Location, Location, Location: Development of Spatiotemporal Sequence Learning in Infancy

Natasha Z. Kirkham
Birkbeck College, University of London,
Centre for Brain and
Cognitive Development

Jonathan A. Slemmer
The Ohio State University

Daniel C. Richardson
University of Reading

Scott P. Johnson
University of California, Los Angeles

We investigated infants' sensitivity to spatiotemporal structure. In Experiment 1, circles appeared in a statistically defined spatial pattern. At test 11-month-olds, but not 8-month-olds, looked longer at a novel spatial sequence. Experiment 2 presented different color/shape stimuli, but only the location sequence was violated during test; 8-month-olds preferred the novel spatial structure, but 5-month-olds did not. In Experiment 3, the locations but not color/shape pairings were constant at test; 5-month-olds showed a novelty preference. Experiment 4 examined ''online learning'': We recorded eye movements of 8-month-olds watching a spatiotemporal sequence. Saccade latencies to predictable locations decreased. We argue that temporal order statistics involving informative spatial relations become available to infants during the first year after birth, assisted by multiple cues.

An important aspect of the ability to perceive the visual environment as coherent and intelligible is an understanding of objects' spatial locations and what their present locations might predict about future events. Acquisition of this type of knowledge is essential for motion perception and for the production of action sequences, because one has to learn not only which actions are appropriate, but also where and when they should be performed. For example, if, while looking out the window of your house, you see your partner walking up the path to the front door, you can reasonably predict that you will see him next in the doorway of your house. You can use this information to guide appropriate anticipatory behavior, such as moving to a location that provides a view of the door to greet your partner as he or she comes inside. In other words, each visual event is temporally related both to the previous event and to the future event and occurs within a spatial context.

Spatiotemporal Sequence Learning in Adults

Recognition and exploitation of patterns of events to guide behavior relies on *spatiotemporal sequence learning*. Adults are competent at learning many kinds of complex spatiotemporal sequences (e.g., Chun & Jiang, 1998; Cleeremans, Destrebecqz, & Boyer, 1998; Conway & Christiansen, 2005; Howard, Mutter, & Howard, 1993; Mayr, 1996). Many of these studies were designed to examine sensitivity to the statistical or probabilistic structure of the input. In serial reaction time studies, for example, observers view a single repetitive stimulus presented sequentially at different locations and respond to each position by pressing a corresponding key (e.g., Nissen & Bullemer, 1987). Stimulus locations may follow a particular spatial and temporal pattern that a participant may be unable to describe explicitly, yet reaction times typically decrease reliably across trials. In contrast, reaction times show no improvement when stimuli are presented in a random order (Cohen, Ivry, & Keele, 1990; Curran & Keele, 1993; Nissen & Bullemer, 1987).

There is evidence, as well, that learning is independent of the specific motor response (e.g., a button press): Mere observation of a sequential pattern can lead to knowledge of serial order (Howard et al., 1993), and there appears to be no special benefit to learning imparted by manual responses, relative to oculomotor responses (Heyes & Foster, 2002).

Perception of scenes can be guided by statistical information and other contextual cues (Chun, 2000). Fiser and Aslin (2001, 2002), for example, presented adults with probabilistically structured sequences of single shapes and shape arrays and found that observers were sensitive to the statistical correlations among multipart objects presented simultaneously as well as to the joint and conditional probabilities of successive shape pairs. Participants viewed a series of multiple shape-pair displays, and were asked to rate familiarity of a previously seen shape pair in either a familiar or novel spatial arrangement during a test phase (Fiser & Aslin, 2001). The spatial structure of the object displays was encoded readily on the basis of presentation sequence, and performance was improved further when more cues were available to guide familiarity choices (e.g., if a novel test pair differed in both local and absolute spatial arrangement). In a subsequent study of temporal sequence learning in adults, Fiser and Aslin (2002) reported sensitivity to both the joint and conditional (predictive) probabilities of successive shape pairs. When joint probabilities were uninformative, participants used conditional probabilities to guide performance.

Temporal Sequence Learning in Infancy

Research concerned with the development of sequence learning has revealed a capacity to pick up temporal patterns under many conditions. Saffran, Aslin, and colleagues have found, for example, that 8-month-old infants can parse a stream of auditory stimuli based solely on the transitional probabilities within and between the syllables (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). Gomez and Gerken (1999) exposed 12-month-olds to a subset of strings produced by one of two artificial grammars and then tested the infants on their ability to discriminate new strings from both the familiar and the unfamiliar grammar. Infants preferred to listen to new strings from their training set relative to strings from the novel grammar, suggesting discrimination of the two grammars based on familiarity. These grammars differed only in terms of the ordering of word pairs: Individual words in the two sets and the starting and ending words were always the same. The only cues to the grammar, therefore, were contained

in word order, implying that the infants encoded the temporal patterns of word co-occurrences. Infants' ability to extract regularities in sequential input does not seem to be a language-specific mechanism, but exists broadly across audition as well. Infants parse auditory streams based on statistical probabilities even when the stimuli are tones (Saffran, Johnson, Aslin, & Newport, 1999), and at least one species of nonhuman primates, cottontop tamarins (which never develop humanlike language skills), can learn statistically structured sounds (Hauser, Newport, & Aslin, 2001).

Domain generality of infants' statistical learning was investigated further by Kirkham, Slemmer, and Johnson (2002), using visual stimuli. Two-, 5-, and 8-month-olds were familiarized with a series of six discrete colored shapes that loomed from the center of a display monitor. Presentation order was defined in part by statistical regularities: The shapes were organized into pairs, and the pairs were ordered randomly. That is, the first shape in a pair reliably predicted the second, but the next shape to appear could be any of the first members of a pair. As in experiments that probe implicit learning in adults and auditory statistical learning in infants and tamarins, the only cue to stimulus sequence lay in its statistical nature: There were no pauses between pairings, each discrete stimulus was the same duration, and there was no a priori relation between stimuli to provide further information for co-occurrence. Infants showed a reliable postfamiliarization preference for random sequences that violated the statistical stimulus grouping viewed initially. Interestingly, there were no age differences in performance, implying that visual statistical learning of simple temporal sequences is available early and may remain constant across development. Performance in this study was likely based on a true sensitivity to the transitional probabilities defining color/shape pairs, rather than to the frequency of occurrence of these pairs (Aslin, Slemmer, Kirkham, & Johnson, 2001).

