The Goldilocks Effect: Infants' preference for stimuli that are neither too predictable nor too surprising

Celeste Kidd (ckidd@bcs.rochester.edu)
Brain & Cognitive Sciences, Meliora Hall
Rochester, NY 14627 USA

Steven T. Piantadosi (piantado@mit.edu)
Brain & Cognitive Sciences, 43 Vassar Street
Cambridge, MA 02139 USA

Richard N. Aslin (aslin@cvs.rochester.edu)
Brain & Cognitive Sciences, Meliora Hall
Rochester, NY 14627 USA

Abstract

Even before birth, infants attend to the statistical properties of their sensory environments to learn about events in world. Tracking these statistics is crucial to mastery of visual, social, linguistic, and cognitive tasks. However, the degree to which their sampling follows prescriptions of rational statistical inference is unclear. Do infants' attentional preferences reflect efficient information gathering? We investigated using an ideal observer model (a Markov Dirichlet-multinomial). We predicted infants' attention to sequential events would be moderated by information content. We tested infants (7-8 months) with 32 unique event sequences (objects popping out of boxes) on a Tobii eye-tracker. Each sequence continued until look-away. Controlling for other variables, we found infants were significantly more likely to look away at either highly informative or uninformative events according to the model. This suggests infants allocate visual attention to maintain intermediate rates of information processing, avoiding committing cognitive resources to either overly predictable or surprising events. This "Goldilocks effect" may reflect a general strategy for efficient learning from environmental statistics.

Keywords: Statistical learning; statistical inference; idealized learner; infant gaze behavior; infant methods; infant eye-tracking; Bayesian modeling; information theory; infant visual attention.

Introduction

Infants have a lot to learn in the first few years of life, and a limited set of resources with which to do it. The world is brimming with potential sources of information, but where among this spatiotemporal array of events should infants begin their learning? From birth, infants survey their sensory environments, sampling the visual data that surrounds them at the incredibly rapid pace of two or more fixations a second during 90% of their waking hours (Haith, 1980). This process, of surveying and sampling, provides infants with rich information from which they can start to learn about the world.

Previous empirical work has demonstrated that infants are able use the statistical properties of their environment in a diverse array of learning tasks pertaining to sounds, words, people, shapes, and objects (Fiser & Aslin, 2002; Maye, Weiss, & Aslin, 2008; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Yu & Ballard, 2007)[1]. In many complex cognitive systems (e.g., object recognition, language), infants obtain the representations of higher-level structures by tracking the low-level statistical coocurrences. For example, newborn infants must track the distribution of acoustical properties of speech sounds in their target language in order to infer its phonological categories (White, Peperkamp, Kirk, & Morgan, 2008). Researchers of visual, conceptual, and social learning find similar patterns. Recent technologies (e.g. eye-tracking, brain imaging techniques) have elucidated many of the mechanisms infants employ in building high-level structures from low-level environmental statistics. Though the topic has been of great interest to researchers, the mechanisms and representations infants employ during the process of collecting environmental statistics are still not well understood.

Amid the unbinned and unsorted masses of sensory data available in the world, an undirected search would be inefficient. Infants have too many things to do—motor actions to program, words to learn, categories to form—to waste time. How then should the infant allocate her visual attention?

Several researchers attempted to unify this work by identifying an overarching stimulus feature that could generally account for all of infants' preferences for various stimulus properties. Sokolov (1960) postulated that the primary driver of infants' attention is stimulus novelty. Consistent with this theory, infants commonly prefer the novel stimuli in preferential looking/listening tasks such as those use in the Fantz paradigm (Fantz, 1964), high-amplitude sucking procedure (Siqueland & DeLucia, 1969), and head-turn preference procedure (Kemler Nelson et al., 1995) . The novelty account is also consistent with habituation behavior, during which infants' attention to recurring stimuli decreases over time. However, the novelty hypothesis does not account for infants' *familiarity* preferences in many preferential looking and listening studies. Notable examples include infants' affinity for their native languages and for faces, especially those of their mothers.

