CHILD DEVELOPMENT

Child Development, November/December 2018, Volume 89, Number 6, Pages e507–e519

Infants Distinguish Between Two Events Based on Their Relative Likelihood

Ezgi Kayhan
Max Planck Institute for Human Cognitive and Brain Sciences and Radboud University Nijmegen

Gustaf Gredebäck and Marcus Lindskog
Uppsala University

Likelihood estimations are crucial for dealing with the uncertainty of life. Here, infants' sensitivity to the difference in likelihood between two events was investigated. Infants aged 6, 12, and 18 months (N = 75) were shown animated movies of a machine simultaneously drawing likely and unlikely samples from a box filled with different colored balls. In different trials, the difference in likelihood between the two samples was manipulated. The infants' looking patterns varied as a function of the magnitude of the difference in likelihood and were modulated by the number of items in the samples. Looking patterns showed qualitative similarities across age groups. This study demonstrates that infants' looking responses are sensitive to the magnitude of the difference in likelihood between two events.

One of the most important features of the human mind is its ability to make inferences and generalizations from sparse data (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). It has been proposed that our brain accomplishes this difficult task by using probability information to represent statistical regularities in our environment and guide our actions (Clark, 2013; Knill & Pouget, 2004). Thus, revealing the origins of how probability information is processed is crucial to understanding how the human mind works. Although much is known about how adults process probability information, less is known about how these abilities develop during infancy. Here, we developed a novel eye-tracking paradigm to investigate how infants of different ages respond to the likelihood of two events presented simultaneously while the relative difference in likelihood between the two samples changes.

Brunswik (1955) was probably the first to suggest that people are "intuitive statisticians" (see also Gigerenzer & Murray, 1987; Peterson & Beach, 1967). The metaphor alludes to an ability to correctly estimate statistical properties in the environment and draw accurate statistical inferences. Recent studies have investigated whether infants are also intuitive statisticians in the sense that they make inductive inferences from a small set of data (Denison, Reed, & Xu, 2013; Denison & Xu, 2014; Lawson & Rakison, 2013; Téglás, Girotto, Gonzalez, & Bonatti, 2007; Téglás, Ibanez-Lillo, Costa, & Bonatti, 2015; Xu & Garcia, 2008). For example, Xu and Garcia (2008) conducted a series of experiments in which they presented 8-month-old infants with a box containing many (e.g., 70) red balls and a few (e.g., 5) white balls. In alternating trials, they drew matching and mismatching samples from the box. The matching samples contained mostly red (e.g., 4) balls and a few (e.g., 1) white balls, whereas mismatching samples had the reverse proportions. Data revealed that infants inferred the probability of the sample given the population (i.e., the likelihood of the sample) and looked longer at the unlikely sample because this sample did not represent what they observed in the population.

Other studies have investigated whether infants already have expectations about single-event probabilities before having experienced any outcome (Lawson & Rakison, 2013; Téglás et al., 2007, 2011). For example, Téglás et al. (2007) presented 12-month-old infants with movies in which three yellow objects and one blue object randomly moved inside a container before one of them exited it.

We would like to thank the children and adults who participated in this study. We would also like to thank members of the Uppsala Child and Baby lab for valuable comments on earlier versions of the article. The European Union Seventh Framework Program Initial Training Network ACT (289404), KAW (2012.0120), and MMW (2015.0055) supported the present study.

Correspondence concerning this article should be addressed to Ezgi Kayhan, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1A, 04103 Leipzig, Germany. Electronic mail may be sent to kayhan@cbs.mpg.de.

DOI: 10.1111/cdev.12970

Results showed that infants looked longer when they observed an unlikely event (i.e., a blue object leaving the container) than when they observed a likely event (i.e., a yellow ball exiting the container). These findings suggest that infants expected that when the objects moved at random, one would likely observe one of the three identical objects falling out of the container rather than the single object.

These findings are crucial, as they suggest that infants infer the likelihood of events and use these inferences to form expectations about future events. This is a remarkable step in infant development especially because this ability potentially guides fundamental cognitive functions such as attention and learning. On a daily basis, infants face the challenge to learn new tasks and develop new skills. Despite distractors, they have to extract the most relevant information and use this information for further learning. If infants have the capacity to estimate how likely an event is as compared to the other events, they could distribute their attentional resources accordingly, which would then facilitate learning. Indeed, in a study by Tummeltshammer and Kirkham (2013), it has been shown that probabilistic relations between events affect infants' attention and enable infants to effectively guide their actions.