Spatiotemporal Sequence Learning in Infancy

These experiments are important in providing evidence for statistical learning of temporal sequences early in development, but leave open the question of the capacity to learn spatial sequences. There is evidence from other paradigms, however, that infants show some sensitivity to spatial relations among repetitive events under certain conditions. For example, young infants learn simple (two-location), predictable spatial sequences in the visual expectation paradigm, which uses oculomotor anticipation as the

index of learning (Haith, 1993). There is some evidence as well for infants' sensitivity to spatial contingency in temporal sequences: Wentworth, Haith, and Hood (2002) presented 3-month-old infants with a spatiotemporal sequence in which a stimulus appeared on the left, in the center, or on the right of a computer monitor. Infants viewed either a fixed or a random pattern of locations, and in some cases there was a contingent relation between the identity of the central stimulus and the location of the next peripheral picture. The fixed sequence of three locations resulted in more eye movement anticipations, and there were more anticipatory saccades to the correct location when there was a contingent relation between central and peripheral events.

Notably, responses to event predictability in the visual expectation paradigm seem to be rather poor considering that infants view spatiotemporal patterns that are perfectly predictable, and there is little improvement across infancy (2 – 12 months): Typically, performance is correct on only 15% – 30% of trials, depending on the specific procedure (Canfield, Smith, Brezsnyak, & Snow, 1997). Young infants' performance in anticipation paradigms is higher when the event is part of a repetitive trajectory, as when an object emerges from behind an occluder at a predictable time and place (Johnson, Amso, & Slemmer, 2003), implying that spatial sequence learning can be facilitated by context, such as a multi-element scene. Further support for this notion comes from experiments that examine infants' responses to statistical structure in displays of greater complexity than the simple two- and three-location events described previously. Upon exposure to a sequence of static multi-element scenes, for example, 9-month-olds appeared to acquire the underlying statistical structure of the scene layout, attending longer following habituation to isolated element pairs that had co-occurred with a higher frequency within the familiar scenes (Fiser & Aslin, 2003). The infants also responded to the predictability (conditional probability) of co-occurrence of individual members of the pair. Additionally, Richardson and Kirkham (2004) demonstrated that 6-month-olds bind audiovisual information to moving locations when presented in a context of repetitive, predictable events. Infants were familiarized with two different animated objects on a computer screen, each accompanied by a distinct sound and appearing in a specific location in one of two square frames. The frames then moved to different positions on the screen and remained empty while the two sounds associated with each of the objects were played sequentially. The infants

looked significantly longer at the frame associated with the test sound, even though it was empty, and occupied a new location in space. This finding suggests that the infants exploited cross-modal regularities to learn about the locations of the events, and that those location representations were updated dynamically.

Rationale for the Present Experiments

As the results of these experiments reveal, spatiotemporal information plays a key role in perceptual development. In all of these studies, however, stimuli were presented either in static scenes or in fixed versus random locations. More subtle statistical regularities, which infants are demonstrably capable of encoding across time (Kirkham et al., 2002; Saffran et al., 1996), have not yet been employed to examine infants' sensitivity to spatial structure.

In the current study, we adapted the Kirkham et al. (2002) visual statistical learning paradigm to examine spatiotemporal statistical learning in which spatial statistics were informative with respect to the location of subsequent elements in the sequence. The goal was to explore whether infants could extract spatiotemporal correlations from a sequence that has distinctive statistics in both spatial and temporal dimensions, and if so, at what age, and under what conditions. Infants as young as 2 months of age appear to be sensitive to the serial order of a set of discrete objects (Kirkham et al., 2002), and by 9 – 11 months of age they encode spatial relations (Fiser & Aslin, 2003) and spatiotemporal action patterns (Baldwin, Baird, Saylor, & Clark, 2001). Can infants also detect the statistics of spatiotemporal sequences when the stimulus locations are changing?

Experiment 1

Method

Participants

Thirty-two full-term infants (20 girls) composed the final sample: sixteen 8-month-olds (M age = 8.2 months, SD = 0.4 months), and sixteen 11-month-olds (M age = 11.3 months, SD = 0.3 months). An additional 3 infants were observed but not included in the analyses, 2 due to fussiness (11-month-olds), and 1 fell asleep during the task (8-month-old). The infants were recruited by letter and telephone from hospital records and birth announcements in the local newspaper. Parents and infants received a small gift (a baby T-shirt or toy) for their participation.

Apparatus and Stimuli

A Macintosh G4 computer and 53-cm color monitor were used to collect looking time data and to present stimuli. An observer viewed the infant on a second monitor and entered looking time judgments with a keypress on the computer keyboard. The observer was unaware of the stimulus sequence viewed by the infant. The computer presented displays, recorded looking times, calculated the habituation criterion for each infant, and changed displays after the criterion was met. Stimuli consisted of a 3 × 2 grid (38.5 × 25.8 cm, 22.9° × 15.5° visual angle) defined by white lines on a black background, and a red circle that appeared sequentially. In each square of the grid, the circle was presented for 1 s and loomed from 4 cm to 24 cm in diameter (2.4° to 14.6°). At no time did the circle loom outside of its box in the grid, so there was never any spatial overlap between individual circles. The stream of stimuli was shown as long as the infant attended to the monitor. A trial ended when the infant looked away for 2 continuous seconds, or had looked for 60 s. Between trials, a beeping target was shown to attract attention back to the screen.