Roder and others attempted to reconcile infants' preference patterns by relating preference and processing load (Hunter & Ames, 1988; Roder, Bushnell, & Sasseville, 2000; Wagner & Sakovits, 1986). Roder suggested that the process of memory formation was responsible for preference. Under this theory, infants would be expected to exhibit a familiarity preference early in processing as they form a memory of

[1]The literature detailing these statistical learning abilities is so large, in fact, that if it were printed and stacked in a pile, it would be more than 140 infants tall (based on 161,000 unique articles cataloged by Google Scholar at time of publication and 26-inch mean height of 8-month-olds).

the stimulus, and a novelty preference later after memory formation was complete. While this account correctly predict age and experience-related shifts in visual preference, it does not on its own account for all types of visual preferences. It does not, for example, make clear predictions of why infants would prefer one novel object to another entirely novel object, since infants would not possess a memory for either item. Kinney and Kagan similarly suggested a processing-based account of preference. Their *moderate discrepency hypothesis* states that infants will preferentially attend to stimuli that are "optimally discrepant", meaning those that are most distinct from the representations they already possess. Like Roder's memory-based account, Kinney and Kagan's theory relates stimulus preferences to stimulus representations established by past experiences (Kinney & Kagan, 1976). The moderate discrepancy hypothesis has the added advantage of accounting for preferences among completely novel stimuli, since it defines the representation formation process as pertaining to the infants' existing representations. Unfortunately, attempts to test this theory behaviorally were hindered by methodological difficulties. First, researchers had no direct access for determining the type, quantity, and nature of infants existing representations, which are crucial to the theory for generating testable predictions. Second, manipulating the identity of stimulus items to test for visual preferences forced researchers to rely on *qualitative* judgments of discrepancy rather than a quantitative metric.

Yet another account, Dember and Earl's *theory of choice/preference*, suggests that *stimulus complexity* drives looking behavior. Dember and Earl posited that every stimulus contains a certain "complexity value, and that each individual[2] has a certain preferred complexity level it seeks to maintain (Dember & Earl, 1957). In this context, complexity can be thought of as information content. The theory predicts that individuals will seek out stimuli containing the ideal level of complexity with respect to their own preferred complexity rates. The amount of information an individual will derive from a stimulus decreases as experience accumulates. Thus, like other processing-based accounts, this complexity-driven one can theoretically predict age and experience-related shifts in visual preference. Berlyne noted that a complexity-driven preference would be an optimal strategy for learning (Berlyne, 1960). It provides a rational solution to the infant learner's problem of deciding where best to allocate attention in the world. As with the attempts to test memory-based theories of attention, the collection of empirical evidence for this theory was hindered by the use of stimuli varied along *qualitative* dimensions rather than quantitative metric.

All prior models of infant visual attention, using the standard 2-second look-away criterion, have been based on hypothetical underlying processes such as information complexity, processing speed, and stimulus salience. Unfortunately, none of these underlying processes were validated by an independent assessment. As a result, the precise way in which these processes were combined could not be estimated, except by observing the outcome of their integrated effect on gaze durations. Here we seek to provide a quantitative model of visual attention to sequential events by systematically manipulating information complexity while holding processing speed and stimulus salience constant.

We used an idealized statistical model—a Dirichlet-Multinomial Markov model—to predict infant looking behavior to a display of sequential events. Our results suggest that infants' behavioral responses to a stimulus are influenced by its information content. Further, we find evidence that infants allocate their attention to maintain a certain information rate under a statistical model of the world. We present this as evidence that infants use rational statistical inference in understanding the world and deciding where to allocate attention and other cognitive resources.

Infant Behavioral Data

Participants

Twenty-five infants (mean = 7.9 months, range = 7.0 - 8.8) were tested. All infants were born full-term and had no known health conditions, hearing loss, or visual deficits, according to parental report. All participating infants completed the study.

Stimuli

We presented each infant with 32 unique event sequences, with the order of the sequences randomized across infants. The events each sequence consisted of were three unique objects that were animated to pop out from behind three occluding surfaces, which simulated an array of boxes. The sequences of object "pop ups" were chosen to vary in their information-theoretic properties (e.g., entropy, surprisal). Thus, some sequences were highly predictable (e.g., *AAAAAAAA*), and others were less predictable (e.g., *CAAABBCABAC*).

For each infant, the Matlab script generated an animated scene based on each of the 32 event sequences. Each event sequence was implemented by creating a scene consisting of three uniquely patterned and colored boxes, each concealing a unique familiar object (e.g., a cookie). The locations of the three boxes for a given sequence were chosen randomly but remained static throughout a scene. The box locations were randomly shuffled between event sequences, but no more than two boxes appeared on either half of the screen. Neither the patterns on the boxes nor the objects were repeated across event sequences so that each object-box pair was independent and unique.

