Anticipating an Effect from Predictive Visual Sequences: Development of Infants' Causal Inference from 9 to 18 Months

Jeffrey K. Bye[1] (jkbye@ucla.edu), Bryan D. Nguyen[1] (bnguyen07@ucla.edu), Hongjing Lu[1,2] (hongjing@ucla.edu), Scott P. Johnson[1,3] (scott.johnson@ucla.edu)
Departments of Psychology[1], Statistics[2], and Psychiatry and Biobehavioral Sciences[3]
University of California, Los Angeles; Los Angeles, CA, USA

Abstract

There has been little research on infants' development of causal inference in the second year after birth. We report an experiment in which 9- to 18-month-old infants viewed visual sequences consisting of three looming shapes, one after another. Half of the sequences (causes) were predictive of an attention-getting reward (effect), and the other half were non-predictive. The statistical complexity of predictive sequences was varied between conditions. We analyzed latencies of infants' eye movements toward the reward location. Older infants yielded more anticipatory eye movements in predictive than non-predictive sequences. Effects of both infant age and complexity of causal sequences were observed. To qualitatively account for these findings, we formulated a Bayesian model based on generic priors favoring simple causal events coupled with noisy shape identification.

Keywords: causal inference; development; Bayesian modeling; perception; infants

Introduction

Understanding cause-effect relations is vital to cognitive development. Imagine, for example, an infant attempting to disambiguate his mother's actions as she starts up the car, adjusts the mirror, scans the surroundings, engages first gear, and turns the steering wheel, at which point the car moves. How do infants make sense of such *causal action* sequences? How do they determine which actions are necessary, and in which order, to produce an effect? How do they use this knowledge to guide their own actions?

We address these questions with a combination of behavioral and computational evidence. We report an experiment in which infants observe causal action sequences for which cause-effect relations are specified by conditional probabilities of varying complexity, and we describe a Bayesian model that simulates the inferences that support identification of causal structure. We reasoned that identifying causal structure incorporates multiple sources of information (e.g., spatiotemporal, statistical) and perceptual/cognitive mechanisms (e.g., visual attention, detection of serial order, working memory, inherent biases) that must operate in concert during the learning process.

Causal Inference in Infants

Previous research has shown that infants make remarkable progress in acquiring sophisticated inferential abilities in the first two years after birth. Within the first few months, for example, infants demonstrate statistical learning in multiple sensory modalities (Frank et al., 2009; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996; Teinonen et al., 2009) and motion-based causal perception given spatial and temporal contiguity (Leslie & Keeble, 1987). By 18-24 months, children begin to expect predictable effects from learned sequences of discrete causal events (Bonawitz et al., 2010), and by 24-48 months, they routinely use the covariation patterns between actions and state changes to categorize objects by causal power and predict the effects of causal actions (Gopnik & Sobel, 2000).

How do infants progress from the perceptual mechanisms of early infancy (statistical and rule-based learning and motion-based causal perception) to the extraction of causal structure from temporal covariation between discrete events? Little is known about learning processes in infancy that support causal inference. Research on causal inference in children under 18 months of age has been limited because the tasks typically used, such as the "blicket" detector paradigm (Gopnik & Sobel, 2000), require vocal ability and/or motor skills. It has been difficult, therefore, to characterize infants' development of causal learning, leaving open the issue of whether early motion-based causal perception is generalized to higher-order causal inference (Michotte, 1946/1963), or whether the former is merely a special case of the latter (Cheng, 1993).

Causal inferences are necessarily constrained by limits in perception, attention, and memory. In particular, potential causal cues and their temporal ordering must be determined before statistical computation can proceed. The literature on animal conditioning is consistent with an inherent connection between detection of causal cues and causal inference. For example, Balleine et al. (2005) showed that increased perceptual discriminability of cues enhances rats' ability to acquire cause-effect associations, influencing sophisticated behavior such as retrospective revaluation of cues. We would expect, therefore, that development of causal inference is constrained by infants' perceptual processing maturity and working memory capacity, as well as the identification of conditional probabilities involved in predicting an effect. Our experiment and our model represent a first step in understanding developments in causal inferences from probabilistic information during this important transitional age range, as well as the perceptual and cognitive constraints on the learning process.