Procedure

Infants were tested individually and sat on a parent's lap 95 cm from the computer monitor. The parent was instructed not to interact with the infant or watch the monitor. The red circles appeared in a continuous stream of randomly ordered location pairs (e.g., Pair 1: top middle, followed by top left; Pair 2: bottom middle, followed by top right; Pair 3: bottom left, followed by bottom right) with only transitional probabilities defining boundaries within and between pairs (see Figure 1). As in the Kirkham
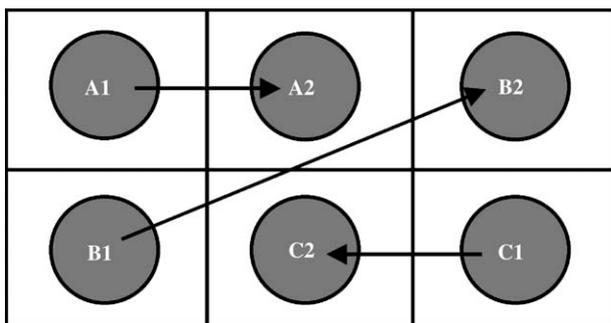

Figure 1. An example of a possible spatial sequence in Experiment 1. Red circles appeared one at a time during the familiarization and test trials, but are shown altogether here for the purpose of illustrating the transitional probabilities.

et al. (2002) study, the transitional probability within pairs was 1.0 and between pairs was 0.33. For each infant the location pairings were randomized by the computer. The initial stimulus location of a location pair always predicted the next stimulus location, and the next stimulus location after a pair was constrained to be the initial location of one of the three allowable pairs. For an individual infant, the location pairs were always the same, but the order of the pairs within the sequence was random. The infants viewed this sequence until habituation of looking occurred or 12 trials had elapsed. The habituation criterion was defined as a decline in looking times across a block of four trials adding up to less than 50% of looking times during the first four trials.

After habituation, infants viewed six test displays alternating between familiar sequences (composed of the same three location pairs) and novel sequences (produced by randomly ordering the same locations). In the novel sequences, the single constraint on stimulus order was that the red circle never appeared in the same location two times in a row. The only difference between familiar and novel sequences was the transitional probabilities between the locations. In the novel sequences there were no predictable location changes. This ensured that any looking time difference observed would necessarily be related to the statistical structure of the sequence. Ordering of test trials was counterbalanced across infants so that half the infants saw a familiar trial first and half the infants saw a novel trial first.

Results

Habituation Looking Times

Eleven infants (two 8-month-olds and nine 11-month-olds) did not habituate. There were no reliable age differences in total looking time during the habituation phase of the experiment (*M* looking time for 8-month-olds = 163.67 s, *SD* = 94.72; *M* looking time for 11-month-olds = 179.32 s, *SD* = 122.48), $t(30) = .40$, *ns*, nor were there differences in test display looking times between infants who habituated and those who did not, in combined data, and within each age group (all $ts < .93$, *ns*).

Test Sequence Looking Times

Infants in the 11-month-old age group, but not in the 8-month-old age group looked longer at the novel sequence, relative to the familiar sequence (see Figure 2). Thirteen of the sixteen 11-month-olds showed this preference (Wilcoxon matched pairs test $z = 2.84$,
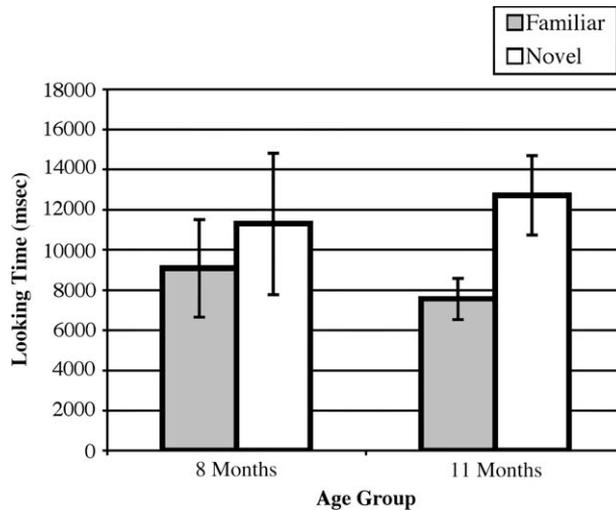
Figure 2. Experiment 1: Mean looking time (in milliseconds), by age group, to test trials.

$p < .01$), but only nine of the sixteen 8-month-olds exhibited this preference ($z = 1.09$, *ns*). A Mann Whitney *U*-test showed that these preferences were significantly different from each other ($z = 2.26$, $p < .05$).

These conclusions were confirmed with parametric analyses. Looking time data in some cells were positively skewed (which is often the case in visual habituation procedures with young infants), and all data in the three experiments in this report were therefore log-transformed prior to analysis (data shown in the figures are based on raw scores). Preliminary analyses did not reveal any reliable sex differences in performance in any of the experiments, nor any pertinent effects of trial block, so these variables were removed from subsequent analyses. A 2 (age group: 8 or 11 months) × 2 (order: familiar or novel sequence seen first after habituation) × 2 (test display: familiar or novel sequence) mixed ANOVA yielded a significant main effect of test display, $F(1, 28) = 11.16$, $p < .01$, the result of longer looking overall at the novel sequence relative to the familiar sequence. There was also a significant interaction between age group and test display, $F(1, 28) = 5.10$, $p < .05$, and no other significant main effects or interactions ($Fs < 1.83$, *ns*). Post hoc comparisons (simple effects tests) revealed a reliable preference for the novel sequence by the 11-month-olds, $F(1, 28) = 15.67$, $p < .001$, but not the 8-month-olds, $F(1, 28) = .59$, *ns*.