The general approach of the probabilistic inference studies in the literature is illustrated in Figure 1A. Here, infants are first familiarized with two populations (P_1 and P_2). Then, a sampling process (α, dashed line) produces a sample (S) from one of the populations (P_1 or P_2). For example, in the studies by Xu and colleagues, the experimenter draws (α) a sample of five balls (S) from the box containing many blue balls (P_1), and both S and P_1 are shown to the infant. The infant is then expected to combine the information from the sample with the information from the population and respond to either the probability of the sample given the population (i.e., $p(S|P)$), or the probability of the population given the sample (i.e., $p(P|S)$), depending on the experimental task.

As previous research suggests, infants estimate how likely an event is based on the population (or reference class) from which the event has been generated. Despite being an important capacity, estimating the likelihood of a single event is not sufficient to make good decisions in many real-world situations. It might be that in some situations children would also benefit from considering and comparing the likelihood of two or more events when making simple decisions. For example, it has been proposed that children develop intuitive theories in a way that can be described by a Markov Chain Monte Carlo search algorithm that explores a space of candidate theories (Ullman, Goodman, & Tenenbaum, 2012). To use such a process, infants would have to be able to compare the likelihood of the available candidate theories. Similarly, in order to distribute attentional resources to facilitate learning, as discussed earlier, infants would need to estimate how likely an event is as compared to other events.

If infants respond to the likelihood of single events, as previous research suggests, it is possible that they could also infer the likelihoods of two or more events presented simultaneously and show sensitivity to a difference in likelihoods. Recent work has begun to demonstrate that indeed children can discriminate between two simultaneously presented samples given the population from which the samples were drawn (Denison & Xu, 2014; Waismeyer, Meltzoff, & Gopnik, 2015). Denison and Xu (2014), for example, presented 10- to 13-month-olds with two jars of lollipops. The two jars included both desirable and undesirable lollipops, but in different proportions. The experimenter drew a lollipop from each jar and placed them in separate cups, after which the infant was encouraged to crawl to the cup to get the lollipop of their choice. Over four experiments, Denison and Xu (2014) showed that the infants consistently reached the cup containing a lollipop that was taken from the jar with the most favorable ratio of desirable to undesirable lollipops. These findings indicate that infants can simultaneously keep representations of two separate likelihoods, and have their actions be guided by these representations. Similarly, Waismeyer et al. (2015) showed that infants could consider two likelihoods simultaneously to make an inference about which of the two items is more likely to activate a machine.

An interesting question arising from these studies is to what extent infants' responses are influenced by the difference between the likelihoods of the two events. In the Denison and Xu (2014) study, for example, it is three times more likely (.75/.25 = 3) that the preferred lollipop is drawn from the jar with the most favorable ratio than from the other jar. Would infants respond differentially if the absolute or relative difference between likelihoods were to change, and at what point would they become indifferent? For example, would the proportion of infants choosing the most favorable jar have increased if the relative likelihood were 10 instead of 3? Beginning to answer such questions is

