Model selection for eye movements: assessing the role of attentional cues in infant learning

Article · January 2012

DOI: 10.1142/9789814340359_0005

5 authors, including:

Daniel Yurovsky
Stanford University
36 PUBLICATIONS **663** CITATIONS
SEE PROFILE

Rachel Wu
University of California, Riverside
65 PUBLICATIONS **651** CITATIONS
SEE PROFILE

Chen Yu
Indiana University Bloomington
179 PUBLICATIONS **4,204** CITATIONS
SEE PROFILE

Natasha Z Kirkham
Birkbeck, University of London
58 PUBLICATIONS **3,279** CITATIONS
SEE PROFILE

MODEL SELECTION FOR EYE MOVEMENTS: ASSESSING THE ROLE OF ATTENTIONAL CUES IN INFANT LEARNING

DANIEL YUROVSKY

Department of Psychological and Brain Science, Indiana University, 1101 E 10th St., Bloomington, IN 47405, USA

RACHEL WU

Centre for Brain and Cognitive Development, Birkbeck,, University of London, Malet Street, London, WC1E 7HX, UK

CHEN YU

Department of Psychological and Brain Science, Indiana University, 1101 E 10th St., Bloomington, IN 47405, USA

NATASHA Z. KIRKHAM

Centre for Brain and Cognitive Development, Birkbeck,, University of London, Malet Street, London, WC1E 7HX, UK

LINDA B. SMITH

Department of Psychological and Brain Science, Indiana University, 1101 E 10th St., Bloomington, IN 47405, USA

A recent study [1] showed that different attention cues (social and non-social) produce qualitatively different learning effects. The mechanisms underlying such differences, however, were unclear. Here, we present a novel computational model of audio-visual learning combining two competing processes: habituation and association. The model's parameters were trained to best reproduce each infant's individual looking behavior from trial-to-trial in training and testing. We then isolated each infant's learning function to explain the variance found in preferential looking tests. The model allowed us to rigorously examine the relationship between the infants' looking behavior and their learning mechanisms. By condition, the model revealed that 8-month-olds learned faster from the social (i.e. face) than the non-social cue (i.e., flashing squares), as evidenced by the parameters of their learning functions. In general, the 4-month-olds learned more slowly than the 8-month-olds. The parameters for attention to the cue revealed that infants at both ages who weighted the social cue highly learned quickly. With non-social cues, 8-month-olds were impaired in learning, as the cue competed for attention with the target visual event Using explicit models to link looking and learning, we can draw firm conclusions about infants' cognitive development from eye-movement behavior.

1. Introduction

1.1. *Multimodal Relationships*

The infant's world is filled with sights and sounds that belong together (e.g., toys falling, people talking, cars driving). Linking these concurrent sights and sounds helps infants organize their cluttered multimodal world [2][3]. Moreover, tracking multimodal relationships (e.g., synced in tempo, motion, and location) helps infants learn about the events: One fundamental advantage of binding sights and sounds is that they allow for the prediction of future events. After learning information in two modalities (e.g., seeing and hearing a favorite toy dancing), it is useful for infants to remember the location of the toy when they only hear the toy's sound across the room. Tracking audio-visual information in this way can help infants organize events in spatial locations. Six-month-olds associate sounds to objects when they are perceived simultaneously in a particular location [4]. When presented with only the sound, infants look to the location where the object had previously appeared. This ability, referred to as spatial indexing, has been documented in infants as young as 3 months of age, with increasing reliability and flexibility over the first year of life [4][5].

1.2. *Social and Non-social cues*

The aforementioned research has established that infants distinguish and track multimodal events when presented in isolation (one multimodal event at a time). In noisy natural environments, unlike the laboratory setting, infants are often presented with multiple streams of cross-modal information. How do infants know which audio-visual event to bind? One way of knowing could be relying on attention cues. Since attending to appropriate events is a fundamental part of learning, infants can rely on attention cues to gain an advantage in learning relevant information. Cues can capture, direct, and sustain attention, though the quantitative and qualitative nature of this shifted attention can vary with each type of cue.

Both social and non-social cues shift infants' attention. Infants follow faces that look in a particular direction [6-8] or when flashing, dynamic shapes appear in a particular location [9-10]. Infants reach adult-levels of attention shifting with non-social cues in the periphery by 4 months of age [10]. With social cues, however, infants begin following eye gaze in simple naturalistic situations by 3 to 4 months (e.g., joint attention [6]) and show dramatically increased reliability by 6 to 8 months of age [8][11].