The objects, boxes, and the order in which the 32 event sequences were presented were randomized across infants. The same 32 event sequences were presented to every infant. This design ensured that differences in looking time across event sequences were not driven by differences in scene items or

[2]Individuals referred to not only baby humans, but also adults and animals

presentation order. Each event in a sequence consisted of an object that popped out of a box (1 s), and then back into the box (1 s). The total duration of each event was 2 s, and events were presented sequentially with no overlap or delay.

Procedure

Each infant was seated on his or her parent's lap in front of a table-mounted Tobii 1750 eye-tracker. The infant was positioned such that his or her eyes were approximately 23 inches from the monitor, the recommended distance for accurate eye-tracking. At this viewing distance, the 17-inch LCD screen subtended 24 X 32 degrees of visual angle. Each of the 3 boxes was 2 X 2 inches. To prevent parental influence on the infant's behavior, the parent holding the infant was asked to wear headphones playing music, lower their eyes, and abstain from interacting with their infant throughout the experiment.

The experiment consisted of 32 trials, one for each event sequence. Each trial was preceded by an animation designed to attract the infant's attention to the center of the screen—a laughing and cooing baby. Once the infant looked at the attention-getter, an experimenter who was observing remotely pushed a button to start the trial.

For each trial, an animated scene depicting one of the event sequences was played. The animated sequence of events—objects popping out of boxes one at a time—continued until the infant looked away continuously for 1 sec, or until the sequence timed out at 60 sec. The 1-sec look-away criterion for trial termination was automatically determined by the Tobii eye-tracking software. If the infant looked continuously for the entire 60-sec sequence, the trial was automatically labeled as a "time out" and discarded before the analysis (3.5% of trials). If the trial was terminated before the infant actually looked away, the trial was labeled by an experimenter as a "false stop" and also discarded. False stops occurred as a result of the Tobii software being unable to detect the child's eyes continuously for 1 sec, usually due to the infant inadvertently blocking the his or her own eyes with head or arm movements (18.5% of trials).

Every infant saw all 32 event-sequence trials. The dependent measure for the subsequent computational modeling was the event at which the infant looked away in each trial; that is, at what point in the sequence did infants look away from the display for more than 1 consecutive second? We predicted that infants were more likely to look away during events that contained either too little or too much information for a particular infants' preferred information-intake rate. We predicted infants would be least likely to look away during events that were "just right"—those that were neither too predictable, nor too surprising. Our Ideal Observer Model was used to determine the amount of information for each event in the event sequences (i.e., which event contained more or less information). If infants' attention to a stimulus is governed by the amount of information it contains, we would expect infants' look-aways to be predictable given the model. We tested our hypothesis by comparing the model's predicted probabilities of an infant looking away for each event in the sequence to the infants' actual look-aways in test.

Ideal Observer Model

We used a Markov Dirichlet-multinomial model (MDM) to evaluate the relationship between the statistical properties of the event sequences and infants' attention to events in that sequence. The model allows us to test the best-fitting set of parameters for predicting from the event sequence whether the infant will continue looking or terminate a trial by looking away from the display. The MDM is a general-purpose statistical model that infers an underlying (multinomial) probability distribution on events, using the history of how many times each event has been observed. The MDM makes parametric assumptions about the form of the prior probability of an event and the likelihood of the event, and is often used in Bayesian statistics because it is computationally simple. Intuitively, infants observe how many times each event in the world occurs, and then use these event counts to infer an underlying probability distribution on events, just as readers extract an underlying word frequency distribution using a set of observations of individual words. An observer who sees only a single event happen would not likely infer that that single event is the only possible event (e.g, has probability 1.0). Instead, observers likely bring expectations to the task. In the version of a MDM used here, this prior expectation is parameterized by a single free parameter, α which controls the prior degree of belief that the distribution of events is uniform (e.g., that all unobserved events are equally likely). As α gets larger, the model has stronger prior beliefs that the distribution of events in the world is uniform; as $\alpha \to 0$, the model believes more strongly that the distribution is closer to empirically observed counts on events.

Formally, if there are three events, A, B, and C, which have been observed to occur c_A, c_B, and c_C times respectively, then the model assigns probability to a distribution on these three events proportional to

$$P(A)^{c_A+alpha-1}P(B)^{c_B+alpha-1}P(C)^{c_C+alpha-1}, \qquad (1)$$

where P is a hypothesized distribution on the events A, B, and C. That is, after observing each event occur some number of times, the infant may form a representation P, which gives the true underlying distribution of events. Every distribution can be "scored" according to Equation 1, allowing one to compute a distribution of beliefs about the state of the world according to the model. This simple model allows us to quantify an ideal observer's degree of belief that any given distribution on events is the true one. Importantly, because of the parametric form of the MDM, statistical measures such as the most likely true distribution of events, can be computed analytically.