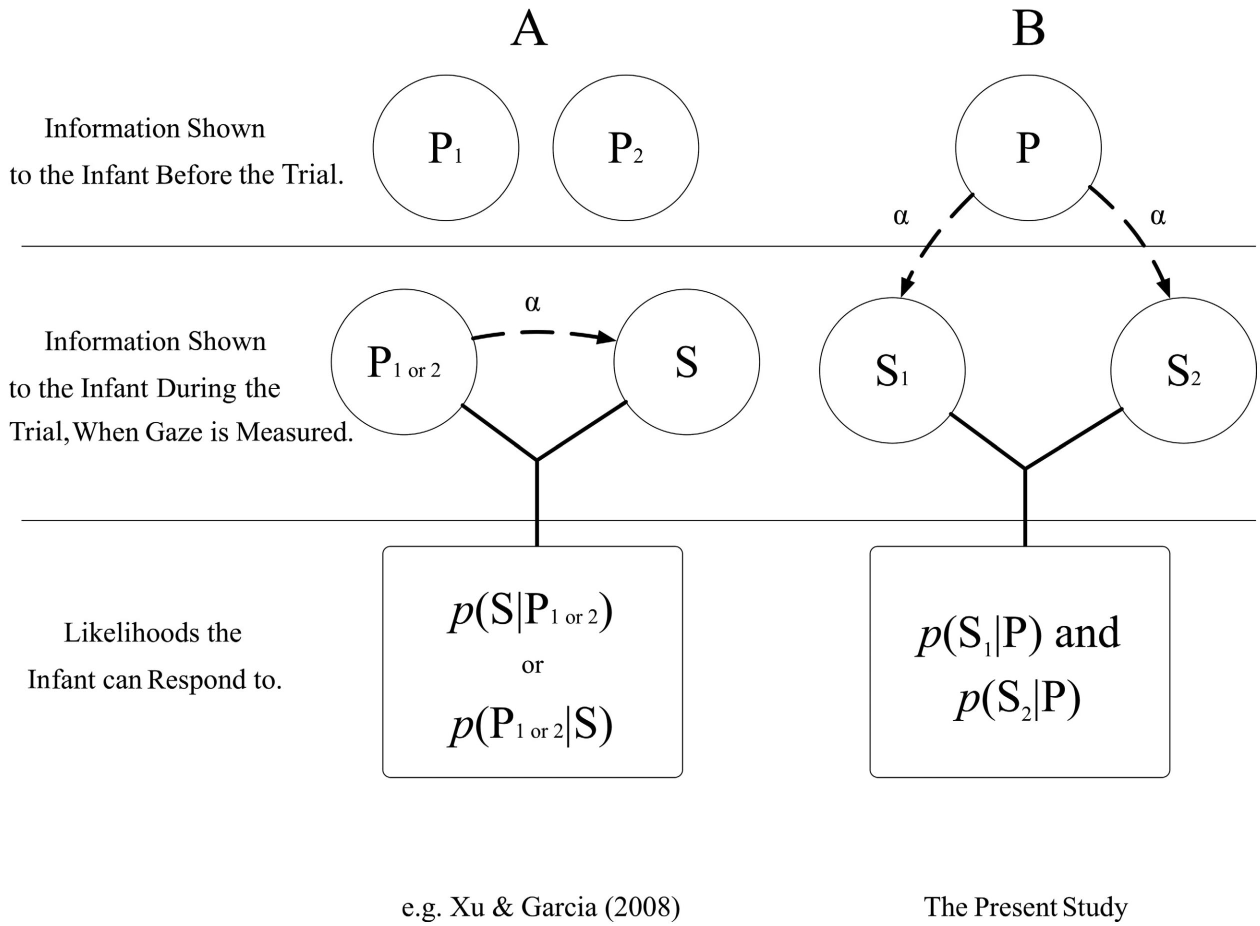

Figure 1. Two approaches to studying probability estimations in infants. Figure 1A illustrates a situation in which a population (P) has generated a sample (S) with some process (α, dashed line) after the infant has been familiarized to two different populations (P_1 and P_2; see Xu & Garcia, 2008). Infants' task is to combine the information in one of the populations $P_{1 \text{ or } 2}$ and a sample S and respond to the magnitude of p (S|$P_{1 \text{ or } 2}$) or p ($P_{1 \text{ or } 2}$|S), as indexed by their looking time. In Figure 1B (the present study), sampling process α has generated two samples from the same P. Infants' task would be to combine the information in S_1 and S_2, when P is no longer available, and respond to the magnitude of $p(S_1|P)$ and $p(S_2|P)$, as indexed by their relative looking time to two outcomes.

an important piece of the puzzle for understanding how infants' probabilistic reasoning develops. Doing so, however, requires a task where infants' responses can scale as a function of the difference between likelihoods.

Figure 1B conceptually illustrates such a task, where an infant could respond differentially to the likelihood of two separate events. Here, a sampling process (α) has generated two outcomes (S_1 and S_2) from a population (P). This would be equivalent to drawing two samples from the population box simultaneously. A situation where the individual estimates of likelihood need to be generated from a memory representation of the population, similar to the study by Denison and Xu (2014), is accomplished by covering up the population before the samples are shown. Now, the infant's task would be to combine the information in S_1 and S_2, with a memory representation of P, to estimate the likelihood of S_1 and S_2, given P (i.e., $p(S_1|P)$ and $p(S_2|P)$). Indexing infants' responses in terms of looking times should result in longer looking times to one of the samples if infants are indeed able to integrate a memory representation of the population with the information in the two samples and use this to distinguish between two events based on a difference in likelihood. Furthermore, if infants are sensitive to the magnitude of the difference, the proportion of looking time to one of the samples should increase with an increasing difference. However, if infants do not distinguish between the two events based on their likelihood, we would expect them to look equally long at the two samples.