Discussion

Only the infants in the 11-month-old age group showed longer looking time to the novel sequence of locations; 8-month-olds did not look longer at either test trial sequence. Given that the only difference between the test trials was the pattern of locations of the identical red circles, it is reasonable to assume that only the 11-month-olds encoded the statistical pattern of locations during the habituation phase of the experiment.

Infants as young as 2 months of age can extract simple statistical associations in a temporal sequence (Kirkham et al., 2002), and so it seems surprising that 8-month-olds cannot extract the same pairwise statistics in a spatiotemporal pattern. But there is an important difference between the two paradigms. In the present experiment, the only available cue to predict the second location in a pair was the previous location, because the visual stimuli were all identical red circles. In contrast, in the Kirkham et al. experiment the stimuli differed in both shape and color (e.g., a turquoise square predicted a purple diamond; a yellow circle predicted a red octagon; a blue cross predicted a green triangle). The infants in that study may have encoded a pattern of colors, a pattern of shapes, or both. Perhaps the younger infants in the present experiment needed both kinds of featural cue in order to discover the pattern, if cue redundancy supported learning of the sequence (cf. Christiansen, Allen, & Seidenberg, 1998).

With this possibility in mind, we designed a second experiment that combined color and shape cues with location, presenting the color/shape pattern of Kirkham et al. (2002) in the location patterns of Experiment 1. Because our interest is in spatiotemporal encoding, color/shape patterns were maintained whereas the spatiotemporal patterns were violated during the novel test trials. Our hypothesis was that multiple feature cues would strengthen the representation of a spatiotemporal pattern and facilitate its acquisition by infants younger than 11 months.

Experiment 2

Method

Participants

Thirty-two full-term infants (13 girls) composed the final sample: sixteen 5-month-olds (*M* age = 5.2 months, *SD* = 0.3 months), and sixteen 8-month-olds (*M* age = 8.4 months, *SD* = 0.4 months). An additional 9 infants were observed but not included in the analyses due to fussiness (2) or procedural error or equipment failure (7). The infants were recruited in the same manner as infants in Experiment 1.

Apparatus, Stimuli, and Procedure

The testing apparatus and procedure were the same as in Experiment 1. Stimuli consisted of a 3 × 2 grid, defined by white lines on a black background, and six colored shapes (turquoise square, blue cross, yellow circle, pink diamond, green triangle, and red octagon) appearing sequentially. In each square of the grid, a colored shape was presented for 1 s and loomed from 4 cm to 24 cm in height (visual angle: 2.4° to 14.6°). The colored shapes appeared in a continuous stream of randomly ordered location pairs. Individual stimulus elements always appeared in the same location during habituation, resulting in two kinds of simultaneous pairing: pairing by color and shape (as in Kirkham et al., 2002) and pairing by location (as in Experiment 1; see Figure 3). The color/shape/location conjunctions were randomized by the computer for each infant. For an individual infant, the color/shape/location pairs were always the same, but the order of the pairs within the sequence was random.

As in Experiment 1, infants were first habituated and then viewed six test displays alternating between familiar sequences, composed of the same three color/shape/location pairs, and novel sequences, produced by randomly ordering of locations. In novel-sequence test trials, as in the familiar sequences, the color/shape pairings were maintained. Importantly, the only change in the novel test trials was the location of the shapes. Because the color-shape pairings did not change across test trials, discrimination of test sequences required sensitivity to the location changes. The single constraint on stimulus order in novel sequences was that shapes did not occur in the same location twice in a row.
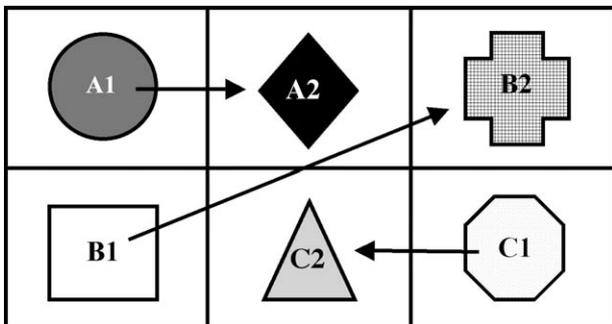
Results

Habituation Looking Times

Eight infants (three 5-month-olds and five 8-month-olds) did not habituate. There were no reliable age differences in total looking time during the habituation phase of the experiment (M looking time for 5-month-olds = 163.80 s, SD = 104.39; M looking time for 8-month-olds = 210.81 s, SD = 104.85), $t(30)$ = 1.27, ns, nor were there differences in test display looking times between infants who habituated and those who did not, in combined data, and within each age group (all ts < 1.7, ns).

Test Sequence Looking Times

Infants in the 8-month-old age group, but not in the 5-month-olds, looked longer at the novel sequence relative to the familiar sequence (see Figure 4). Thirteen of the sixteen 8-month-olds showed this preference (Wilcoxon matched pairs test z = 2.37, p < .05), as opposed to only seven of the sixteen 5-month-olds (Wilcoxon matched pairs test z = .53, ns). These preferences were significantly different from each other (Wilcoxon matched pairs test z = 2.37, p < .05). A 2 (age group: 5 or 8 months) × 2 (order) × 2 (test display) mixed ANOVA yielded a significant main effect of age group, $F(1, 28)$ = 4.23, p < .05, resulting from longer looking overall by the 8-month-olds, and a significant interaction between age group and test display, $F(1, 28)$ = 5.41, p < .05. There were no other reliable effects (Fs < 3.51, ns). Post hoc comparisons (simple effects tests) revealed a reliable preference for the novel sequence by the

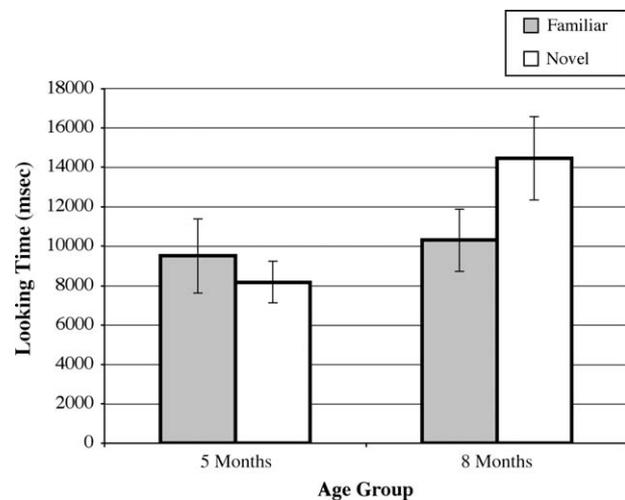Figure 3. Example of the spatial structure of the stimuli in Experiment 2.

