Quantitative Linking Hypotheses for Infant Eye Movements

Daniel Yurovsky
dyurovsk@indiana.edu
Department of Psychological
and Brain Sciences
Indiana University

Shohei Hidaka
shhidaka@jaist.ac.jp
School of Knowledge Science
Japan Advanced Institute of
Science and Technology

Rachel Wu
r.wu@bbk.ac.uk
Centre for Brain and
Cognitive Development
Birkbeck, University of London

Abstract

The study of cognitive development hinges, largely, on the analysis of infant looking. But analyses of eye gaze data require the adoption of *linking hypotheses*: assumptions about the relationship between observed eye movements and underlying cognitive processes. We develop a general framework for constructing, testing, and comparing these hypotheses, and thus for producing new insights into early cognitive development. We first introduce the general framework – applicable to any infant gaze experiment – and then demonstrate its utility by analyzing data from three studies investigating the role of attentional cues in infant learning. Finally, we discuss general implications for construction and testing of quantitative linking hypotheses.

Keywords: eye movement data; infancy methods; Bayesian data analysis; learning; attention

Introduction

The study of infant cognitive development hinges, largely, on the analysis of infant looking data (Aslin, 2007). Since Fantz's (1964) landmark demonstration of visual memory in 2-month-old infants, researchers have used his habituation technique, and other eye-movement methods, to ask deep theoretical questions about the ontogeny and development of human cognition. But analysis of eye-movements, like analysis of other high-dimensional cognitive measures (e.g. fMRI, EEG) carries particular challenges (Yu, Yurovsky, & Xu, 2012). In order to connect observed eye-movements to underlying cognitive processes, one must define a linking hypothesis that relates them (Aslin, 2007; Teller, 1984).

Every eye gaze paradigm used to study infant cognition commits to a particular *linking hypothesis*. In habituation studies, decreased looking is hypothesized to indicate encoding, and recovery of looking indicates discrimination of a novel stimulus (Gilmore & Thomas, 2002). In violation of expectation studies, increased looking is hypothesized to indicate noticing a surprising event. Intermodal preferential looking studies hypothesize that a difference in looking time to one sound-object mapping over another indicates a difference in their associations. Critically, these linking hypotheses are *qualitative*; they assert that a relationship exists, but do not specify its *quantitative*, metric properties.

Why should we prefer quantitative linking hypotheses? They help us, in several ways, to move from asking *if* a phenomenon occurs, to asking *how* and *why*. First, quantitative linking hypotheses allow researchers to clearly and unambiguously specify the assumptions and mechanisms in their theories. As theories grow in complexity, correctly deriving their (sometimes counterintuitive) predictions can become difficult. Formalizing theories makes such prediction tractable (Shiffrin, 2010). Second, without quantitative predictions it can be impossible to distinguish competing theoretical accounts of the same data, fueling debates about "rich" (conceptual) vs. "lean" (perceptual) theoretical explanations (e.g., Spelke, 1998). Third, quantitative linking hypotheses allow researchers to test the same theoretical model across experiments, integrating multiple datasets within one self-consistent framework (Aslin, 2007; Schöner, & Thelen, 2006; Shiffrin, 2010).

Developmentalists who measure eye-movements, however, face several challenges to the construction of quantitative linking hypotheses. First, control of eye-movements is complex, and saccades are moderated by multiple systems (Aslin, 2007). Thus, quantitative linking hypotheses may need to integrate interacting mechanisms. Second, although fixation duration is likely related to learning, their relationship may not be a simple linear one. Instead, learning and looking may be linked non-monotonically, with a preference for familiarity appearing first, and a preference for novelty developing with further experience (Hunter & Ames, 1988). Linking hypotheses must be flexible enough to accommodate this kind of complexity. Third, early development is a time of rapid change, and the variability among infants of the same age may be surprisingly high. Thus, using the same linking hypothesis for each infant may distort true relationships in the data (Siegler, 1987). Because one cannot know apriori whether one's data is best analyzed as one group, or two, or three or more, construction of linking hypotheses must adaptively accommodate this kind of variability.