Paradigm

An experiment by Buchsbaum et al. (2011) with 41- to 70-month-old children provides a paradigm that we adapted to create a causal inference task that can be performed by younger infants. The children in Buchsbaum et al.'s study

observed an adult performing a series of five 3-action sequences on a toy (e.g., knocking, stretching, rolling), such that certain sequences were followed by a desirable outcome (a musical jingle). Afterwards, the children were given the toy, and asked to reproduce the effect.

Three conditions were examined, varying the statistical complexity of causes. In the ABC condition, the music played only when a specific set of three actions (A, B, and C) were performed in order; in the BC condition, the music played whenever the final two actions (B and C) were performed in order, regardless of what action was performed in the first position; and in the C condition, the music played after every sequence (all sequences ended in action C, but the first and second positions varied).

The results of the experiment demonstrated that when given the toy, significantly more children in the BC and C conditions were willing to imitate shorter action sequences that had never been shown in isolation, rather than a 3-action sequence as they had been shown during learning. The children appeared to infer the conditional probabilities between the actions that are predictive of the effect, and used that information to make rational choices about which actions performed by the adult needed to be imitated in order to produce the effect. These results, and the Bayesian model formulated by Buchsbaum et al. (2011), suggest that the children combined this conditional probability information (statistical learning) with prior beliefs about causal events in making decisions to imitate specific actions.

Experiment

The goal of our study was to investigate perceptual and cognitive constraints on infants' learning of causal action sequences, in particular developments related to identification of conditional probabilities governing causal structures. To examine the development of causal inference in infants, we combined methods from studies of statistical learning (e.g., Kirkham et al., 2002) with a perceptual analog of the causal inference paradigm described previously (Buchsbaum et al., 2011). We presented infants with 48 sequences of three looming, colored shapes that were predictive (or not) of the appearance of an attractive "reward," a colorful, animated attention-getting stimulus shortly after the final shape's offset. We operationalized the learning of causal structure as a greater tendency to anticipate the reward (using eye movements as the dependent measure) in *true* (predictive of the reward) vs. *false* (non-predictive) sequences, and we examined the time course of learning in each condition across trials.

Statistical learning experiments involve passive detection of conditional probabilities from a continuous stimulus stream. Our approach differs by requiring the detection of cues' statistical regularity in predicting a reward that is qualitatively and spatially distinct from the cues themselves. This design and the use of an eye tracker allow us to study the mechanisms characteristic of causal inference in a younger age range than what has been previously reported outside of motion-based causal perception.

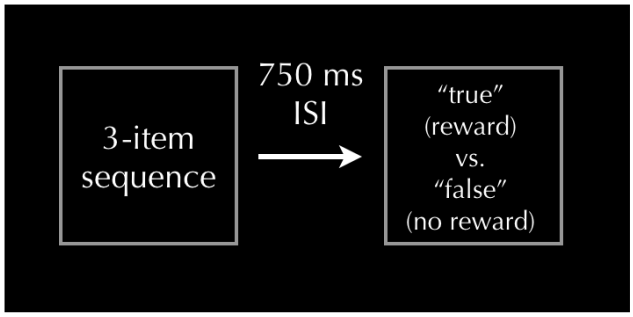

Figure 1: Schematic depiction of the stimulus layout.

Method

Participants Sixty-three infants (31 female), ranging from 9- to 18-months-old (*M* age = 13.6 months), were recruited from a database of new parents and tested at the University of California, Los Angeles (UCLA) Baby Lab. An additional 28 infants were observed but eliminated from data analysis due to excessive fussiness (13), inattention (12), or experimenter error (3). Infants were given a toy as a thank-you gift in exchange for participation in the experiment. Twenty-five infants were assigned to the ABC condition, 19 to the BC condition, and 19 to the C condition (described subsequently).

Procedure Infants were seated in a parent's lap; the adult was instructed not to interact with the infant or look at the stimulus monitor. Infants were placed 60 cm from the monitor. Eye movements were recorded with an SR EyeLink 1000 (SR Research, Ltd) at 500 hz.