Figure 4. Experiment 2: Mean looking time (msec), by age group, to test trials.

8-month-olds, $F(1, 28) = 8.42$, $p < .01$, but not the 5-month-olds, $F(1, 28) = .15$, *ns*.

Discussion

As predicted by the multiple cue hypothesis, the 8-month-olds we observed discriminated between the two test trials when the familiarization phase included color and shape cues in addition to the location pattern. It is worth stressing that during the test phase, the color/shape patterns were never violated; only spatiotemporal predictability was disrupted. Apparently the 8-month-olds were able to use the redundancy across color, shape, and location to discriminate familiar from novel spatiotemporal patterns. It is not clear whether successful performance was based on a true sensitivity to location statistics or on detection of a violation of expected location of individual stimulus elements (or both). Either way, the results of Experiment 2 suggest a burgeoning ability to keep track of stimuli over numerous locations by 8 months of age, when stimuli appear in a context of dynamic and complex multi-element scenes.

In contrast to 8-month-olds, 5-month-olds did not appear to recognize violations of the spatiotemporal pattern seen during habituation despite the availability of multiple cues, although 5-month-olds have been shown to be sensitive to statistical information from color and shape when location is held constant (Kirkham et al., 2002). This age difference suggests that spatiotemporal sequence learning may be limited in young infants by difficulty in accessing spatiotemporal information. Adults represent locations using a variety of spatial reference frames (e.g., egocentric and allocentric), and coordinate those reference frames across modalities (Stein, 1992). Gilmore and Johnson (1997, 1998) have proposed, in contrast, that infants younger than 6 months represent location only relative to their current eye position. Locations encoded within this *retinocentric* reference frame become invalid once a saccade is made, which would make spatiotemporal sequence learning difficult, if not impossible. An *egocentric* reference frame combines retinal and eye position information to represent the location of targets relative to the body, and is the bare minimum required in order to keep track of spatial locations of objects in dynamic events.

Gilmore and Johnson (1997) used a "double-step" saccade paradigm, in which two visual targets were flashed briefly in a dark visual field, to document the emergence of egocentric reference frames in infants. The temporal and spatial parameters of target appearance were more complex than the simple two- and three-location sequences presented in the visual expectation paradigm described in the Introduction (Haith, 1993). Targets were flashed one after the other with an interval short enough that the second target had already disappeared before the saccade to the first target had finished, so that a retinocentric reference frame would result in an incorrect saccade to the location of the second target. Gilmore and Johnson found that a progression from predominantly retinocentric reference frames to predominantly egocentric reference frames takes place between 4 and 6 months of age.

Analysis of our task at this level provides a ready explanation for the failure of the 5-month-olds to respond to location statistics in Experiment 2, and leads to a prediction. If 5-month-olds cannot recruit an appropriate reference frame to predict the location in which a stimulus element will appear in a complex sequence, then spatial position is actually irrelevant to spatiotemporal sequence learning in young infants. This prediction was tested in Experiment 3. Five-month-olds were habituated to the color/shape/location display employed in Experiment 2, and then viewed two test trial sequences. The familiar sequence was identical to the habituation sequence, as in the design of the first two experiments. The novel sequence, in contrast, preserved the location pairings and randomized the color/shape pairings. If the 5-month-olds' failure to discriminate familiar from novel sequences in Experiment 2 was centered specifically in the spatiotemporal aspect of the task, they should discriminate between the two test trials in the present experiment. This is because location statistics are hypothesized to provide no contribution to event predictability for this age group, whereas color and shape are cues to which 5-month-olds are known to respond in a visual statistical learning paradigm (Kirkham et al., 2002). In the present case, therefore, violation of location statistics should have no negative impact on performance. This possibility was examined in Experiment 3.

To be confident that the results we have obtained are informative with respect to infants' (in)sensitivities to statistically probabilistic spatiotemporal patterns, it is necessary to rule out another alternative hypothesis. It is possible that our experiments tap into infants' ability to pick up on shape-location correlations, rather than sensitivity to a higher order probabilistic pattern of location change. Perhaps the 8-month-olds showed a novelty preference because they noticed a violation of the shape-location pairings in the novel test trials (e.g., the blue star was no longer in the top right position), and not because they noticed

a higher-order pattern of location changes. This alternative explanation was investigated in Experiment 4.

Experiment 3

Method

Participants

Sixteen full-term 5-month-old infants (4 girls) composed the final sample (*M* age = 5.3 months, *SD* = .6 months). One additional infant was observed but not included in the analyses due to persistent inattention to the displays. The infants were recruited in the same manner as infants in Experiments 1 and 2.

Apparatus, Stimuli, and Procedure

The testing apparatus and procedure were the same as in the first two experiments. The habituation sequence was the same as in Experiment 2, as well as the familiar test sequence (the statistics of both spatial location and color/shape of the stimulus elements were preserved). In the novel test sequence, location statistics were maintained but the color/shape pairing was randomized, with a single constraint on stimulus order: Shapes did not occur in the same location twice in a row. That is, stimulus elements appeared in location-based pairs, the second member of which was predictable from the first, but which color/shape combinations composed the members was random.