We used two different forms of the MDM. In the first, the events A, B, and C correspond to events in the behavioral experiment (objects appearing from behind the occluding boxes). This model does not represent the transitions between events in the world; that is, the sequence *AAABBBCCC* would have the same expectation as *ABCABCABC*. In the second model, we treated the events *A*, *B*, and *C* as transitions (or bigrams): for each object, we created a separate MDM for events that happen next. This model represents three separate MDMs that capture the transitions between events.

Both of these forms of the MDM provide an estimate of what an ideal observer would infer about the structure of the world. However, a model of infant's beliefs alone is not sufficient to predict their behavior: what is needed additionally is a set of linking assumptions that relate beliefs to actions. Here, we assume that the infant's looking behavior is at least partially determined by the information-theoretic properties of the model. Specifically, we test whether the predictability of a stimulus according to an idealized learning model influences infants' looking behavior. Formally, we use the negative log probability of the current event according to the model, conditioned on observing all the previous events. As this negative log probability increases, the current event is more surprising: for instance, after seeing a long sequence of *A*s, a *B* would have a high negative log probability. Negative log probability is a convenient measure because it corresponds to the number of *bits* of information conveyed by the stimulus. Thus, negative log probability provides a measurement at each point in time of the unpredictability of an event, using a measure that is typically used as a measure of information content. Because of the form of the MDM, the model roughly predicts that events in the future will tend to occur with their already-observed probability. However, the model essentially adds a small amount of smoothing—parameterized by alpha—that prevents unseen events from having probability zero.

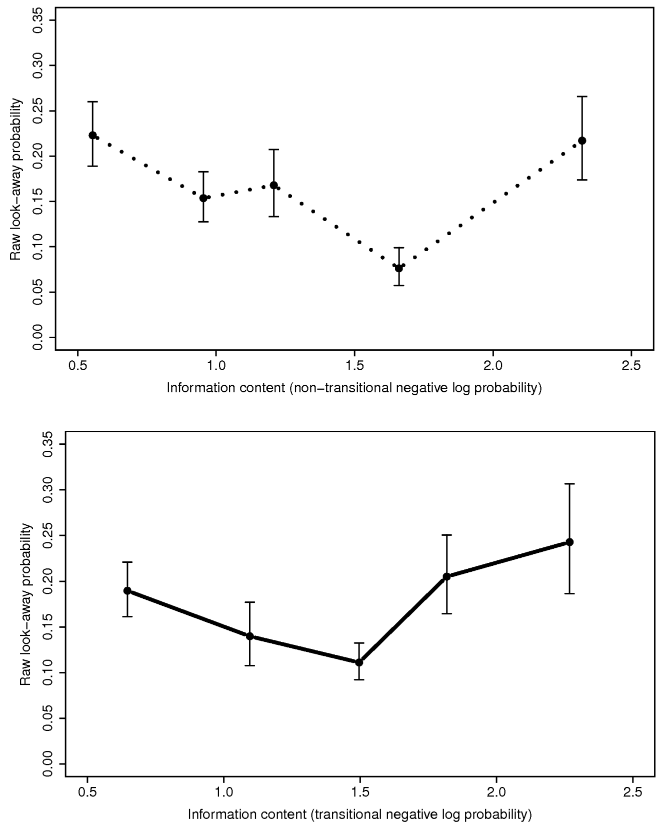

Figure 1: Infant look-away probabilities as a function of non-transitional surprisal (top) and transitional surprisal (bottom).

Results & Analysis

At each event in a sequence, infants make an implicit decision to either look away or keep looking at the scene. Figure 1 shows their raw probability of looking away at each item, as a function of that item's negative log probability according to the model, and collapsing across infants and sequences. The blue line shows the results for the non-transitional model, and the red line shows results for the transitional model. Both show a U-shaped relationship between raw look-away probability and model-based estimate of surprisal, with infants looking away to events that are especially surprising or especially predictable. There is a "Goldilocks" value of surprisal around 1.5, corresponding to infants' preferred rate of information in this task[3] which corresponds roughly to the point in the graph where infants have the lowest raw probability of looking away.