Building on a growing body of statistical tools in Bayesian non-parametrics, this paper presents a rigorous, principled, empirically successful framework for the construction of quantitative linking hypotheses that meets the three challenges reviewed above. To demonstrate the utility of this framework, we analyze data from a set of experiments investigating the role of social and non-social cues in infant multi-modal learning (Wu & Kirkham, 2010). This analysis shows how quantitative linking hypotheses can provide leverage in understanding the development and operation of infant learning mechanisms. We begin by presenting the general framework, demonstrating its robustness in simulation studies, and then present the empirical data.

General Model Framework

In any eye-tracking experiment, infants are exposed to stimuli that encode some structure of theoretical interest, and the researcher measures the influence of this structure on their behavior. For instance, in word-learning experiments, infants are exposed to consistent pairings between words and objects, and their discrimination for consistent vs. inconsistent mappings is measured (e.g. Yu & Smith, 2011). However, we are typically interested not in the change in observed behavior, but rather in the cognitive processes it implicates (Aslin, 2007). Quantitative linking hypotheses let us describe these processes directly.

For each infant, on each trial, the researcher observes some eye-gaze data (D), and the researcher's goal is to determine the model (M) that best explains these observed eye-movements ($P(M|D)$). This can be formalized as a problem of Bayesian inference. The researcher can specify several possible models, each making different predictions about the gaze data likely to be observed ($P(D|M)$). The researcher may also prefer simpler models apriori, in accord with Ockham's razor ($P(M)$). These properties can then be combined via Bayes' rule to infer the model that best describes the infants' cognitive processes (Equation 1).

$$P(M|D) \propto P(D|M)P(M) \qquad (1)$$

We present a graphical model (Figure 1) for connecting hypothesized cognitive models to observed eye gaze data. On each trial of an experiment, an infant (i) is exposed to some experimental stimuli (e) and produces observed eye movements (d). This observed gaze data is encoded as proportion of dwell time over a set of hypothesized areas of interest (AOIs). The inference framework discovers the set of underlying cognitive processes (s) that operate on the stimuli to generate the observed data. Intuitively, this is essentially a regression problem: inference finds the relationship between predictor variables (s,e) and observed outcomes (d). Because gaze data are a distribution over AOIs rather than a single continuous variable, we connect predictors to outcomes via the Dirichlet distribution (θ).
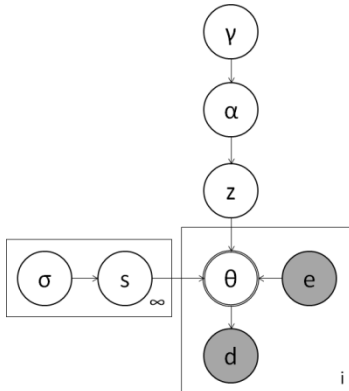

Figure 1: A graphical model for inferring the cognitive processes (s) responsible for generated eye movements (d) under particular experimental conditions (e).

The introduction identified three challenges for quantitative linking hypotheses: multiple processes may drive eye-movements, linking functions may be complex, and a group of infants may be heterogeneous. This framework meets all three challenges. Because s can encode any hypothesized cognitive model, the contributions of multiple processes can be estimated together without forcing a dichotomy (Anderson, 2011). Nonetheless, if a process has little effect, this is found via the prior on parameter values σ (Figuerido, 2002). Second, cognitive processes and observed eye movements need not be linked in a simple, linear way. In this framework, the cognitive model s can encode any functional link. For simplicity, and to minimize assumptions, we do so through arbitrary degree polynomials (see Jackson & Sirois, 2009). Again, the model parameter prior (σ) facilitates discovery of the most parsimonious linking function, penalizing complex polynomials.