1.3. *Learning from Attention Cues*

How do different attention cues affect learning? In other words, how useful are these cues in helping infants' cognitive development? [1] found that depth of learning audio-visual events from social and non-social attention cues was dependent on the age of the infant and the nature of the attention cue. This study measured gaze behavior of 4- and 8-month-old infants when they were presented with dynamic audio-visual events (i.e., cats moving to a *bloop* sound and dogs moving to a *boing* sound) in white frames in the corners of a black background. An object's appearance in a spatial location consistently predicted a location-specific sound. On every familiarization trial, infants were shown identical audio-visual events in two diagonally opposite corners of the screen (i.e., two valid binding locations). To test the effects of attentional cueing on audio-visual learning, either a social (i.e., a real face) or non-social (i.e., colorful flashes) cue shifted infants' attention to one of the two identical events on every trial. For the social cue, a face appeared, spoke to the infant, and turned to one of the lower corners containing an object. For the non-social cue, a red flashing square wrapped around the target frame appeared and disappeared at a regular interval (i.e., flashed continuously) throughout the familiarization trial. During the test trials, only the four blank frames were displayed on the screen while one of the sounds played (Figure 1).

Within one familiarization trial, two locations could have been associated with the sound, and across all familiarization trials, two locations were cued (one for each object type). The design of this paradigm allowed for the discrimination of two types of learning: 1) learning from attention cues, and 2) audio-visual learning. They measured where infants predicted the objects would appear. There were four possible learning outcomes: 1) Infants could predict that objects would appear in only cued locations (lower corners) regardless of multimodal information, 2) infants could predict that objects would appear in valid binding locations regardless of where they were cued, 3) infants could use both information from the attention cue and multimodal presentation by predicting objects would appear in cued correct object locations, or 4) infants could use neither set of information and look in incorrect frames or equally to all frames.

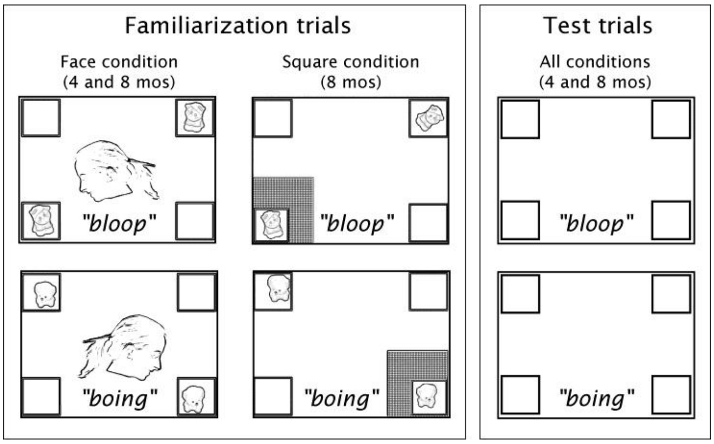

Figure 1: Familiarization and test trials from the Face and Square cueing conditions from [1]. The stimuli were in full color; the frames were white and presented on a black background. The shaded areas represent red flashing squares.

Results showed that infants displayed qualitatively different looking patterns at test, despite fixating for equal time to the correct locations across different training conditions. [1] showed that while both cues led infants to attend preferentially to cued locations during familiarization and test across all age groups, cross-modal contingencies were learned only by older infants exposed to social cues. Only 8-month-olds cued by the face during familiarization looked longer to *correct* cued object locations during test, whereas the 4-month-olds in the same condition anticipated events in both correct and incorrect cued locations, perhaps ignoring the multimodal information. The 8-month-olds cued with flashing squares (non-social cue) also looked longer to cued locations than to non-cued locations regardless of object–sound mappings. These findings suggest that specific multimodal learning is dependent on the nature of what orients attention as well as the age of the infant using the attention cue.

2. Microanalysis: A Model Selection Approach

2.1. *Motivation*

Using a preferential looking time method, [1] demonstrated the impact of attention cues on learning during infancy. This paradigm used only a behavioral measure (i.e., looking time) and collapsed the data within a condition for each age group. As a result, three issues remain unresolved. First, the distribution of learning rates within each condition is unknown. For example, did every infant learn in the same way, or are there clusters of different learner types within each condition? Second, whether the learning differences between conditions were quantitative or qualitative is unknown. If they were due to a quantitative measure, a factor such as training length could eventually match the learning effects across conditions. This would not be the case for a qualitative difference. Third, we do not know the exact mechanisms underlying the learning behavior. What factors drove the anticipatory looking? To address these three issues, we analyzed the temporal dynamics of eye movement data at a finer level.