Results

Habituation Looking Times

Three infants did not habituate. There were no differences in test display looking times between infants who habituated and those who did not (all *t*s < .45, *ns*).

Test Sequence Looking Times

Thirteen of the sixteen 5-month-olds looked longer at the novel test display (Wilcoxon matched pairs test $z = 3.00$, $p < .01$). A 2 (order) × 2 (test sequence) mixed ANOVA yielded a significant main effect of test sequence, $F(1, 14) = 31.52$, $p < .0001$: Looking time was longer to the novel test sequence than to the familiar sequence (see Figure 5). The effect was more pronounced in the group that viewed the novel sequence first (order × test display interaction $F(1, 14) = 8.36$, $p < .05$). Post hoc comparisons (simple
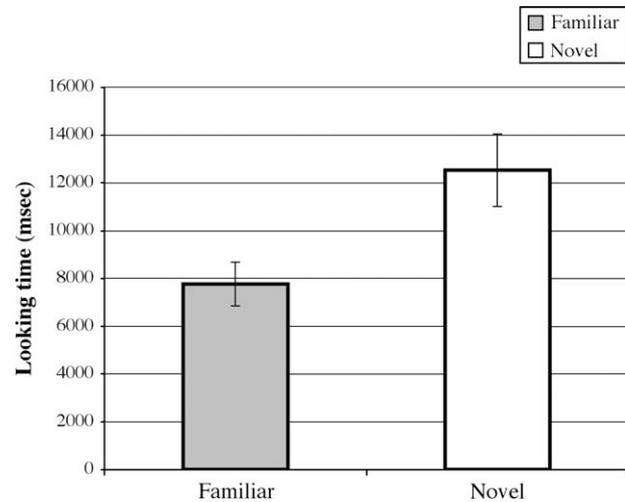

Figure 5. Experiment 3: Mean looking time (in milliseconds) to test trials.

effects tests) revealed that the preference for the novel sequence was reliable in infants who viewed it first, $F(1, 14) = 36.18$, $p < .001$, and marginally reliable in infants who viewed it second, $F(1, 14) = 3.71$, $p = .07$. However, there were no reliable differences in preference between the two groups for either the novel or the familiar sequence ($Fs < 1.57$, *ns*). The main effect of order was not statistically significant, $F < .07$, *ns*.

Discussion

In a replication of Kirkham et al. (2002), and as predicted, 5-month-olds demonstrated sensitivity to the statistical structure of the color/shape pairings and ignored violations of location statistics. Their failure to discriminate familiar and novel spatiotemporal patterns in Experiment 2, therefore, may have been due to an immature spatial reference system incapable of processing location pairings, a spatial reference system that is sufficiently functional by 8 months to support computation of shape/color/location statistics.

Experiment 4

Previously, we suggested that the 8-month-olds' performance in Experiment 2 demonstrated sensitivity to the probabilistic pattern of location changes. It is possible, however, that the infants responded on the basis of specific shape-location pairings, and showed a novelty preference when those pairings were violated. For example, say in Experiment 2 an infant was habituated to a blue cross in the top left followed by a red square in the bottom right. During the test phase

the infant may have viewed a blue cross in the top left followed by a red square in the bottom *left*. The infant may have looked longer to this display either because she learned a spatiotemporal pattern that was violated, or because she had never seen a red square in that location. To distinguish between these two possibilities, we designed an experiment in which we recorded infants' oculomotor latencies as they viewed spatiotemporal sequences. We reasoned that detection of spatiotemporal patterns would facilitate gaze shifts to predictable locations. In the structured sequence pairs employed in Experiments 1–3, the second shape was always predicted by the first. Therefore, if infants keep track of spatiotemporal patterns, they should show shorter saccade latencies to the location of the second shape in a pair.

In Experiment 4, 8-month-old infants were presented with the structured sequence used in Experiment 2. After they had seen 24 pairs, we began to record their eye movements. The infants saw the same spatiotemporal structure throughout the whole experiment (there was no violation of shape-location pairings). Therefore any differences in fixation latencies to the predictable and unpredictable events must be due to online learning of spatiotemporal structure, rather than other differences between responses to structured and unstructured sequences. Note that we do not expect infants to make anticipatory eye movements to the upcoming locations (as shown by Johnson et al., 2003). In the current display, stimuli loom one after another without pause, and so an anticipatory eye movement would require the infant to look away from a very salient event toward a blank region of the screen. Rather, we predicted that infants would have some expectation of the upcoming location while looking at the first member of a stimuli pair, and that this spatiotemporal knowledge would show itself in decreased fixation latencies to the second.

Method

Participants

Sixteen full-term 8-month-old infants (8 girls) composed the final sample (*M* age = 8.2 months, *SD* = 0.3 months). Twelve additional infants were observed but not included in the analyses due to persistent inattention to the displays or problems in calibrating the eye tracker to their gaze. The infants were recruited from a database of parents who had previously agreed to participate in research. They received and infant T-shirt and/or toy as a parting gift.

Apparatus, Stimuli, and Procedure

The stimuli were presented with a Macintosh G5 computer and 152-cm rear projection screen. Infants sat on their caregiver's lap 180 cm away from the screen. An Applied Science Laboratories Model 504 corneal reflection eye tracking system was used to collect eye movement data as infants were shown the stimulus displays. A remote pupil camera with a pan/tilt base was placed on the table below the projection screen. The stimuli viewed by the infant were imported directly into the eye tracker from the Macintosh. The eye tracker also fed a signal into a mini digital video recorder in the form of crosshairs superimposed on the stimulus. Stimulus dimensions were the same as in Experiments 1–3.

Each session required two experimenters, an "eye-tracking experimenter" (ET experimenter) and a "stimulus presenter," both of whom sat behind the rear projection screen out of sight of the infant. The ET experimenter worked the eye-tracking system, watching an image of the infant's pupil, the infant's point of gaze (POG), and the stimulus on a split screen monitor. If and when the infant's eye moved away from the presented stimuli, the ET experimenter used a remote control to redirect the pupil camera. The stimulus presenter controlled presentation of the calibration and experimental stimuli based on when the infant was attending to the projection screen.