The computer screen displayed two gray square frames, 9.4 cm per side (9.0˚ at the 60-cm viewing distance), side-by-side on a black background and separated by 8.6 cm (8.2˚) (Figure 1). One frame (the left in Figure 1, but counterbalanced across subjects) was designated for the shape sequences, while the frame on the other side displayed the reward (when applicable). The shapes were presented for 750 ms each (2250 ms for the total sequence), followed by a 750-ms interstimulus interval (ISI) after the third shape. For the trials with the reward (true trials), a dynamic attention-getter appeared after the ISI in the opposite frame for 1500 ms. Example videos available: http://www.babylab.ucla.edu/index.php?page=causal-action.

Analogously to the paradigm employed by Buchsbaum et al. (2011), each infant was randomly assigned to one of three conditions: ABC, BC, or C. These conditions correspond to the number and position of the shapes that were predictive of the reward (Figure 2). In the ABC condition, all 24 true sequences consisted of the same three shapes in the same positions, such that shape A was always first, shape B second, and shape C third. In the BC condition, the 24 true sequences consisted of 8 trials each of ABC, DBC, and EBC, such that only shapes B and C in the second and third positions (respectively) predicted the reward (i.e., the first position's shape was variable). In the C

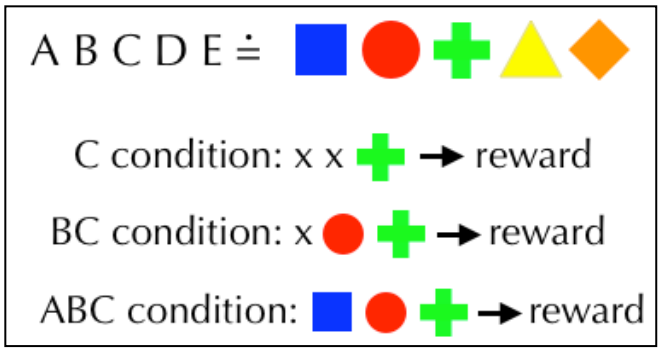

Figure 2: Schematic depiction of causal action sequences. Distinct colored shapes correspond to items A-E. In the actual experiment, shape assignment (i.e., which shapes corresponded to which letter roles) was randomized across subjects. The letter *x* indicates variable shape positions (randomized each trial).

condition, the 24 true sequences all ended with shape C in the third position, but the first two positions were randomly sampled (without replacement).

For all three conditions, the 24 false sequences were random permutations of the 5 shapes (A-E) with the constraints that in the BC condition there were no BC*x* trials (i.e., B and C did not appear consecutively except in the second and third positions), and in the C condition, there were no C*xx* or *x*C*x* trials (i.e., C did not appear outside the third position). Shapes were never repeated within a trial (e.g., there were no AAC or BDB trials) for either true or false sequences. The 48 sequences were presented in a randomized order.

For each infant, shapes were randomly assigned to role A, B, C, D, or E. The shapes were a blue square, red circle, green cross, yellow triangle, and orange diamond (Figure 2), such that color and shape information were redundant (to increase distinctiveness and facilitate identification).

Eye tracking data were analyzed to extract *predictive gaze shifts* (PGS), defined as directing the point of gaze away from the frame displaying the shape sequences toward the reward's frame 200 ms or less prior to the actual onset of the reward (inferred to be anticipatory). The 200-ms criterion was based on the estimated time to program an eye movement (Gredebäck, Johnson, & von Hofsten, 2010). Gaze shifts to the reward's frame that were completed after 200 ms past its onset were not classified as PGS (inferred to be reactive). For each subject, we computed the difference between the number of PGS for true trials and for false trials. We used this PGS difference score as an indicator of how well the infants learned what sequences predicted the effect.

We hypothesized that (1) older infants would produce more PGS for true vs, false trials (higher PGS difference scores) than younger infants, regardless of condition, due to development in perceptual processing, attention, and/or memory, and that (2) infants would produce the highest PGS difference scores in the C condition, followed by BC, followed by ABC, due to the added complexity of tracking conditional probabilities in BC and multiple sets of conditional probabilities in ABC.