In order to make claims about infants' learning mechanisms, inferences must be made from the available data: looking behavior. This requires a *linking hypothesis* to connect looking and learning. In the preferential looking paradigm, [1] implicitly committed to one such hypothesis: as infants learn about the associations between sounds and locations, they will prefer to fixate those locations when they hear the matching sounds. The same hypothesis used frequently in early word learning research [12]. Other infancy studies, focusing on other topics (e.g., visual perception [13]), however, sometimes adopt the opposite hypothesis: as infants learn more about an event, they preferentially fixate new locations to search for novel events (i.e., looking to the original event wanes, habituation). Moreover, these mechanisms that could drive infants' real-time looking behavior are complex. For example, infants' habituation profiles may be non-monotonic, with a preference first for familiarity and a later preference for novelty emerging as learning progresses [14][15]. This suggests two troubling implications for analysis of eye movements – one at the individual infant level, and one at the group level. First, if infants first show a familiarity preference and then a novelty preference over the course of learning, they may progress through a period of no preference in between. An infant who shows no preference, then, may in fact have learned an association. Second, if individual infants learn at different rates – and evidence suggests that they do [16] – then group analyses may average together data from infants who have a novelty preference with those who have a familiarity preference. A group null result thus becomes difficult to interpret [17][18]. We account for these factors in a model selection framework, using infant eye movements in the course of learning to demonstrate and interpret differential learning from attention cues.

2.2. *Model Selection for Eye Movements*

The main issue with using a linking hypothesis is initially choosing the appropriate one. As suggested by [19], this study relied on the data to specify the linking hypothesis. A model selection approach to this problem is a four-step procedure. First, we generated a formal description of the learning task. This included the input available to the infant on each trial of the experiment and the output categories (i.e., regions of interest) as they were measured in the original data analysis. Second, we defined a null model for the task. This null model was the function expected to generate eye movements for infants who did not learn about the input. Third, we defined a set of possible linking functions. The next section presents one principled way of generating a set of such functions (for an alternative, see [20]). Finally, we defined a method of selection, a way of determining which linking function was most likely to have generated each infant's recorded eye movements. For the selection, we used the Akaike Information Criterion (AIC) [21], a heuristic that trades off increasing the fit to data with increasing the number of parameters in the model. Other heuristics are possible, (e.g., BIC). Akaike's criterion, however, has been shown to have lower error when the true model is not among the set of candidates [22] – a situation very likely given the simplifying assumptions in our analysis. This issue will be further discussed in the context of specific models below.

The following sections review the four steps of the model selection framework for the cued attention task presented above, and demonstrate its utility for drawing inferences about infant learning at multiple levels.

3. Modeling the Cued Learning Task

3.1. *Formalizing the Task*

On each trial of the experiment, infants saw a black screen with four white-framed boxes, one in each corner of the screen. These four boxes, or locations (Loc_1, Loc_2, Loc_3, Loc_4), were defined as our regions of interest. On each trial, the model was asked to predict the duration of looking time to each of these four locations. For mathematical convenience, we used the log odds of looking to each location [23]. Using the odds ratio form of the exponentiated Luce choice axiom [24], we proposed that odds of looking to each location were the odds ratio of their theoretical activation functions (defined below).

$$OddsLook(Loc_i) = \frac{\exp Act(Loc_i)}{\sum_{j \neq i} \exp Act(Loc_j)} \tag{1}$$

The four boxes, however, were unlikely to be equally interesting because objects appeared in two of the four boxes. We will refer to any box which contained an object as *salient*. Formally,

$$Salient(Loc) = \begin{Bmatrix} 1 & Loc\ contains\ an\ object \\ 0 & Otherwise \end{Bmatrix} \tag{2}$$

In addition, one of the locations was *cued* – either by a centrally-located face (Face condition) or by a flashing square around the box (Square condition).

$$Cued(Loc) = \begin{Bmatrix} 1 & Loc\ was\ cued \\ 0 & Otherwise \end{Bmatrix} \tag{3}$$

Finally, a sound was heard on each training trial. Since the sound was not a component of the visual display, we proposed that it did not have a direct effect on fixation patterns, but guided looking indirectly through sound-location associations. We return to the role of the sound on fixations in Step 3.

In contrast to the training trials, test trials contained neither objects nor visual cues, only white boxes and the sound. Because the test trials did not differ from the training trials in any other way and because the boxes were on screen in the presence of sounds for the same length of time as in training trials, we described both types of trials with the same formal vocabulary. The only difference was that the *salient* and *cued* functions always had the value '0' for all locations on the test trials.

3.2. *A Null Model*

After formalizing the structure of each trial, we defined a null model for the task. This null model defined the activation function (above) for infants for whom looking was not guided by learning. In the absence of learning, we suggest that looking was guided by two potential sources: 1) The on-screen cue's direction of attentional shift (*Cued*), and 2) the presence or absence of an object in each box (*Salient*). The null model was thus a function of these two factors:

$$Act(Loc) = c \times (Cued(Loc)) + s \times (Salient(Loc)) \tag{4}$$

The constant c weighted the importance of the cue, and similarly, s weighted the importance of the salient objects.