In the current study, we explored this novel situation in an eye-tracking paradigm and investigated whether infants differentially respond to the likelihood of two samples when the population from which they were drawn is no longer visible. We presented infants with a population of balls comprising two colors. After covering the population, two samples, one likely and one unlikely with respect to the population, were drawn and presented to the infants as we measured their relative gaze duration.

We targeted three auxiliary aspects of the task. First, using a within-subjects design, we investigated whether infants' looking times scale continuously to the differences between event likelihoods by manipulating the magnitude of the relative difference in likelihood of the two events. Previous research has indicated that infants look more toward an unlikely than a likely event when events are presented separately (e.g., Xu & Garcia, 2008). If infants represent the likelihood of two events presented simultaneously in a similar manner, and if their responses are influenced by the difference in likelihood between the two events, here measured in relative terms, we predicted a gradual monotonic change in looking preferences for the unlikely event over the likely event when the relative likelihood changes (i.e., the looking preference is a function of relative likelihood that entirely increases or decreases). Put differently, infants should look more toward the unlikely than the likely event and even more so when the relative difference between the two likelihoods is larger. We graded likelihood information to ensure that our dependent measure actually targeted the key concepts under investigation instead of other low-level features, such as luminance, color, or contrasts (see Aslin, 2012 with respect to infant eye tracking and Xu & Garcia, 2008 and Xu & Denison, 2009 with respect to probability estimations).

Second, we investigated whether the number of objects in the samples would influence infants' responses. We compared a small sample set size (i.e., six items in each sample) to a large sample set size (i.e., eight items in each sample) while keeping the relative likelihood constant (see Table 1). The sample set size of six was motivated by previous research (e.g., Xu & Garcia, 2008) indicating that infants can estimate probabilities in a one-sample situation (Figure 1A) with this sample set size. We chose the sample set size of eight with two constraints in mind. First, the number of balls in this set should be greater than the number of balls in the small sample set while keeping the relative likelihoods the same. Second, the number of minority balls in the samples should not exceed the number of minority balls in the population, which would render a situation with impossible samples. It should be noted that the set size manipulation was primarily intended to investigate if different sample size were differentially processed. To avoid issues with infants' set size limitations, infants were not required to remember the samples, rather they were in full view once they had been presented.

Finally, even though previous research has indicated developmental differences with respect to infants' probability estimations (Denison et al., 2013), the developmental trajectory of infants' responses to the likelihood of several events has never been investigated. By using the same task to test 6-, 12-, and 18-month-old infants, we aimed to examine the developmental trajectory of this ability over the first one and a half years of life.

Method

Participants

Twenty-five 6-month-olds (M = 184.54 days, SD = 15.79 days; 14 girls), twenty-five 12-month-olds (M = 363.13 days, SD = 12.42 days; 10 girls), and twenty-five 18-month-olds (M = 551.75 days, SD = 10.41 days; 15 girls) were included in the

Table 1
Relative Likelihood of the Samples (625, 81, 25, and 9) Used in the Study, the Probability of the Likely and the Unlikely Samples, the Parameters Used to Create These Likelihoods (Sample and Population Ratios), and Sample Set Sizes

		Population ratio			
		1:5		1:3	
Sample set size	Sample ratio (likely/unlikely)	Probability (likely/unlikely)	Relative likelihood	Probability (likely/unlikely)	Relative likelihood
6	1:5/5:1	.4018/.0006	625	.3559/.0043	81
6	2:4/4:2	.2009/.0080	25	.2966/.0329	9
8	2:6/6:2	.2604/.0004	625	.3114/.0038	81
8	3:5/5:3	.1041/.0041	25	.2076/.0230	9

sample. One 6-month-old and one 12-month-old were tested but excluded from the sample during the analysis due to a lack of data (< 4 trials). Data were collected during October–December 2014. Participants were recruited from a database of volunteer families. Families received gift cards in return for their participation.