Results

As noted, a greater number of PGS in true vs. false trials was taken as evidence for causal structure learning. We first used a median split to divide the sample into two age groups, and found that older infants (M = 15.6 months) produced reliably more positive PGS difference scores overall than younger infants (M = 11.4 months), M PGS difference = .65 vs. -.86, respectively, $t(61) = 2.29$, $p = .026$. However, as revealed by within-condition t-tests, older infants' performance significantly exceeded chance only in the C condition, ($t(9) = 2.50$, $p = .034$), while it was marginally significant in BC ($t(8) = 1.98$, $p = .084$) and not significant in ABC ($t(14) = -.335$, $p = .742$).

Analysis of correlations between age and PGS difference confirmed these observations. We found a significant overall correlation between infants' age and PGS difference ($r = .28$, $p = .025$). Within each condition (Figure 3), the correlation between age and PGS difference was statistically significant for the C condition (r = .48, $p = .036$), marginally significant for the BC condition (r = .39, $p = .10$), and non-significant for the ABC condition (r = .16, $p = .45$).

To examine learning across trials, we computed a series of generalized estimating equations with PGS difference as the dependent variable. We found a marginally significant age x trial type x trial interaction in the ABC condition ($p = .09$), but the interactions in BC and C conditions were not statistically significant ($ps = .15$ and .77, respectively).

The overall positive correlation between age and PGS difference supports our first hypothesis that older infants

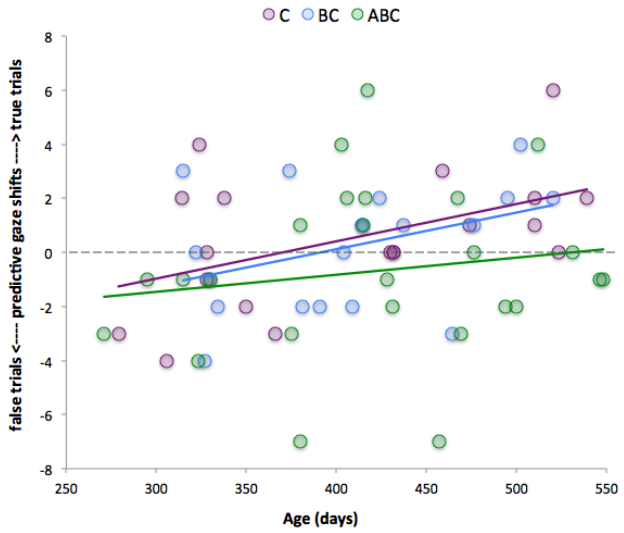

Figure 3: Differences in predictive gaze shifts (PGS) in true vs. false trials as a function of condition and age. Points correspond to individual infants, plotted by their age and the number of their PGS during true trials – the number of their PGS during false trials. Chance level performance = 0.

would outperform younger infants. The within-condition correlations suggest that development in this period is strongest for the C condition, followed by BC, which provides tentative support for our second hypothesis, implying that infants are better able to make causal predictions when the number of conditional probabilities they need to track is relatively low. Performance did not improve across trials in the BC and C conditions as infants gained experience in the task. There was some suggestion of age-related improvement across trials in the ABC condition, perhaps owing to older infants' ability to detect multiple conditional probabilities in this sequence with repeated exposure.

Bayesian Model

We developed a Bayesian model to simulate the causal inference in each trial and to predict the occurrence of the reward (effect). The model incorporates generic priors favoring simple causes, a constrained causal hypothesis space (subset of all possible hypotheses) to reflect the limited working memory of infants, and a noisy function for correctly identifying and memorizing the shapes in each sequence to emulate developing perception and memory systems in infancy.

To enable the inference, the model first defined a constrained causal space (set of all possible causal structures to be considered) using all permutations of the five candidate shapes in each of the three positions, with allowance for variable shape position(s). For example, the causal structure in Figure 4a would predict the effect following a sequence of the blue square in the first position, red circle in the second, and green cross in the third; the structure in Figure 4b predicts the effect following the red circle in the second position and green cross in the third, regardless of the first shape. This first example would correspond to an ABC cause in the behavioral paradigm, and the second a BC cause. Unlike Buchsbaum et al. (2011), non-terminal subsequences (e.g., AB*x*) are considered possible causes.