Survival analysis

Although the MDMs in Figure 1 provide a revealing picture of the relationship between indexes of surprisal and looking durations, there are likely other factors that influence infant look-aways. For instance, it might be the case that events in sequences generally become higher probability as infants form a picture of the statistical properties of the stimulus. If infants generally looked for a fixed amount of time, rather than paying attention to the statistical properties of the stimulus, then generally increasing predictability could make it look as though they preferred a certain information rate. To address this, we performed a regression analysis to control for the influence of other factors on look-away probability.

When infants look away, their trial ends and they provide no more additional data for that sequence [4]. This means that there is only a data point for an infant at time t if they have not looked away before t. We used a type of regression that respects this statistical relationship between look-aways and future data called a *survival analysis*. The type of survival analysis we used, Cox regression, measures the log linear influence that predictors have on the probability of a look-away at each point in time, but controls for a baseline look-away distribution. In the variety of survivial analysis we used, the baseline looking distribution is fit nonparametrically to the data, meaning that the analysis conservatively removes the

[3]This information rate must be interpreted relative to the frequency with which events in the sequence are presented, one every 2 seconds

[4]In the statistical literature, this type of data is called *censored*.

influence of an "average" distribution of looking times, before testing the significance of predictors.

We used a stepwise procedure for the Cox regression that tested whether each of several variables improved the model fit (AIC). Thus, at each iteration, the regression only added variables if they contributed positively, and at the same time removed variables if they contributed negatively. We included the following predictors in the survival analysis as control covariates.

- TRIAL-NUMBER: The number of sequences the child has already observed
- FIRST-APPEARANCE: A boolean factor corresponding to whether this event is the first time an object has been observed.
- UNSEEN-ITEMS: The number of objects that have not yet been observed.
- SAME-EVENT : A boolean factor for whether or not the current event is the same as the one that just happened.

The primary predictors we included in the survival analysis is the negative log probability of the event according to the MDM. Table 1 revealed that this variable is likely related to look-away probability quadratically, so we also included the squared negative log probability of the event according to the model[5]. A significant effect of squared predictability tests the significance of the U-shaped effect observed in Figure 1. As discussed above, we formed both transitional and non-transitional versions of the model, corresponding to models that treat each event independently, or each transition independently. Because the predictions of these two models are highly-correlated, we performed separate analyses on each.

Figure 2 shows the results of the survival analysis, including all predictors that were added via the stepwise procedure. These results can be interpreted by multiplying each coefficient by the value of the covariate and then exponentiating. This number represents an amount by which the probability of looking away is scaled, according to the best-fitting model. For instance, the coeffient of TRIAL-NUMBER is 0.033, meaning that by the 10'th sequence the child sees, they have a $exp(10 * 0.033) = 1.39$ greater factor of looking away. This effect of TRIAL-NUMBER is a plausible effect of fatigue. The results also show a significant effect of SAME-EVENT: children are a factor of $exp(0.316) = 1.37$ more likely to look away when the event is a repeat of the most recent event. This effect is also plausible: infants search for other things to keep their interest when the experiment shows a repeating–and therefore boring–event.

The regression results also reveal significant effects of NEG-LOG-PROB-SQUARED. Because these variables were standardized, the outcome can be interpreted as the response to changing the negative log probability by one standard deviation from those seen throughout the entire experiment. If the negative log probability of the event changes by one standard deviation, the probability of looking away changes by a factor of $exp(0.099) = 1.10$ for the non-transitional model and $exp(0.194) = 1.21$ for the transitional model. That is, infants are a factor of 1.1 to 1.21 more likely to look away on events that are either highly surprising or highly non-surpising according to an idealized statistical model for learning the structure of the sequences they observe.

The predictions of the the transitional and non-transitions models are difficult to distinguish because they are closely related: the information content of both models are correlated at $R = 0.62$ ($p < 0.001$). However, if both are entered into a stepwise Cox regression, the transitional NEG-LOG-PROB-SQUARED is significant at $p < 0.001$ (coef=0.25, $z = 5.74$)[6], while the non-transitional information content is not significant $p > 0.1$. This provides strong evidence that infants track transitional probabilities, but the null result for the non-transitional model is difficult to interpret due to its correlation with the transitional model and the noise inherent in infant data.

Conclusions & Discussion

These results have explicitly tested two interrelated hypotheses related to infants' looking behavior. First, we constructed a rational, statistical model that used observed events or transitions between events to form probabilistic expectations about what events are most likely in the future. This model embodies a simple, but non-trivial learning theory under which infants follow at least approximately rational statistical inference in inferring properties of the world. Second, we used this model to test whether infants have a preferred information rate in deciding where to allocate attention. The model was necessary in determining what information content a stimulus should convey, to an idealized observer. A failure of either theses assumptions—the probabilistic model or the linking assumption of the relevance of information content—would have yielded a null result.