Stimuli

We created eight 20-s animated movies in 1,280 × 1,024 screen resolution using Anime Studio Pro (Figure 2A–C). As shown in Figure 2A, the infants were presented with the population when colored balls (diameter 0.5 visual degrees) fell from the top of the screen into a rectangle-shaped container (6.5 × 5.3 visual degrees in the form Width × Height for a viewing distance of 60 cm). Once all the balls piled up inside the container (8 s), the population container was covered with a light gray occluder (2 s). Next, infants' attention was drawn to the bottom of the container (Figure 2B) by a flashing light accompanied by a siren sound (2 s). Immediately after, two lids at the bottom of the population container were opened where two samples of balls were released into smaller separate containers (4.6 × 2.9 visual degrees) at the bottom of the screen (2 s). In the following 3 s, the samples were completely visible to the infants to compare them, as the balls reached the small containers and remained static in the containers (3 s) until the end of the movie (Figure 2C). The total time between the appearance of the first ball in the sample and the end of the trial was 8 s. The total number of balls and the proportions of the two colors in the population and samples for each condition are described in Table 1. Infants were presented with a new population and two new samples on each trial without any familiarization period. An example stimulus movie and descriptive statistics for the first 12 s of the stimulus

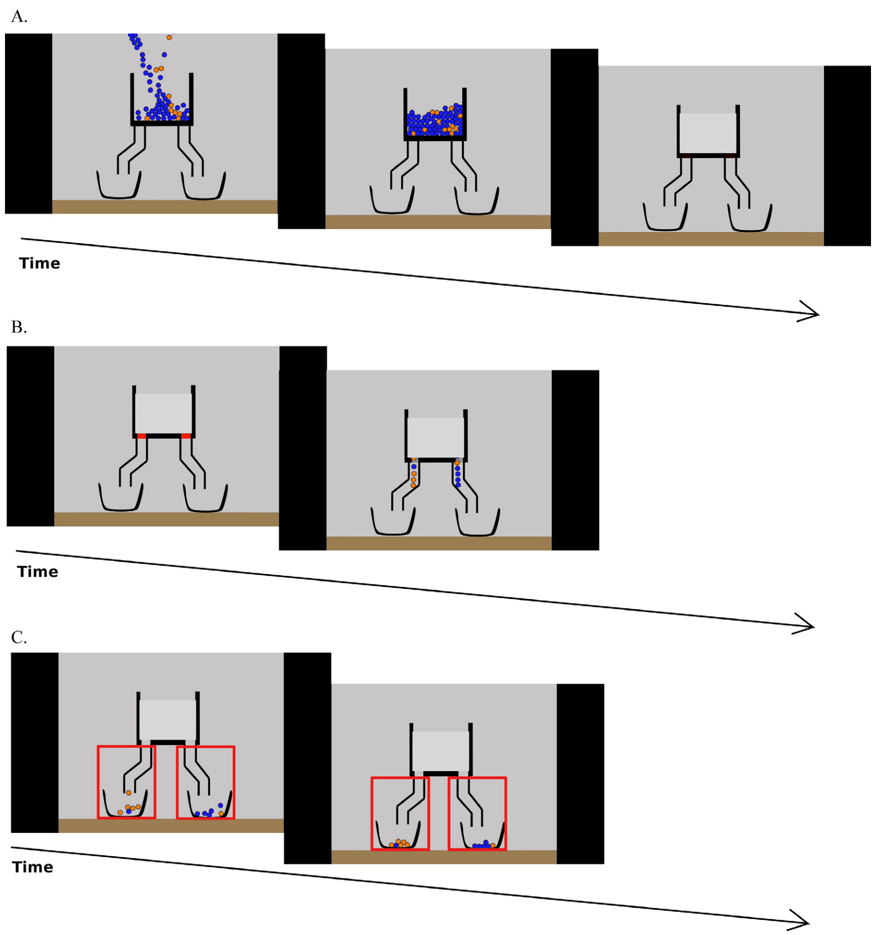

Figure 2. Snapshots from an example stimulus movie. Figure 2A shows the ratio of the balls in the population box and the occlusion of the box. Figure 2B demonstrates the sampling event. Figure 2C presents sampled outcomes. The red rectangles in Figure 2C illustrate the approximate position and size of the areas of interest used during data analysis.

presentation can be found in the Supporting Information.

The colors of the balls were changed in different trials to keep the infant's attention on the events and avoid potential confounding of specific colors. The direction from which the balls entered the scene during the presentation phase, the color of the minority and majority balls, and the location of the likely and unlikely samples were counterbalanced across movies. The positions of the balls in the population and samples were pseudorandomized for each movie. Each infant observed the same stimulus movies in a counterbalanced order. Short unrelated attention grabbers (each lasting for 5 s) separated each movie. Movies were identical in terms of the timing of the events in order to ensure that infants had the same amount of exposure to each stimulus movie. The entire stimulus presentation took 7 min (i.e., sixteen 20 s movies each separated by 5 s attention getters).