3.3. *A Learning Model*

After defining the null model, we specified the linking function, which defined the activation function for infants whose gaze patterns were driven by their learning. First, we proposed that infants may have remembered which screen locations they frequently fixated. They may then have preferred not to fixate those locations on future trials (i.e., habituating to them). Efforts to characterize infant habituation functions [15][19][20] have modeled them with polynomial functions or bounded exponentials. Because habituation functions are non-monotonic [14], for simplicity we used polynomials of up to degree 2 (arbitrary orders are possible in principle). Thus, we defined habituation to a screen location as a polynomial function of cumulative looking time to that location. Thus, the function could be increasing, decreasing, or both, and could be either linear or faster-than-linear. Formally, if N_h is the maximum order of the infant's habituation function,

$$Habit(c\_time_{Loc}) = \sum_{n=1}^{N_h} h_n \times c\_time_{Loc}{}^{n} \qquad (5)$$

where h_n is the parameter of the n^{th} term of the infant's habituation function. Parameters were selected by regression to best account for the infant's looking data. Then, model selection (below) was used to determine which order of habituation was appropriate for each individual infant.

Second, we proposed that infants may have learned the relationship between the sounds and the on-screen locations of the objects (in two locations) that were presented in synchrony. Binding the audio and visual events was of primary interest to [1]. To determine whether infants learned this relationship, we formalized how this knowledge could drive looking. As with the habituation function above, we proposed that looking on a current trial was driven by a polynomial function of cumulative looking. However, in this case, we used cumulative looking to a location in the presence of the sound to indicate this type of learning. That is, the polynomial is a function of the subset of looking during which the association sound was heard. Equation 6 represents this dependency as a conditional probability notation. In addition, we again used up to second degree polynomials (though any order is possible in principle). Thus, if N_a is the maximum order of association for a given infant,

$$Assoc(c\_time_{Loc} \mid sound) = \sum_{n=1}^{N_a} h_a \times c\_time_{Loc \mid sound}{}^{n} \qquad (6)$$

In sum, the goal of the above models is to produce an explicit account of where infants should look on each trial given one of a set of possible models. Each of the models must predict the amount of time an infant will spend looking at each of the four possible screen locations on each trial. This amount is represented as a fraction of total looking time using a log odds representation.

These fractions are then stored to represent cumulative looking time, which is the variable on which learning functions operate.

The models of learning used in this analysis are functionally similar to neurally inspired learning models [15], but are not intended to link directly to any brain structures. We take this approach primarily for efficiency and simplicity. The functions which are considered here are easy to parameterize, compare, and analyze, which are the questions of critical importance to this analysis. Tying these models to underlying neural mechanisms would be an interesting question for future research.

3.4. *Selecting the Best Model*

Up to this point, we have defined a set of possible learning models, as well as a null model for the task. In order to determine whether (and how) each infant learned, we only needed to determine which function was most likely to have generated the observed looking patterns. Therefore, we picked a criterion (Akaike Information Criterion [21]) to judge the correspondence between the model and the observed data. AIC trades off minimizing the divergence between predictions and observations (in the form of sum of squared errors) with increasing the complexity of the model. Intuitively, as the model becomes more complex (higher orders of association and habituation), it will produce an increasingly better fit to the data. As a result, however, there will be diminishing returns: eventually, the corresponding increase in goodness of fit for adding a new parameter will become small relative to the increase in model complexity. As per Occam's razor, we wanted the simplest model with a 'good enough' fit.

3.5. *Model Training*

After formalizing the learning task, defining a null model and a set of candidate learning models, and describing how to select among them, the next step was to fit the functions to each infant's looking behavior to find the most suitable one. Fitting functions is an optimization problem – a process of parameterizing the functions to fit the data.

The form of the *OddsLook* function supports easy parameter fitting via linear regression, for which a closed form solution is known. Thus, for each infant, parameters were selected for each function to best predict the data. Then, the most parsimonious of these models was selected with AIC. Because there was no principled reason to separate training trials from test trials, and because including training trials greatly increased the amount of data available, models were fit to the entire course of looking behavior during the experimental session for each infant.

4. Results and Discussion

We present the results in two steps. First, we walk through an in-depth analysis of the 8-month olds in the Face condition to show how the outcome of the model selection analysis can be interpreted. Second, we present a cross-condition comparison using the model selection framework. This analysis highlighted the differences between the experimental conditions, and allowed us to ask how learning differed across cue conditions and age groups.

4.1. *Data Analysis of 8 months Face Condition*

Data from 28 8-month-old infants were analyzed using the model selection framework described above. Infants were exposed to four blocks, each consisting of six training trials followed by one test trial. Over 28 trials, infants fixated each of the four locations for some length of time. Each model was required to fit this looking data. The large number of trials (training and test) per infant in these experimental sessions provided significantly more data than is typically used to assess learning via preferential looking analysis (only test trials).

To justify and draw conclusions from the modeling analysis, we first established that the models fit a significant proportion of each infant's looking data. The best model for each infant matched closely to that infant's actual data (mean Pearson's $r = .72$, $p < .001$, see Figure 2). Having confirmed that the chosen linking functions characterized each infant's looking data; we could draw further inferences from these functions.