Formally, each cognitive model parameter s is modeled as a draw from a 0-mean normal distribution whose standard deviation has a non-informative prior, making high values unlikely (Jeffreys, 1961). Each infant's data are modeled as a draw from a Dirichlet distribution over the AOIs whose parameters are defined as the exponentiated product of the cognitive model parameters and experimental settings e (Equations 2). This allows model parameters to be negative. A specific formulation is presented in the next section.

$$\sigma \sim \frac{1}{|\sigma|}$$
$$s \sim N(0, \sigma^2)$$
$$\theta_z = e^{(s_z \times e)} \qquad (2)$$
$$d \sim Dir(\theta_z)$$

Finally, infants in a sample may come from two or more different kinds of groups, (e.g. slow and fast learners: Yu & Smith, 2011). This framework automatically and adaptively determines the number of groups of infants, and the infants who belong to each group. Each distinct group is best represented by a different cognitive model. The estimation of unique groups is performed using the Chinese restaurant process (Aldous, 1985), which has been used successfully to determine unique groups in adult experiments (Navarro, Griffiths, Steyvers, & Lee, 2006). Clusters are discovered in this process by treating participants by analogy to customers in a Chinese restaurant. As each customer enters, he sits at each occupied table (z) with probability proportional to the number of occupants, but also chooses a new table with some small probability (α). This implements a rich-get-richer scheme in which groups that account for the behavior of many infants become favored, and the most parsimonious number of groups is discovered. A hyper parameter (γ) prevents us from having to make a direct decision about the probability of choosing a new table (Equations 3). Each cluster has different cognitive parameter values (s).

$$\alpha \sim Exp(\gamma)$$
$$z \sim CRP(\alpha) \qquad (3)$$

Case Study: Attentional Cues and Infant Learning

To demonstrate how this framework's utility in a concrete case, we applied it to data from a set of studies investigating the role of attentional cues in infant multi-modal learning. In each experiment, 8-month-old infants watched videos in which sounds and objects' on-screen locations were reliably related. When objects appeared in the top-left and bottom-right boxes, one sound was heard. When they appeared instead in the top-right and bottom-left boxes, a different sound was heard (Figure 2). In some conditions, infants were cued to one of the two objects. Subsequently, infants were exposed to test trials on which they heard a sound from training, but all four boxes were blank. If infants had learned sound-location regularities, they were expected to attend preferentially to locations consistent with each sound.

Submitting these test preferences to ANOVAs, Wu and Kirkham found reliable multi-modal learning only in the presence of the Face cue (2a), but not when infants were cued with a flashing square (2b) or received no cue (2c). We reanalyze this data to reveal significantly more structure, and to provide new insights into infant learning.

To this end, we define quantitative linking hypotheses for these experiments, formally specifying the connection between the observed eye-movement data (d), observable experimental conditions (e), and the unobservable, hypothesized cognitive processes (s). By analogy to regression, the data are the dependent variable, experimental conditions are the independent variables, and the cognitive processes parameterize these independent variables. On each trial of the experiment – whether training or testing – infants saw a black screen containing four boxes, one in each corner of the screen (Figure 2). Thus, we define five areas of interest (AOIs): one for each of the four boxes, and a fifth to capture all other looks (including off-screen looks).
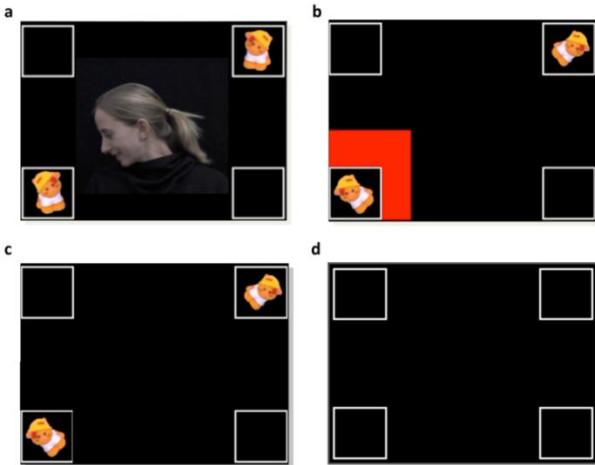

Figure 2: Training and testing trials from Wu & Kirkham (2010). In the Face condition (2a), a centrally-located face directed infants' attention to one of the boxes. In the Square condition (2b), a red flashing square highlighted one of the boxes. In the No Cue condition (2c) only the multi-modal regularity was present. On test trials (2d), all boxes remained empty while infants heard a sound from training.