Experimental Design

The total number of balls ($n = 72$) in the population was kept constant across movies. We manipulated the relative likelihood of the two samples within subjects by changing the proportion of the two colors in the population and the samples to obtain four levels of relative likelihood: 625, 81, 25, and 9 (see Table 1). Relative likelihood was calculated as $p(S_1|P)/p(S_2|P)$ with S_1 and S_2 being the likely and unlikely sample, respectively. Thus, a relative likelihood of 625 means that the likely sample is 625 times more likely to be randomly drawn from the population than the unlikely sample (e.g., the relative likelihood for a 1:5/5:1 sample ratio drawn from a 1:3 population ratio was calculated as follows: $(3/4)^5 \times (1/4)^1 \div (1/4)^5 \times (3/4)^1 = 81$).

We also manipulated the total number of balls in the samples to utilize a large ($n = 8$) and small ($n = 6$) sample set. Each participant was presented with two trials for each stimulus, resulting in a total of 16 trials (2 × 4 likelihood levels [relative likelihood of 625, 81, 25, and 9] × two sample set sizes [6 and 8 items in each sample]). The four likelihood levels were created by varying both the ratio in the samples and the ratio in the population, as described in Table 1.

Experimental Setup and Procedure

Gaze data were recorded at 120 Hz by a corneal reflection eye tracker (Tobii 120; Tobii Technology, Danderyd, Sweden) calibrated using a 5-point procedure (Gredebäck, Johnson, & von Hofsten, 2010). The procedure was repeated if four or fewer calibration points were detected. The 16 stimulus presentations were intermixed with brief attention grabbing movies that served to redirect the infant's attention to the screen. Each experimental session lasted an average 30 min and consisted of an introduction, testing, and debriefing. Six- and 12-month-olds were seated in a Maxi Cosi car seat placed on their parent's lap. Eighteen-month-olds were seated directly on their parent's lap. All participants viewed the testing material from a distance of 60 cm.

Measures and Data Reduction

We measured the looking times toward the small containers for each trial. We excluded trials in which the total looking time for the entire screen was < 25% during the analysis interval. We determined one area of interest (5.6 × 7.5 visual degrees) for each container (Figure 2C). Initial data analyses of the final 6 s of the stimulus presentation revealed that infants lost interest in looking at the two samples approximately 3 s after they had been released, resulting in poor data quality and a large percentage of invalid data points. Therefore, we defined our time of interest as the first 3 s after the balls were released and visible.

Data Analysis

Due to the incomplete data matrix (23% of the trials were missing for 6-month-olds, 14% for 12-month-olds, and 8% for 18-month-olds) and the standard assumption of sphericity likely not being met, we chose to analyze the current data set using a general linear mixed model (GLMM) instead of the standard approach of repeated measures analysis of variance. See Hoffman and Rovine (2007) for details about the benefits of using GLMM for incomplete data sets and assumptions of sphericity arguments.

For statistical analyses, we calculated a proportion score (PS) for each trial as the proportion looking time spent looking at the unlikely sample over the total looking time spent looking at both likely and unlikely samples. Data processing was performed in MATLAB (MathWorks, Friedrichsdorf, Germany) using the open source analysis tool TimeStudio (Nyström, Falck-Ytter, & Gredebäck, 2016, version 1.1.). The actual analysis, settings, and source code for our analysis can be downloaded from uwid: ts-29d-55d within the TimeStudio environment.

Results

All GLMMs were fitted using the lme() function from the nlme package in R (Pinheiro, Bates, DebRoy, & Sarkar, 2015). To ease the interpretation, we present the effects of the fixed factors in each model evaluated with conditional *F* tests. In addition, the details of the full model, including all predictors, are shown in Table 2.

Age, Likelihood, and Sample Set Size as Predictors

First, we analyzed the effects of age, relative likelihood, and sample set size on PS using a GLMM with age, relative likelihood, and sample set size as fixed factors and participants as a random factor. This analysis revealed a significant two-way interaction between relative likelihood and sample set size, $F(3, 221) = 13.90$, $p < .001$, and a significant main effect of relative likelihood, $F(3, 208) = 5.97$, $p < .001$. No other effect reached significance (Fs < 2 and ps > .16). Initially, we also included order as a fixed factor in the model. Because the data revealed no order effects, this factor was accordingly dropped from the model to ease the presentation.