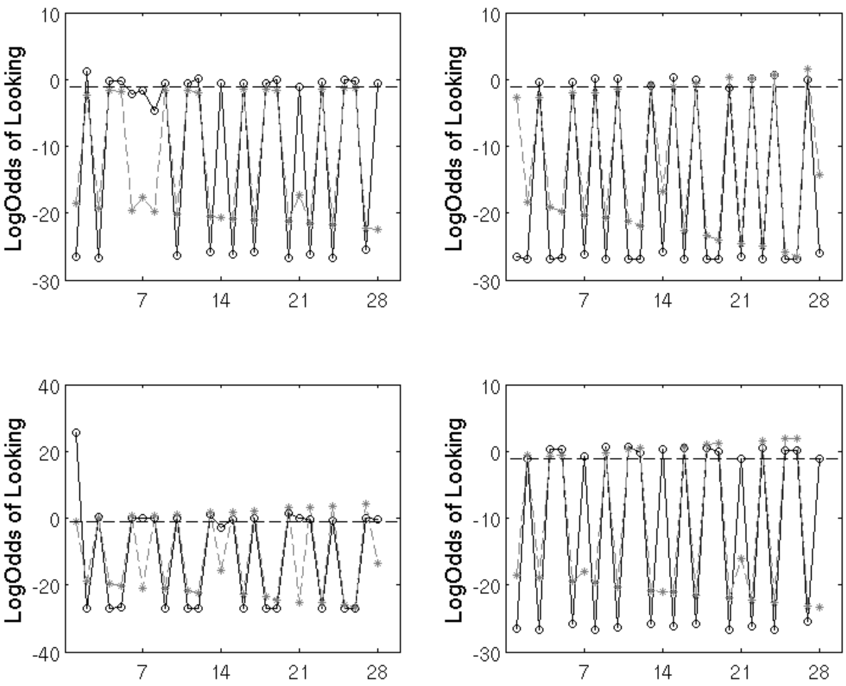

Figure 2: Looking data (black circles) and model fits (gray circles) for one 8-month-old in the Face condition. The four graphs represent the four locations on each screen, and each data point represents log odds of looking on one trial. Marked trials on the x-axis indicate test trials.

Because each infant was modeled individually, we can determine whether or not each infant learned. As the main factor of interest is the nature of associations formed by the infants, we focused on parameters of infants' association functions (N_a). Of the 28 infants, 22 were found to have learned a relationship between sounds and screen locations ($N_a > 0$). Of these, the great majority learned at a faster-than-linear rate ($N_a > 1$, 17/22). Thus, multi-modal learning in this condition was both frequent and rapid. Moreover, if the model was correct, we expected to find a relationship between infants' looking preferences at test and the order of their association functions. Indeed, the strength of an infant's preference for the correct location at test correlated significantly with that infant's order of association (r = .41, p < .05). Figure 3 displays both the distribution of habituation and association orders, and the correlation between association and looking preference at test.

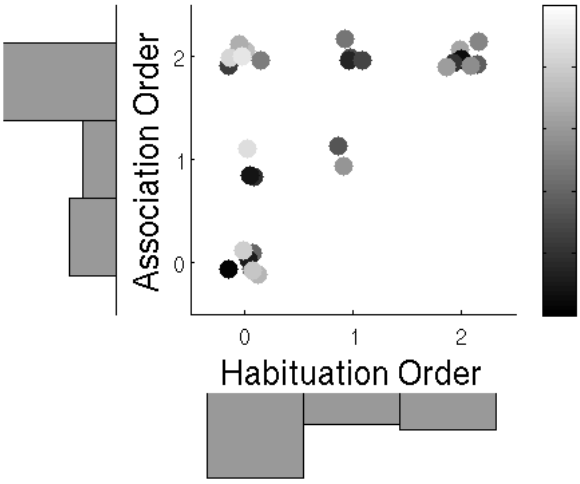

Figure 3: A scatter plot of infant habituation and association orders. Association orders greater than zero imply that an infant learned the relationship between sounds and locations. When infants are ordered by the strength of their preference at test (lightest = strongest preference), this correlates with their order of association.

4.2. *Comparison Across Conditions*

The real power of the model selection framework is the leverage it provides for comparing across multiple conditions. We thus applied the same four-step procedure of the framework to each of the two other conditions – 4 months Face (i.e., face cue) and 8 months Square (i.e., flashing square cue). As in the previous condition, the most parsimonious model for each infant accounted for a significant proportion of the looking behavior ($r_4 = .68$, $p < .001$, $r_8 = .69$, $p < .001$). Therefore, we were licensed to make further inferences on the basis of the linking functions.