The total data (d) for an individual infant is thus the entire set of gaze proportions observed on each trial. Formally, this is a matrix in which rows correspond to trials, columns to AOIs, and each cell to the proportion of looking to a particular AOI on a particular trial. This whole matrix is the outcome to be predicted from the experimental conditions (e) and hypothesized cognitive processes (s).

Next we specify the experimental conditions on each trial. While all four boxes were empty on test trials, on training trials two of the four boxes contained pictures of animals (Figure 2a-c). These are coded with binary indicator variables $salient$ specifying whether a box (b) contains a picture. Further, in the Face and Square conditions (Figure 2a-b), one of the boxes was highlighted by an attentional cue. We similarly define an indicator variable $cued$.

$$salient(b) = \begin{cases} 1 & box\ b\ contains\ stim \\ 0 & otherwise \end{cases}$$
$$cued(b) = \begin{cases} 1 & box\ b\ cued \\ 0 & otherwise \end{cases} \tag{4}$$

In addition to the visual stimuli, each trial also included a sound that could alter looking patterns if infants learned sound-location contingencies. To formalize this learning process (below), we encode infants' experience with these contingencies in the experimental conditions (e). Thus, we let $contingent$ be the cumulative time an infant has fixated a given box (b) in the presence of a particular sound (n). So, on trial t that plays sound n_t and on which we observe data d_t, where $\delta(i, j)$ is Kronecker's delta function,

$$contingent_t(b, n_t) = \sum_{r=1}^{t-1} \delta(n_r, n_t) \cdot d_r(b) \tag{5}$$

Last, we define cognitive processes that act on experimental condition variables to produce observed gaze data. First, infants may have a baseline preference for some locations over others. Thus, we include a preference constant p_b for each AOI, allowing the contributions of the other variables to be estimated relative to proper baselines. Second, an infant's preference for a box may be altered by the presence of an object ($salient$), or the presence of an attentional cue ($cued$). We let the strength of these factors be linearly scaled by parameters l and c respectively, which function like slope terms in linear regression.

Finally, in these experiments, the question of interest is whether infants learn to associate sounds with objects/locations. We define the effect of association between a sound and location as a change in preference for that location through exposure to the contingent sound. Specifically, we define association between a sound and location as a function of time spent fixating that location in the presence of that sound ($contingent$). To avoid making assumptions about the association function (e.g. that it is linear, or monotonic), we let association between box b and sound n on trial t be an arbitrary degree polynomial function of cumulative looking time to b while hearing n.

Since polynomials can approximate any functional form, this is a general solution (Jackson & Sirois, 2009). As in testing for higher-order terms in standard regression, polynomial coefficients are pushed down to zero by model priors if they do not contribute to predictive power.

$$assoc_t(b, n) = \sum_{o=1}^{O} a_o \cdot [contingent_t(b, n)]^o \qquad (6)$$

After formally specifying the experimental conditions and hypothesized cognitive process that act on this input, we can infer the effect of each hypothesized factor on infant the gaze data. As in regression, differences in parameters across conditions help us understand whether and how different cues affect infant multi-modal learning. To infer parameter values, we perform Bayesian inference in the model specified in Figure 1. Because this model has non-conjugate priors, we use an MCMC sampling algorithm that alternates Metropolis-Hastings updates with Split/Merge steps for cluster assignment (Jain & Neal, 2007). Sampling estimates the true distribution for each of these parameters, producing a set of credible intervals (similar to confidence intervals) that can be used to determine the likelihood that parameters are non-zero, as well as their likely range (Kruschke, 2011). Clustering was relatively insensitive to γ, so we let $\gamma = 1$.

Simulations

In order to ensure that it behaves as expected, we validate the analysis in a set of simulation studies by generating gaze data from a known cognitive model and trying to recover its parameters. In these simulations, we show that this framework can deal with all three challenges for quantitative linking hypotheses: non-homogenous samples of infants, interactions among multiple cognitive processes, and non-linear functions linking learning to looking.

Infants in Wu & Kirkham's (2010) study were simulated in training and testing trials like those in their experiments. Each simulated infant was exposed to four consecutive blocks, each consisting of six training trials and a test trial. On each training trial, objects appeared in two of the boxes (top-left and bottom-right, or top-right and bottom-left), and the lower box was cued. Each configuration of objects also co-occurred with a unique sound. Each of the two configurations occurred three times in each block of training trials, and order was randomized within a block. After all six training trials, infants saw one test trial where the screen was empty, but one of the two sounds was heard. Simulated infants then saw three more blocks, and each sound was tested twice in random order across the four test trials.