Table 2
*Beta Weights (β), Standard Errors (*SE*), and* p *Values (*p*) for the Fixed Factors in the Full General Linear Mixed Model With Age, Relative Likelihood (RL), and Sample Set Size as Predictors*

Effect	β	*SE*	*p*
Age: 12 months	.10	.08	.49
Age: 18 months	.01	.08	.93
RL: 81	−.09	.08	.50
RL: 25	−.34	.08	.01
RL: 9	−.31	.08	.02
Set size: large	−.42	.08	.01
Age: 12 Months × RL: 81	−.05	.11	.69
Age: 18 Months × RL: 81	.06	.11	.58
Age: 12 Months × RL: 25	−.08	.11	.48
Age: 18 Months × RL: 25	−.10	.11	.38
Age: 12 Months × RL: 9	−.10	.11	.38
Age: 18 Months × RL: 9	−.06	.11	.62
Age: 12 Months × Set Size: Large	−.02	.11	.88
Age: 18 Months × Set Size: Large	.08	.11	.60
RL: 81 × Set Size: Large	.20	.11	.17
RL: 25 × Set Size: Large	.42	.11	.00
RL: 9 × Set Size: Large	.31	.12	.03
Age: 12 Months × RL: 81 × Set Size: Large	−.06	.16	.65
Age: 18 Months × RL: 81 × Set Size: Large	−.09	.15	.43
Age: 12 Months × RL: 25 × Set Size: Large	−.02	.16	.84
Age: 18 Months × RL: 25 × Set Size: Large	.05	.16	.70
Age: 12 Months × RL: 9 × Set Size: Large	.06	.16	.62
Age: 18 Months × RL: 9 × Set Size: Large	.02	.16	.87

When the dependent variable is a proportion calculated from trial to trial, it is under some conditions more appropriate to use an empirical logit transform on the proportion measure before applying a linear regression. We therefore also ran all of our analysis on the empirical logit transformed proportion measure. These additional analyses gave highly similar results and did not change any of the conclusions driven from the untransformed data.

The Effect of Age and Likelihood in the Small and Large Sample Set Size Conditions Separately

The interaction between relative likelihood and sample set size (Figure 3) was further explored by examining each condition separately using two models with age and relative likelihood as fixed effects and participants as a random effect.

For the large sample set size, we found no significant effects (ps > .35). For the small sample set size, we found a significant main effect of relative likelihood, $F(3, 176) = 21.3$, $p < .001$. As illustrated in Figure 3, this main effect might be because infants' looking preferences for the unlikely event compared to the likely event decreased, as the relative

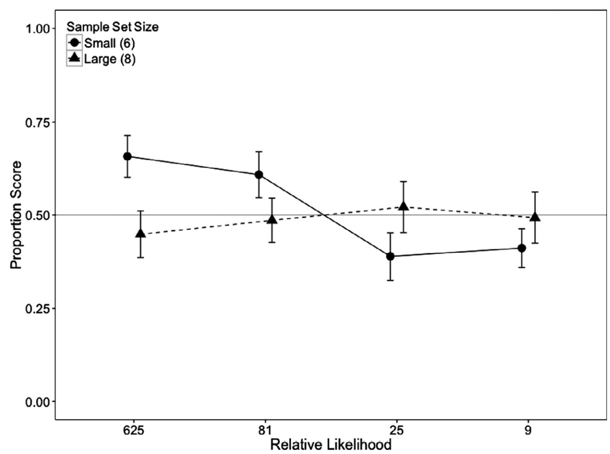

Figure 3. Mean proportion score for the small (circle) and large (triangle) sample set sizes as a function of relative likelihood pooled over the three age groups. Proportion scores above 0.50 indicate looking longer at unlikely sample, whereas values below 0.50 represent looking longer at the likely sample. The gray horizontal line represents chance. Vertical bars represent 95% confidence intervals.

likelihood of the two samples decreased, suggesting a monotonic change in the infants' looking preferences.

None of the other effects in the small (Fs < 1.0) or large (Fs < 1.05) sample set size reached significance. Follow-up post hoc tests (Tukey) showed that the pairwise difference between the 625 and 25 conditions was the only one that reached significance (p = .03), though the differences between the 625 and 9 conditions approached significance (p = .06; all others ps > .12).