The hypothesis space, H, is defined over the causal space, such that a given hypothesis, h, is drawn from a set of causes. The current model assumes that all hypotheses consist of only one cause, reflecting the limited working memory in infancy.

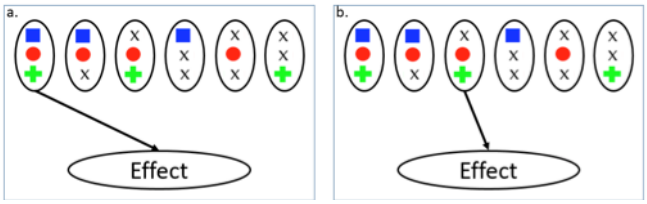

Figure 4: Two possible hypotheses (a subset of the constrained hypothesis space), each positing a single cause of the effect. For each candidate cause, the specified shape must appear in that position (e.g., a blue square first). The letter *x* indicates variable shape positions.

To make a comparison with infants' behavioral measure (PGS), the model predicts the probability of a reward given the current shape sequence and all previous data. The computation sums over all hypotheses h in the hypothesis space H to compute the probability of the reward (r) given some current shape sequence (s) and all previous trials' sequence and reward data (D):

$$P(r \mid s, D) = \sum_{h \in H} P(r \mid s, h) P(h \mid D) \,. \qquad (1)$$

The posterior distribution, $P(h \mid D)$, represents the most current belief about the cause of the reward given all prior data, and is easily computed from Bayes' rule:

$$P(h \mid D) = \frac{P(D \mid h) P(h)}{\sum_{h'} P(D \mid h') P(h')} \,.$$

The prior probability of a hypothesis, $P(h)$, is based on the prior distribution defined in Buchsbaum et al. (2011, Eq. 3), modified as:

$$P(h) = P(c \in h) = \frac{1}{1 + \exp\left(-\beta(|c| - 2)\right)} \,, \qquad (2)$$

where $|c|$ is the length of causal sequence c in hypothesis h. In our experiment, this is the number of specified shapes, 1 to 3. In Eq. 2, β represents a prior favoring causal sequences of a certain length. $\beta < 0$ gives preference to shorter sequences.

The likelihood of observing the data given a specific hypothesis, $P(D \mid h)$, is computed as the product of the likelihood probability for all previous trials' sequences and the presence or absence of the reward:

$$P(D \mid h) = \prod_{s,r \in D} P(r \mid s, h) \,.$$

The conditional probability of the reward given a specific shape sequence and causal hypothesis, $P(r \mid s, h)$, also computed in Eq. 1, is determined by two factors: the probability of a candidate causal event c generating a reward r, and the probability of identifying causal event c based on the observed shape sequence s,

$$P(r \mid s, h) = \sum_{c} P(r \mid c, h) P(c \mid s) \,. \qquad (3)$$

The two components in Eq. 3 can be conceptualized, respectively, as the *prediction* of the reward's presence or absence given the constituent causes of the sequence, $P(r \mid c, h)$, and the *identification* of each candidate cause from the sequence of shapes, $P(c \mid s)$.

The prediction component computes the probability of the effect following the identified sequence and corresponds to *causal reasoning*. In this model, it is assumed that causes are deterministic, i.e., the reward always occurs if the candidate causal event in the hypothesis is identified.

The identification component involves a noisy process of recognizing potential causal cues, resulting from *attention*, *perception*, and *memory constraints* in infancy. We introduce probabilistic noise to allow a certain probability of misidentifying or misremembering cues. This probability weights the deterministic causal reasoning to yield a probabilistic likelihood function in Eq. 3. The identification probability can be calculated using Bayes' rule as:

$$P(c \mid s) = \frac{P(s \mid c)P(c)}{\sum_{c'} P(s \mid c')P(c')}. \quad (4)$$

The prior probability of each candidate cause c is computed as in Eq. 2. The likelihood of s given c is defined using a multinomial distribution as:

$$P(s \mid c) = \theta^m \left(\frac{1-\theta}{4}\right)^{|c|-m}, \quad (5)$$

where θ represents the probability of correctly identifying and remembering a given shape ("matching probability") and m is the number of specified shapes in c that match (in the same position) those in s. Given that there are five unique shapes, $\theta = 0.2$ corresponds to matching at chance.