Simulation 1

In Simulation 1, we generated gaze data from known models in which the infants in a sample were drawn from a mixture of one, two, three, or four distinct groups. Formally, the 30 infants on each run were drawn from a multinomial distribution with equal probability for each group.

Parameters for each group were drawn randomly without replacement from $cued - c$: $\langle 0, 1, 2, 3 \rangle$, $salient - l$: $\langle 0, 1, 2, 3 \rangle$, and $contingent - a_1$: $\langle 0, .2, .4, .6 \rangle$. Baseline AOI preferences for each box were drawn uniformly from $[-2, -1]$ and off-screen preference was drawn from $[.5, 1.5]$. These values were representative of those found in the empirical analysis (next section).

Across all 120 simulations (30 runs at each group size), the correct number of groups was identified in all but 1. On one run at group size 4, the analysis identified only 3 clusters. Further, individual infants were almost always assigned to the right group. Group assignment was perfect when the number of true groups was 1 or 2, and less than a quarter of one percent ($<.0025$) of infants were misclassified at the higher group numbers. Thus, this framework deals well with heterogeneous groups of infants.

Simulation 2

Simulation 2 tested the framework's ability to recover correct quantitative parameter values when multiple processes interacted to produce eye movements. This time, all infants were drawn from one group, but group parameters were parametrically manipulated to sample the space of parameters recovered in the analysis of Wu & Kirkham's empirical data. Six unique parameter values were chosen for each hypothesized cognitive processes, and one simulation was run at each combination. Baseline preferences on each run were drawn as in Simulation 1. Figure 3 (next page) shows parameter estimates and true values for each combination of parameter values. Inference was successful: r^2 values were exceedingly high.
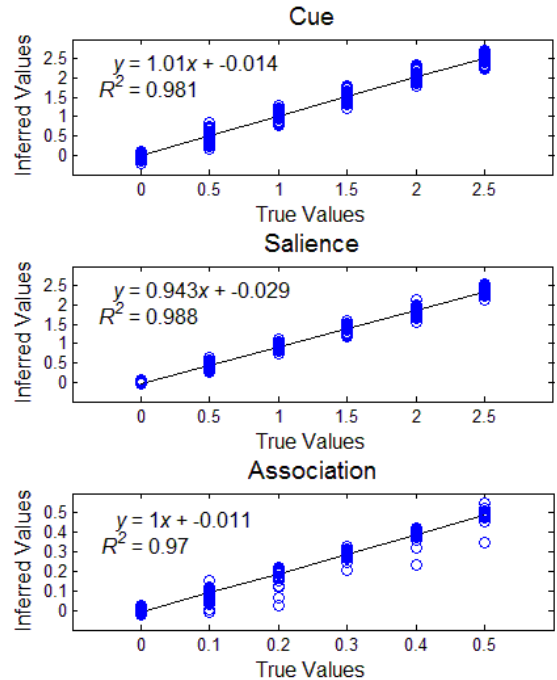

Figure 3: Best fit lines for true and inferred parameter values for each of the three factors hypothesized to affect infant gaze patterns in the experimental data.
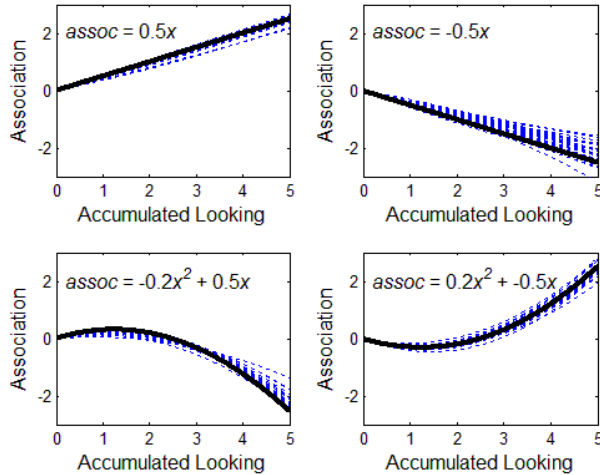
Figure 4: True functions (solid) and 30 inferred functions (dashed) for each learning function tested in Simulation 3.