Analysis of the orders of association for the 8 months Face condition demonstrated that the vast majority of infants learned the multi-modal relationship, and that the majority of infants who learned did so at faster-than-linear rates. We compared this distribution of association orders to those found in the other two conditions. As can be seen in Figure 4, the distributions are different across conditions. Eight-month old infants cued with a flashing square instead of a face were much less likely to learn cross-modal contingencies ($N_a > 0$). In addition, infants who did learn the contingencies did so more slowly than those in the Face condition. The ratio of quadratic learners dropped from 3:1 (Face condition) to 2:1 (Flashing square condition). Four-month-olds in the

Face condition exhibited different behaviors from the other two conditions. The number of learners was in between those of the two 8 months conditions. More interesting was the distribution of linear and quadratic learners. In contrast to the 8-month-olds, the majority of whom were best described by quadratic functions, 4-month-olds were much more likely to be linear learners. Thus, the rate of learning from the face cue seemed to increase with age. Moreover, it seems that the face cue was easier to learn from than the flashing square cue.

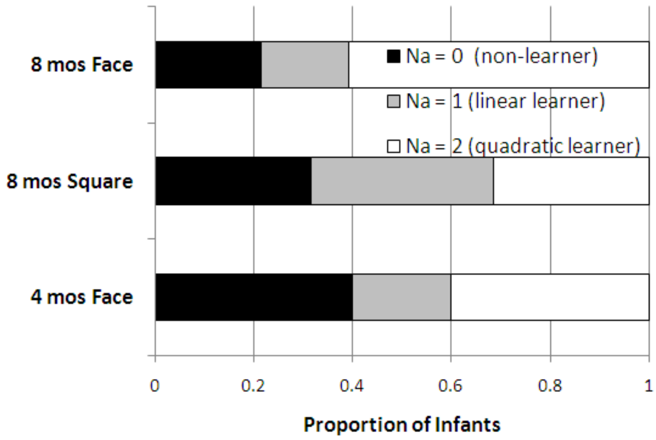

Figure 4: Distribution of association order across conditions. Eight-month-olds learned more rapidly from the face than from the square, and 4-month-olds showed a different ratio of linear (N_a=1) to quadratic (N_a=2) learners than 8-month-olds.

For completeness, we performed a similar analysis for infants' habituation functions. Although habituation is not the question of primary interest in this analysis, it is worth making explicit we have included it as a mechanism which may have driven looking behavior. In fact, infants across all three conditions showed very similar habituation order distributions (Figure 5). Slightly less than half of the infants in each condition produced eye movements which implicated a habituation function. The rate of habituation (at least in terms of linear vs. quadratic) did not appear to differ across conditions or ages. Thus, although habituation is important for capturing some of the behavior in this task, it does not appear to be important for describing differences across cues. We thus focus further analysis on association orders.

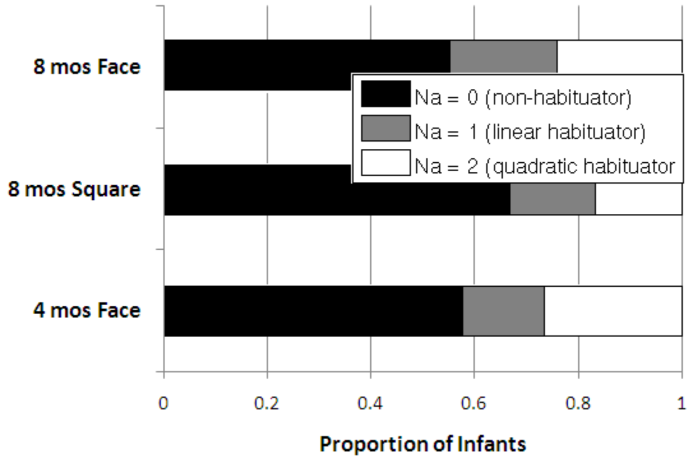

Figure 5: Distribution of habituation order across conditions. Unlike the distribution of association orders across conditions (Figure 4), the habituation parameters were consistent across ages and cue types. Thus, habituation plays a role in directing eye movements in this task, but does not do so differently for different types of cues.

How did the learners in the different conditions learn? In addition to the learning rates found to best describe each infant, the models included weights for both cue and salience strength (eq. 4). We investigated the relationship between these parameters across conditions. Because the cue and object salience compete directly, we should expect a negative correlation between these two parameters across infants. The relationship of these parameters to learning rates, however, is not specified by the model's form. We thus investigate this relationship empirically. Did infants who learned have eye movements driven more by cues (face or square) or by object salience (object presence)? Figure 6 shows the cue and salience parameters for each infant (standardized to z-scores). Infants who did not learn (N_a = 0), linear learners (N_a = 1), and quadratic learners (N_a = 2) are represented by different markers. These distributions differed considerably across conditions.

In the 8 months Face condition, there was a clear separation between learners and non-learners. The successful learners were characterized by a strong preference for cue over salience. The infants whose looking was driven more by the cue than by the salient objects may have been more able to learn the relationship isolated by the cue. Thus, there was a strong positive correlation between the difference between cue and salience weights, and having a non-zero association order ($r = .395$, $p < .05$).

