Milner and Goodale, *The Visual Brain in Action*, vol. 2. Oxford: Oxford University Press, 2006.

[3] Haith, M.M., *Rules That Babies Look By: The Organization of Newborn Visual Activity*. New Jersey: Erlbaum, 1980.

[4] G. Bronson, *The Scanning Patterns of Human Infants: Implications for Visual Learning*. Norwood, NJ: Ablex, 1982.

[5] S.P. Johnson, J.A. Slemmer, and D. Amso, "Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds," *Infancy*, vol. 6, pp. 185-201, 2004.

[6] D. Amso and S.P. Johnson, "Learning by selection: Visual search and object perception in young infants," *Developmental Psychology*, vol. 42, pp. 1236-1245, 2006.

[7] M. Schlesinger, D. Amso, and S.P. Johnson, "Increasing spatial competition enhances visual prediction learning," *Proceedings of the First Joint IEEE Conference on Development and Learning and on Epigenetic Robotics*, New York, 2011, pp. 1-6.

[8] M. Schlesinger and D. Amso, "Image free-viewing as intrinsically-motivated exploration: Estimating the learnability of center-of-gaze image samples in infants and adults," *Frontiers in Cognitive Science*, vol. 4, pp. 1-12, 2013.

[9] M. Schlesinger, S.P. Johnson, and D. Amso, "Prediction-learning in infants as a mechanism for gaze control during object exploration," *Frontiers in Perception Science*, in press.

[10] J. Colombo and D.W. Mitchell, "Infant visual habituation," *Neurobiology of Learning and Memory*, vol 82, pp. 225-234, 2009.

[11] W.R. Charlesworth, "The role of surprise in cognitive development," in *Studies in Cognitive Development: Essays in Honor of Jean Piaget*, D. Elkind and J. Flavell, Eds. Oxford, UK, 1969, pp. 257-314.

[12] E.N. Sokolov, *Perception and the Conditioned Reflex*. New York: Pergamon, 1963.

[13] B. Goldstein, *Sensation and Perception*. Belmont, CA: Wadsworth, 2010.

[14] J.L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, pp. 179-211, 1990.

[15] G. Baldassarre and M. Mirolli, "Intrinsically motivated learning systems: An overview," in *Intrinsically motivated learning in natural and artificial systems,* G. Baldassarre & M. Mirolli, Eds. Berlin: Springer-Verlag, 2013, pp. 1-14.

[16] M. Schlesinger, "Investigating the origins of intrinsic motivation in human infants," *Intrinsically motivated learning in natural and artificial systems*, G. Baldassarre & M. Mirolli, Eds. Berlin: Springer-Verlag, 2013, pp. 367-392.

[17] M. Schlesinger, D. Amso, and S.P. Johnson, "The neural basis for visual selective attention in young infants: A computational account," *Adaptive Behavior*, vol. 15, pp. 135-148, 2007.

[18] M. Schlesinger, D. Amso, and S.P. Johnson, "Simulating the role of visual selective attention during the development of perceptual completion," *Developmental Science*, vol. 15, pp. 739-752, 2012.