Simulation 3

Simulation 3 tested the framework's ability to recover non-monotonic learning functions, for instance, a preference for familiarity followed by novelty. On each run of the simulation, 30 infants were generated with one of four possible learning functions: linear increasing, linear decreasing, u-shaped up, and u-shaped down (Figure 4). All other model parameters were drawn as in Simulation 2. Bayesian 95% credible intervals for estimated parameters were analyzed to determine how often a 0-valued parameter was estimated to be non-zero (0%) and how often a non-zero parameter was estimated to be 0 (2.5%). Thus, both Type I and Type II error rates were low (Kruschke, 2011).

Empirical Analysis

We now apply the same inference procedure to gaze data from real infants. Instead of comparing the effects of different attentional cues on raw looking patterns, as in standard analyses (e.g. ANOVAs), inferring cognitive model parameters for each condition lets us analyze the effects of cues directly on attention and learning.

Method

Inference was performed using the same model as in the Simulations above. Infants each saw a series of training and testing trials, and gaze data on each trial were coded as a proportion of looking to each on-screen box as well as a fifth AOI for all other looks (Figure 2). As before, inference recovered the joint distribution for all parameters (s) explaining looking as a function of experimental conditions. Full parameter descriptions are in the Simulation section. Gaze data included 26 infants in the No Cue condition, 29 infants in the Face condition, and 30 infants in the Square condition (see Wu & Kirkham, 2010 for full details).

Before presenting the results, we review Wu & Kirkham's ANOVAs for test trial looking. These analyses showed associative learning only in the Face condition. In contrast, infants in the No Cue condition showed no learning, and

infants in the Square condition preferred the cued locations, but did not learn to sound-location associations.

Results and Discussion

Inference yields full posterior distributions for all cognitive model parameters, estimating the contribution of each factor in the context of all other factors. We focus on two key factors: attention to the cue (c) and the association function ($assoc$). Figure 5 shows estimated parameter values for both factors for infants in each experimental condition.

First, in no condition were infants best described as a single homogeneous group. Two distinct groups were identified in the Face and No Cue conditions, and four groups were found in the Square Condition. Thus, even within one condition, infants learned and used cues differently. Second, all learning functions were linear; credible intervals for all association coefficients ≥ 2 overlapped 0 in all conditions. Thus, Figure 5 shows the first-order association coefficient (a_1) for each group.

Finally, we turn to the parameter values and their implications. First, all infants in the No Cue condition appeared to be learning ($a_1 > 0$), although approximately ⅔ had low association values, indicating that they learned slowly. The Face condition had a comparable number of equally fast learners, and these fast learners did not show evidence of using the cue ($c \approx 0$). However, the larger, slow group of learners did use the cue, and learned faster than the slow learners in the No Cue condition. Learners had two routes into learning the regularity: quickly and directly, or slowly and indirectly. This detailed level of structure underlies and explains Wu & Kirkham's coarser analysis.

The Square condition also had a small group of fast learners who used the cue. However, in contrast to the other conditions, approximately ½ of the infants did not learn, and these infants all used the cue ($c > 0$). These results directly confirm Wu and Kirkham's hypothesis that the flashing square may interfere with learning by competing for attention, and that only the fastest learners may be able to learn from these kinds of competing cues. Together, these results both confirm the major findings from the standard analysis and provide deeper insight into how attentional cues guide (or interfere with) infant multi-modal learning.
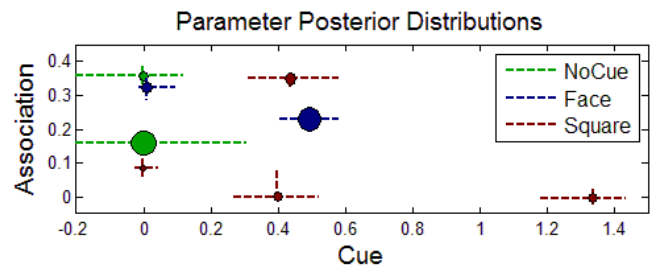

Figure 5: Posterior distributions for *cue* (c) and *association* (a_1) parameters for infants from Wu & Kirkham (2010). Each circle indicates a cluster, and its size indicates the proportion of infants in that condition in that cluster. Circles are centered at median parameter values, and dashed lines indicate 68% credible intervals, akin to +/-1 SE.