At first glance, the 8-month-old Square condition looks similar. Again, infants who were drawn too strongly by the salient objects relative to the cue did not learn successfully ($r = .440$, $p < .05$). A second cluster, however, is apparent on the graph: the infants who learned more slowly (linear) were also those who assigned the most weight to the cue relative to the salient objects. Perhaps these infants were drawn to the cue, rather than the objects, and therefore did not learn the multimodal relationship. If they were drawn to the cue, infants must have attended to the cue covertly (without fixating on the cue) because total looking time to the cued object was similar between the 8 months Face and 8 months Square conditions.

Finally, the 4-month-old Face condition showed two qualitatively different groups of learners. The linear learners were predominantly those whose looking was driven more by the salient objects than by the cue. In contrast, the quadratic learners showed predominantly the opposite pattern (similar to the 8 months Face condition). This distinction suggests that the 4-month-old infants may have been learning in two different ways, one way more sophisticated than the other. As in the 8-month-old Face condition, but not in the 8-month-old square condition, the best learning was exhibited by infants who attended predominantly to cued objects rather than to both objects. Nonetheless, some 4-month-olds cued by the face and some 8-month-olds cued by the square tended to ignore the cue and still learn the multimodal association. This alternative strategy perhaps requires more cognitive effort (learning associations in two locations) and more time. This analysis shows precisely why the model selection framework is powerful – it avoids the ‘perils’ of averaging across strategies [25].

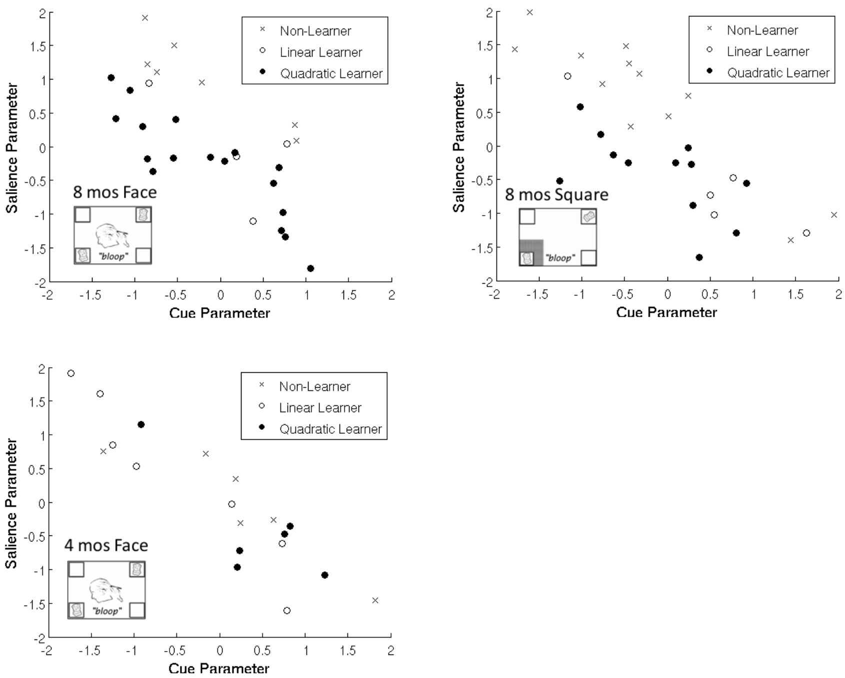

Figure 6: Weights for *cued* and *salient* locations across conditions labeled by learning score. The negative slope is a direct consequence of our null function, but the relationship of the parameters to learning rates is not. In fact, weighting the cue over the salient locations led to better learning in both 8 months conditions, with an eventual decline in the face condition. Two distinct clusters of learners were found among the 4-month-olds.

In summary, we found that the weights of the two parameters in the learning function (*Cued* and *Salience*) were related to learning in both conditions for both age groups. In general, learners were most successful when they paid more attention to the cue than to the salient objects. However, the relationship was non-monotonic in the Square condition, with too much or too little attention to the cue leading to slower learning [26]. This result suggests a qualitative difference between the Face and Square conditions: infants may have processed the two cues differently. A crucial difference between the central social cue and the peripheral non-social cue is that the peripheral cue was wrapped around the target box. Perhaps this cue was harder to disengage from compared to the central cue, which occupied a separate spatial location from the cued event (for discussion, see [1]). The linear learners in the Square condition may have been compelled to look at the correct location, but focused on the cue itself rather than the multi-modal relationship. If there was an increased focus on the cue, it must have been covert (without fixating on the cue) because total fixation length to the target object was similar between the two 8 months conditions [1].