In our analysis, we we used a Cox regression survival analysis, which allowed us to test the predictions of the model controlling for potential confounds such as the number of items that have not appeared yet, item repeats, and an arbitrary baseline distribution of look-away probabilities. To our knowledge, the hypothesis that infants prefer a fixed information rate has not been tested controlling for these other variables; nor has previous work used this type of formal model in measuring information rate. As such, this work provides several methodological advances. Rather than predicting infants' average looking time to a stimulus, our analysis attempted to predict the precise item in a sequence that an infant would look away on. We found that the information-theoretic properties of a formal model were a significant predictor of infant

[5] Covariates were standardized before including them in the analysis and before squaring them.

[6] The Variance Inflation Factors are small for these variables (< 3.1), suggesting that collinearity is not a substantial problem in computing statistical significance.

Non-transitional model				
Variable	**Coef.**	**Std. Error**	z	**p-value**
TRIAL-NUMBER	0.033	0.006	5.867	0.000
SAME-EVENT	0.213	0.100	2.140	0.032
NEG-LOG-PROB-SQUARED	0.099	0.049	2.024	0.043
Transitional model				
Variable	**Coef.**	**Std. Error**	z	**p-value**
TRIAL-NUMBER	0.033	0.006	5.791	0.000
SAME-EVENT	0.316	0.114	2.772	0.006
NEG-LOG-PROB-SQUARED	0.194	0.047	4.134	0.000
UNSEEN-ITEMS	-0.175	0.089	-1.959	0.005

Figure 2: Included variables using a stepwise Cox regression analysis to predict infant look-aways. In predictions of both transitional and non-transitional models, the squared (standardized) negative log probability is a significant predictor of look-aways.

look-aways, over and above the effects of other variables, but that their effect was U-shaped. Thus, the Cox regression validates the trend observed in Figure 1, showing that it does not result from other confounds.

We take these results as strong evidence for the theory that infants are the Goldilocks of the “blooming, buzzing confusion,” preferring stimuli with a certain moderate level of information, and are at least approximately rational in their decisions about where to allocate attention.

Acknowledgments

The first and second authors were supported by Graduate Research Fellowships from the National Science Foundation. The research was supported by a grant from the National Institute of Health (HD-37082). We thank Johnny Wen for his help with Matlab; Holly Palmeri, Laura Zimmermann, and Kathryn Schuler for their help preparing stimuli and collecting infant data; Suzanne Horwitz, Katheryn Lukens, Alyssa Thatcher, Lindsay Woods, and Rosemary Ziemnik for their help recruiting and scheduling subjects; and Collin Bannard, Michael S. DeFreitas, Noah Goodman, T. Florian Jaeger, Elissa Newport, Josh Tenenbaum, Ed Vul, Katherine S. White and members of CoCoSci, and the Aslin and Newport labs for comments and suggestions.

References

Berlyne, D. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.

Dember, W. N., & Earl, R. W. (1957). Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review*, *64*, 91-96.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *1964*, 668-670.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academcy of Sciences*, *99*(24), 15822-15826.

Haith, M. M. (1980). *Rules that babies look by: The organization of newborn visual activity*. Lawrence Erlbaum Associates.

Hunter, M. A., & Ames, E. W. (1988). Advances in infancy research. In L. P. Lipsitt (Ed.), (p. 69-95). New York: Academic Press.

Kemler Nelson, D. G., Jusczyk, P., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The headturn preference procedure for testing auditory perception. , *18*, 111-116.

Kinney, D. K., & Kagan, J. (1976). Infant attention to auditory discrepancy. *Child Development*, *47*(1), 155-164.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Dev Sci*, *11*(1), 122–134.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preference for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491-507.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 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*(1), 27-52.

Siqueland, E. R., & DeLucia, C. A. (1969). Visual reinforcement of nonnutritive sucking in human infants. *Science*, *165*(898), 1144–1146.

Wagner, S. H., & Sakovits, L. J. (1986). Advances in infancy research. In L. Lipsitt & C. Rovee-Collier (Eds.), (Vol. 4, p. 195-217). Nordwood, NJ: Ablex.

White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, *107*(1), 238-265.

Yu, C., & Ballard, D. H. (2007). A unified model of word learning: integrating statistical and social cues. *Neurocomputing*, *70*(13-15), 2149-2165.