1207

Non-monotonic learning functions. Simulation 3 showed that this framework can recover non-monotonic functions linking learning to looking when appropriate for the data (Hunter & Ames, 1988). However, no such functions appear in the Empirical Analysis above. Why? One possibility is that non-monotonic linking functions arise in a different kind of experiment or at a different age. An alternative possibility is that non-monotonic linking functions are seen when infants' baseline preferences are not controlled. In our analyses, we included a set of parameters p_b to encode baseline preferences for each location. When these parameters were not included, we *did* find non-monotonic linking functions in all conditions. Thus, we propose that, at least in some cases, observation of non-monotonic linking functions may be an artifact of different baseline preference rather than a core property of the learning system itself.

Competing Hypotheses. One strength of quantitative linking hypotheses is that they facilitate direct comparison of competing theories for the same data. In the previous sections, we argued that changes in looking preferences over the course of these experiments arise from associations between heard sounds and fixated locations, and modeled this learning with the *assoc* function. Alternatively, preferences could change over time through habituation; infants' preferences could change as a function of looking to a location independent of the concurrent sound. We tested this directly, by modeling habituation as an arbitrary-degree polynomial function of cumulative looking time to a location (Equation 7). However, 95% credible intervals for *habit* parameters overlapped 0 in all conditions, out this explanation for the data. Thus, quantitative looking hypotheses allowed us to directly compare two hypothetical explanations of this data and to choose the best alternative.

$$habit_t(b) = \sum_{o=1}^{O} h_o \cdot \left(\sum_{r=1}^{t-1} d_r(b) \right) \qquad (7)$$

General Discussion

Infant researchers have made tremendous progress by using eye gaze data to ask questions about early cognition and development. The majority of this work has used *qualitative* linking hypotheses, but we propose that even faster progress can be made through model-based analyses using *quantitative* linking hypotheses (Aslin, 2007; Teller, 1984). While quantitative linking hypotheses have been proposed for specific experiments (e.g. Gilmore & Thomas, 2002; Yu & Smith, 2011), this paper presents a general framework applicable to all eye movement experiments. We hope this work will facilitate asking and answering future questions about early cognitive processes and their development.

Acknowledgments

References

Aldous, D. (1985). *Exchangeability and related topics. In École d'été de probabilités de Saint-Flour, XIII—1983* (pp. 1–198). Berlin: Springer.

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*, 1-8.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*, 48-53.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*, 668–670.

Figuerido, M. A. T. (2002). Adaptive sparseness using Jeffreys prior. *Advances in Neural Information Processing Systems*, *14*, 722-729.

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

Hunter, M.A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In L.P. Lipsitt (Ed)., *Advances in child development and behavior* (pp. 69-95). New York: Academic Press.

Jackson, I., & Sirois, S. (2009). Infant cognition: going full factorial with pupil dilation. *Developmental Science*, *12*, 670-679.

Jain, S., & Neal, R. M. (2007). Splitting and merging components of a nonconjugate Dirichlet process mixture model. *Bayesian Analysis*, *2*, 445-472.

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299-312.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet Processes. *Journal of Mathematical Psychology*, *50*, 101-122.

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*, 273-299.

Shiffrin, R. M. (2010). Perspectives on modeling in cognitive science. *Topics in Cognitive Science*, *2*, 736-750.

Siegler, R. S. (1987). The perils of averaging over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250-264.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, *21*, 181-200.

Teller, D. Y. (1984). Linking propositions. *Vision Research*, *24*, 1233-1246.

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*, 118-136.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*, 165-180.

Yu, C., Yurovsky, D., & Xu, T. (2012). Visual data mining: An exploratory approach to analyzing temporal patterns of eye movements. *Infancy*, *17*, 33-60.