Assessment

Under the assumptions of single-cause hypotheses and deterministic causes, we evaluated the performance of the model under different values for model parameters: the probability that an infant will correctly identify and remember a given shape (θ) in Eq. 5 and the degree of the preference favoring shorter causal sequence lengths (β) in Eq. 2. These parameters correspond to constraints on perception and simplicity preference in causal reasoning, respectively.

The model predicted the probability of a reward, $P(r \mid s, D)$, for each trial in each individual infant's session. Then we averaged $P(r \mid s, D)$ across all true trials (i.e., predictive of the reward) and all false trials, separately. To compute the model prediction analog to the PGS difference score for infants, we calculated the difference in expected frequencies of PGS for both true and false trials, to fit each individual's data. We then averaged across subjects within each of the three conditions.

Figure 5 shows the model's performance as a function of β, the degree to which simple causes are favored, when $\theta = 1$ (i.e., perfect shape identification). When $\beta < 0$, favoring simple causes, the model achieves the best performance in the C condition, followed by BC, then ABC (the latter two track closely in the range $-1 < \beta < 0$), consistent with older infants' performance as shown in Figure 3. However, when $\beta > 0$, indicating a bias favoring longer shape sequences (as used in Buchsbaum et al., 2011), the model predicts the opposite pattern, with the ABC condition yielding the best performance, and the C condition the worst. Accordingly, we conclude that a negative β, favoring simple causes with shorter shape sequences, is essential to account for the overall main effect of better performance for infants in the C condition relative to ABC.

We also examined the influence on model performance of the second model parameter, θ, the probability of correct shape-matching. We assume that infants are able to identify shapes better (i.e., adopt higher θ values) with increasing age. Figure 6 depicts the model performance as a function of matching probability when $\beta = -1$ (preference for shorter causes). The model exhibits better performance with the increase of shape matching probability for the C condition,

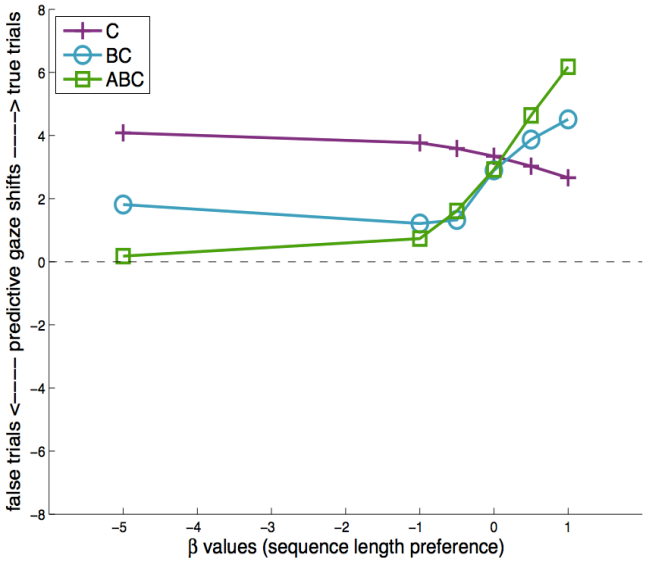

Figure 5: Model results. The expected PGS difference scores as a function of β values: causal sequence length bias. $\beta > 0$ favors longer causal sequences; $\beta < 0$ favors shorter causal sequences ($\beta = 0$ is uniform). This simulation assumes the other model parameter $\theta = 1$ (i.e., perfect shape identification and memory).

consistent with the significant age effect observed for infants in the C condition. For the BC condition, a smaller improvement was revealed by the model, in agreement with the marginal age effect in infant performance. In contrast, the ABC condition did not show monotonic improvement over the range of matching probabilities, and also exhibited