Regarding the ability to disengage from the cue, the 4-month-olds who focused on the cued locations were the fastest to learn, perhaps engaging with the task most like the 8-month-olds. This notion suggests that more practice would allow more of these younger infants to exhibit successful learning. Combined with the results from the two 8 months conditions, these results suggest an optimal cue-weight parameter – a point of maximal return depending on the cue type and age of the infant. While 8-month-olds who focus too much on the cue do not learn as quickly, 4-month-olds are slower processors of complex scenes and thus need relatively higher cue weights. Alternatively, because infants have just started following social cues at this age [6], perhaps infants who did not learn from the cue do not yet know *how* to learn from this cue (i.e., did not follow the cue purposefully). Future research should identify the underlying reasons driving looking behavior between and within groups.

5. Conclusions

Following [1], which found that different cues elicit different audio-visual learning, we developed a model to show a more detailed picture of the learning. Importantly, this model proposed two underlying mechanisms of such differential learning. We found that both habituation and association played a role in learning from the cues, and that these processes were best characterized by different functions across age groups and cues. Eight-month-olds learned less quickly from a flashing square than from a face, perhaps because too much covert attention to the square led to difficulties processing the cued stimuli. Four-month-olds fell into two clusters – those who learned quickly by following the face cue, and those who learned more slowly by ignoring it. These results suggest that at four months, infants may be just on the cusp of learning from the face.

A model selection method investigating individual infant behavior supports a deeper analysis of infant learning than statistical tests between infant groups in the standard preferential looking paradigm. As a result, the underlying mechanisms can be specified more completely, and distributions of individual differences across conditions and age groups can provide more data than binary tests of statistical significance. In order to understand infant learning we must move beyond documenting *what* can be learned, and begin asking *why and how* it is learned. Answering these questions will require us to make deeper inferences from eye movement data – to understand the processes which generate fixations and ultimately lead to learning. With the model selection framework, we can begin to unravel the contributions of internal and external drivers of this behavior. We can make eye movements make sense.

Acknowledgments

This work was supported by a NSF Graduate Research Fellowship to DY, NIH Grant R01HD056029 to CY, and a BPS Postgraduate Study Vists Award to RW. The authors would like to thank Shohei Hidaka for help developing the modeling methodology and Thomas Smith for data processing code.

References

1. R. Wu and N. Z. Kirkham, *J. Exp Child Psych*. **107**, 118 (2010).
2. L. E. Bahrick, In L. Nadel, *Encyclopedia of Cognitive Science* (Nature, 2004)
3. D. J. Lewkowicz, *Psych. Bull.* **126**, 281 (2000).
4. D. C. Richardson and N. Z. Kirkham, *J. of Exp. Psych.: Gen.* **133**, 46 (2004).
5. N. Z. Kirkham, D. C. Richardson, R. Wu, and S. P. Johnson, (under review).
6. G. Butterworth, in A. Slater and G. Bremner, *Theories of infant development* (Wiley Press, 2004).
7. B. M. Hood, J. D. Willen, and J. Driver, *Psych. Sci.* **9**, 131 (1998).
8. A. Senju and G. Csibra, *Curr. Bio.* **18**, 1 (2008).
9. M. H. Johnson, *Att. and Perf*. **15**, 291 (1994).
10. M. H. Johnson, M. I. Posner, and M. K. Rothbart, *J. of Cog. Neuro.* **3**, 335 (1991).
11. D. Poulin-Dubois, T. L. Demke, and K. M Olineck, in R. Flom, K. Lee, and D. Muir, *Gaze-following: Its Development and Significance* (Lawrence Erlbaum, 2007).
12. R. M. Golinkoff, K. Hirsh-Pasek, K. M. Cauley, and L. Gordon. *J. Child Lang.*: **14**, 23 (1987).
13. N. Z. Kirkham, J. A. Slemmer, and S. P. Johnson. *Cognition* **83**, B35 (2002).
14. M. A. Hunter and E. W. Ames. *Advances in Infancy Research* **5**, 69 (1998).
15. S. Sirois and D. Mareschal. *Trends. Cogn. Sci.* **6**, 293 (2002).
16. C. H. Cashon and L. B. Cohen, *Infancy* **1**, 429 (2000).
17. W. K. Estes, *Psychol. Bull.* **53**, 134 (1956).
18. D. Yurovsky, S. Hidaka, C. Yu, and L. B. Smith. *Proc. Cog. Sci.* (2011).
19. H. Thomas and R. O. Gilmore, *Psychol. Methods* **9**, 70 (2004).
20. R. O. Gilmore and H. Thomas, *Infant Behav. Dev.* **25**, 399 (2002).
21. H. Akaike, *IEEE T. Automat. Contr. Infancy* **AC-19**, 716 (1974).
22. Y. Yang, Biometrika. 92, 937 (2005)
23. D. Barr, *J. Mem. Lang.* **59**, 457 (2008).
24. R. D. Luce, *Individual Choice Behavior* (Wiley Press, 1959).
25. R. S. Siegler, *J. Exp. Psychol. Gen.* **116**, 250 (1987).
26. C. Kidd, S. T. Piantadosi, and R. N. Aslin, *Proc. Cog. Sci.* (2011).