Before the stimulus presentation, the room lights were turned off and the infant was shown a cartoon clip to engage attention. During this time, the pupil camera was directed toward the infant's left eye. After the infant's left eye was in view, the ET experimenter placed the eye-tracking computer in "automatic" mode, during which the camera remained directed at the pupil despite small displacements of the infant's head (via an algorithm built into the eye tracker). If the infant moved his or her head more quickly than the camera could follow, such that the pupil was lost from view, the ET experimenter changed the computer to manual mode, located the pupil in the camera, and automatic control was resumed.

Following acquisition of the pupil image, as the infant watched the cartoon, adjustments were made on the eye tracker to maximize robustness of the POG. This varied somewhat from infant to infant with respect to reflectance of infrared and visible illumination (corneal and pupil reflection, respectively). The infant was then shown a series of looming cartoon movies that made fun noises (e.g., cartoon-style musical noises) to attract her attention and calibrate the POG. The eye tracker was calibrated on each

infant's left eye using a two-point calibration routine (i.e., the POG for upper left and lower right locations were used; other locations were interpolated by the computer) and then checked against two different points at the end of the calibration. If the calibration check was successful, the experiment began.

The stimulus sequence was identical to that described in Experiment 2.

Each infant was first familiarized with a sequence of 48 looming shapes. This sequence contained eight exposures of each of the three pairs of color/shape/ location conjunctions. Pilot data suggested that this 48-s exposure was long enough perhaps for the infants to learn something of the statistical structure of the sequence, yet short enough to allow further testing before they grew restless. The looming sequence continued seamlessly after the familiarization period, and we recorded infants' eye movements until they became disinterested.

Results

Infants watched an average of 72 looming shapes ($SD = 9.6$) following the familiarization trials. We compared saccade latencies to the first member of the pair, the location of which was unpredictable, to saccade latencies to the second member of the pair, which was predictable. Latencies to the predictable locations ($M = 569$ ms, $SD = 111$) were faster than those to the unpredictable locations ($M = 645$ ms, $SD = 82$), $t(15) = 2.27$, $p < .05$. Eleven of the 16 infants produced faster saccades to the predictable location (Wilcoxon matched pairs test $z = 1.82$, $p < .05$).

What were these infants learning? One possibility was that the infants were only sensitive to such spatiotemporal contingencies over a short time span, when the same pair of shapes was repeated in the sequence. When such instances of repetition were excluded from the analysis, however, the latency to predictable trials remained over 70 ms faster than unpredictable trials, and this difference was significant, $t(15) = 2.13$, $p < .05$. Another possibility was that infants had just learned a single sequence pair, and this limited spatiotemporal learning accounted for the overall differences we observed. We calculated that 59% of the pairs had shorter latencies for the predictable member, and hence showed evidence of learning ($p < .05$ with a sign test).

Discussion

After watching a probabilistic sequence for a short period of time, 8-month-olds responded differently to unpredictable and predicable events: Saccade latencies were faster to a new shape when its location had been predicted by the previous shape. This provides evidence that the infants responded to the higher order pattern of location changes, not simply the novelty of seeing the particular shapes occur in new locations.

General Discussion

These four experiments present evidence concerning a fundamental cognitive skill in infancy: the ability to learn probabilistic event sequences across space and time. We also obtained evidence of important developmental limitations in learning: Only the oldest infants we observed (11-month-olds) responded solely on the basis of location statistics, showing a posthabituation novelty preference to a display in which the positions of stimulus elements were randomly placed. The youngest infants we tested (5-month-olds), in contrast, appeared largely insensitive to location statistics, although they were able to detect probabilistic sequences based on a combination of color and shape. Infants at an intermediate age (8 months) provided evidence of learning location statistics only when color and shape contributed additional (redundant) cues for the spatiotemporal sequence. This latter finding was supported by an on-line eye-tracking experiment and accords with a view stressing an important role for integration of multiple information sources in sequence learning. In the present case, these information sources are multiple visual features. A mechanism capable of learning simple temporal associations, therefore, appears to be available even to very young infants (Kirkham et al., 2002), but extraction of purely location-based associations seems to be more complex, perhaps requiring the maturation of dedicated spatial representation systems.

We propose a developmental theory of multiple cue integration, which highlights acquisition of spatiotemporal information within a statistical learning framework. In the adult vision literature, theories of multiple cue integration, or weighted sum models, have been offered as explanations for how observers combine redundant stimulus property cues (e.g., Kinchla, 1977) and multiple depth cues (e.g., Landy, Maloney, Johnston, & Young, 1995) while viewing visual stimuli. Investigations of depth perception, for example, have determined that the wide range of available cues (such as object motion, disparity, and texture gradients) are integrated effortlessly and quickly to produce coherent visual percepts of objects,

surface layout, and scenes (e.g., Jacobs, 2002; Mayhew & Frisby, 1980; Sperling, Landy, Dosher, & Perkins, 1989; Todd & Akerstrom, 1987). Recent findings from infants converge to reveal early development of sensitivities to some kinds of spatiotemporal event structure (Canfield et al., 1997; Haith, 1993; Kirkham et al., 2002; Saffran et al., 1996; Johnson et al., 2003; Smith, Loboschefski, Davidson, & Dixon, 1997). A question that surfaces when considering these sensitivities concerns the role of the precise spatiotemporal context when learning about structure in the visual environment. As adult observers inspect a series of visual scenes, the spatiotemporal information that is available across repeated exposures composes a larger context within which the observers have continually greater access to available structure (Chun & Jiang, 1998; Olson & Chun, 2001). It may be that rich scenery, with many cues available to highlight environmental structure, is processed more efficiently by infants than a scene with sparser features. If this is so, then this implies that nascent spatiotemporal sensitivities might require more contextual support in order to glean structure from the environment. The results from the present experiments are consistent with this suggestion, with 11-month-olds outperforming the younger infants in a comparatively barren context.