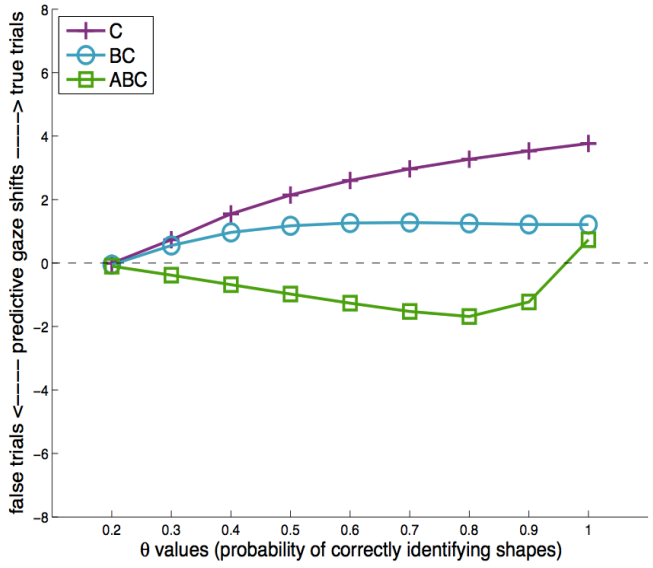

Figure 6: Model results. The expected frequencies of PGS difference as a function of θ values: the matching probability for shape identification ($\theta = 0.2$ is chance). This simulation assumes the other model parameter $\beta = -1$ (i.e., favoring simple causes with shorter shape sequences).

performance at or below chance level (i.e., same frequency of PGS for predictive and non-predictive sequences) for all values except $\theta = 1$, when shapes were perfectly recognized and memorized. Such chance-level performance in the ABC condition when the shape identification process is noisy is consistent with the observed performance of the infants, suggesting the infants' failure to learn in the ABC condition may be due to their noisy perception and limited attention and memory for identifying and remembering the shapes.

General Discussion

The experiment and model provide an important first step toward assessing the developmental changes in the second year of infancy that increase sensitivity to more statistically complex cause-effect relations. What developmental mechanisms could be at work? Our experimental results imply that performance was constrained by statistical complexity, in particular computation of conditional probabilities in the BC and ABC causal sequences, and our modeling results in turn imply that these computations may be constrained by difficulties identifying individual items and remembering their ordering in sequence. It remains for future studies to examine these possibilities, and to pinpoint the ages at which infants are able to track item identity and ordering so as to examine their contributions to detecting multiple conditional probabilities in predicting an effect.

It might be objected that the task we used is solvable by purely associative properties (i.e., statistical regularity). However, it is unclear whether standard associative accounts would be able to explain why performance was better in the C condition than in the other two conditions, since the statistical regularity was comparable across all three conditions. As previous research has shown, statistical regularity is necessary but not sufficient for causal inference: the learner must also have an *a priori* notion of causes (Cheng, 1997) and prior knowledge to enable efficient inference of causal relations (Griffiths & Tenenbaum, 2009; Lu et al., 2008).

The current study provides converging evidence for the importance of generic causal prior knowledge and highlights constraints based on attention, perception, and memory in guiding high-level causal reasoning in infancy. The novel experimental paradigm goes beyond the limitations of previous statistical learning and causal perception research, offering a promising new approach to determine how these processes develop in infancy. Planned future experiments using this paradigm, along with corresponding extensions of the Bayesian model, will contribute to better understanding of this developmental trajectory.

Acknowledgments

Special thanks to all of the infants and their parents, and the members of the UCLA Baby Lab. Patricia Cheng provided helpful suggestions. Funding for this project was provided by NIH grant R01-HD73535.

References

Balleine, B. W., Espinet, A., & González, F. (2005). Perceptual learning enhances retrospective revaluation of conditioned flavor preferences in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(3), 341-350.

Bonawitz, E. B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A. N., Woodward, J., & Schulz, L. E. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition*, *115*(1), 104-117.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*, 331-340.

Cheng, P. W. (1993). Separating causal laws from casual facts: Pressing the limits of statistical relevance. *The Psychology of Learning and Motivation*, *30*, 215-264.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367-405.

Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental Science*, *12*(4), 504-509.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205-1222.

Gredebäck, G, Johnson, S. P., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*, 1-19.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334-384.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. *Cognition*, *25*(3), 265-288.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*, 955-984.

Michotte, A. E. (1963). *The perception of causality*. New York: Basic Books. (Original work published 1946).

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

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21.