Nevertheless, even young infants integrate multiple perceptual cues to organize the visual world. Johnson et al. (2003) found that 4-month-olds' perception of object trajectories is improved dramatically when cues of temporal and spatial predictability were provided. In addition, when viewing partially occluded objects, 2- to 4-month-olds integrate available cues such as edge alignment, synchronous motion, and depth to support perception of unity (Johnson, 1997, 2004; Smith, Johnson, & Spelke, 2003). That is, young infants exploit the regularities in their environment to complete occluded objects, and to predict their trajectories.

Concurrently, researchers investigating language acquisition have provided similar kinds of multiple cue hypotheses as possible answers to the speech stream segmentation problem (e.g., Christiansen et al., 1998; Johnson & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999). Christiansen and colleagues have proposed that the integration of multiple acoustic cues can help infants parse an ostensibly continuous stream of sound into word segments. The combination of these cues provides evidence about aspects of linguistic structure that is not available from any single source of information, a notion that receives added support from connectionist modeling: Providing a simple recurrent network with a combination of cues supported correct word segmentation, but none of the cues in isolation was effective in this task (Christiansen & Dale, 2001).

Given the probabilistic nature of the environment across all perceptual domains, it is not surprising that there might be commonalities between language learning and visual learning. Christiansen et al. (1998) speculated that associations between visual and auditory stimuli may provide possible cues for more sophisticated kinds of segmentation relative to parsing the speech stream. For example, if an infant notices the temporal correlation between the speech sound "ball" with the actual physical object when playing with a parent, he or she could use that information to isolate "ball" from the speech stream during subsequent encounters.

It seems natural, then, that infants would take advantage of the regularities in the environment, and use them to support understanding of a dynamic spatiotemporal situation. The somewhat lengthy developmental trajectory in perceiving spatiotemporal structure revealed by our experiments is striking in this context. We speculate that the richness of environmental stimulation may provide a significant challenge to developing perceptual systems, whose task is to discern statistical and abstract patterns across a wide range of input. When viewing relatively simple displays or events typical of much research on perceptual development, young infants often are able to detect information relevant to task success with little difficulty. In the present experiments, in contrast, spatiotemporal statistics in isolation were insufficient to guide perception of event structure until 11 months. This seems surprising given the importance of spatial location to visual guidance of behavior, and suggests that development of spatial cognition may be more complex than previously envisioned.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Aslin, R. N., Slemmer, J. A., Kirkham, N. Z., & Johnson, S. P. (2001, April). *Statistical learning of visual shape sequences.* Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, MN

Baldwin, D. A., Baird, J. A., Saylor, M. M. & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708–717.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the Visual Expectation Paradigm. *Monographs of the Society for Research in Child Development, 62*, whole no. 250.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13,* 221–268.

Christiansen, M. H., & Dale, R. A. C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220–225). Mahwah, NJ: Erlbaum.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4,* 170–178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2,* 406–416.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17–30.

Conway, C., & Christiansen, M. H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31,* 24–39.

Curran, T., & Keele, S. W. (1993). Attentional and non-attentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 189–202.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499–504.

Fiser, J. & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458–467.

Fiser, J., & Aslin, R. N. (2003). Statistical learning of new visual feature combinations by infants. *Proceedings of The National Academy of Sciences, USA, 99,* 15822–15826.

Gilmore, R. O., & Johnson, M. H. (1997). Egocentric action in early infancy: Spatial frames of reference for saccades. *Psychological Science, 8,* 224–230.

Gilmore, R. O., & Johnson, M. H. (1998). Learning what is where: Oculomotor contributions to the development of spatial cognition. In F. Simion & G. Butterworth (Eds.), *The development of sensory, motor, and cognitive capacities in early infancy: From perception to cognition* (pp. 25–47). East Sussex, UK: Psychology Press.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Haith, M. M. (1993). Future-oriented processes in infancy: The case of visual expectations. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 235–264). Hillsdale, NJ: Erlbaum.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78,* B53–B64.

Heyes, C. M., & Foster, C. L. (2002). Motor learning by observation: Evidence from a serial reaction time task. *The Quarterly Journal of Experimental Psychology, 55A,* 593–607.

Howard, J. H., Mutter, S. A., & Howard, D. V. (1993). Serial pattern learning by event observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1029–1039.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6,* 345–350.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.

Johnson, S. P. (1997). Young infants' perception of object unity: Implications for development of attentional and cognitive skills. *Current Directions in Psychological Science, 6,* 5–11.

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

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences, USA, 100,* 10568–10573.

Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. *Perception and Psychophysics, 22,* 19–30.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy. *Cognition, 83,* B35–B42.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modelling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465–494.

Mayhew, J. E., & Frisby, J. P. (1980). The computation of binocular edges. *Perception, 1,* 69–86.

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 350–364.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1–32.

Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1299–1313.

Richardson, D. C. & Kirkham, N. Z. (2004). Multi-modal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133,* 46–62.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Smith, P. H., Loboschefski, T. W., Davidson, B. K., & Dixon, W. E. (1997). Scripts and checkerboards: The influence of ordered visual information on remembering locations in infancy. *Infant Behavior & Development, 20,* 549–552.

Smith, W. C., Johnson, S. P., & Spelke, E. S. (2003). Motion and edge sensitivity in perception of object unity. *Cognitive Psychology, 46,* 31–64.

Sperling, G., Landy, M. S., Dosher, B. A., & Perkins, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 826–840.

Stein, J. F. (1992). The representation of egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences, 15,* 691–700.

Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 242–255.

Wentworth, N., Haith, M. M., & Hood, R. (2002). Spatio-temporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy, 3,* 303–321.