The Effect of Likelihood in the Small Sample Set Size for Each Age Separately

No significant effect of age was observed in any of the above analyses. However, we wanted to ensure that the effect of relative likelihood in the small sample set was not driven primarily by one of the three age groups. Therefore, for the small sample set, we examined each age group separately using models including relative likelihood as a fixed effect and participants as a random effect. For all three age groups we found a significant main effect of relative likelihood, 6-month-olds: $F(3, 52) = 3.1$, $p = .03$; 12-month-olds: $F(3, 58) = 7.2$, $p < .001$; 18-month-olds: $F(3, 66) = 14.4$, $p < .001$, indicating that all three age groups exhibited the main effect of relative likelihood (Figure 4).

Discrimination of Likelihoods

The pattern of results illustrated in Figure 4 indicates that, although all three age groups discriminated between the two samples at a relative likelihood of 625 (confidence intervals do not overlap .5), only the 18-month-old infants did so for the other three levels of relative likelihood. Supplementary analysis with single sample t tests revealed that the PS for the 18-month-olds was significantly different from .5 for all four levels of relative likelihood. However, at 6 and 12 months of age, only the largest relative likelihood differentiated from .5 (only marginally at 6 months), suggesting more individual variability at younger ages. The t–test results were corrected for multiple comparisons within age groups according to the method proposed by Benjamini and Hochberg (1995).

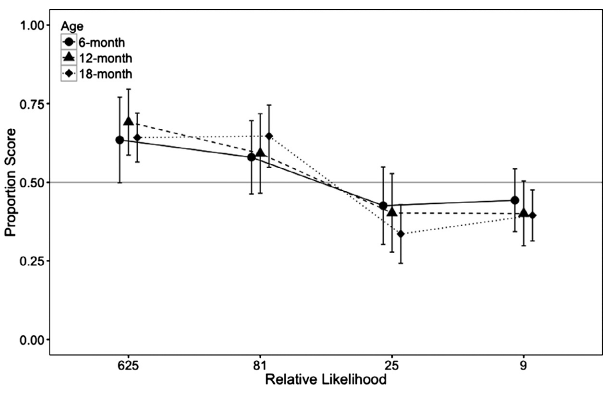

Figure 4. Mean proportion score in the three age groups as a function of relative likelihood. The gray horizontal line denotes indifference between the two events. Proportion scores above 0.50 indicate looking longer at unlikely sample, whereas values below 0.50 represent looking longer at the likely sample. Vertical bars represent 95% confidence intervals.

Discussion

Estimating the likelihood of two events simultaneously is an important ability when dealing with the uncertainty of life. Previous research has shown that infants form expectations about the probability of single events and are surprised when their expectations are violated (Téglás et al., 2007; Xu & Garcia, 2008). Recent work (e.g., Denison & Xu, 2014; Waismeyer et al., 2015) also indicates that infants can consider two likelihoods simultaneously. Based on these findings, however, it was unclear whether infants' responses, when presented with two events at the same time, are influenced by differences in the likelihood between the events.

Here, we extended this research by introducing a novel task where infants' responses scale to the difference, in relative terms, between the likelihoods of two events. To approximate how such estimates are often framed in real-world situations and in previous paradigms (e.g., Denison & Xu, 2014), we tested infants with a task in which they had to integrate information from two visible samples with a memory representation of the population from which the samples were drawn.

To achieve our main goal, we developed an eye-tracking paradigm that implemented the approach illustrated in Figure 1B. If infants differentiated between the two samples based on their likelihood and if they responded to the magnitude of the relative difference between the two samples, we expected it to be evident in the time spent looking at the samples (i.e., relative looking time). Our data indicate that infants differentially responded to the difference in likelihood between two events and allocated more looking time to one of the two samples. This finding suggests two important conclusions. First, it provides novel insights into the scope of infants' abilities as "intuitive statisticians," particularly into the ecologically important task of comparing the likelihood of two events. Second, because the population is occluded after the presentation phase, alternative explanations such as the possibility that infants distinguish between the two samples by referring each of them visually to the population can be ruled out. This design feature alludes to the need for a mental representation of the statistical properties of the population. Research in other domains has shown that infants already form mental representations of objects moving behind an occluder (Gredebäck & von Hofsten, 2007) or the structure of a causal system (Sobel & Kirkham, 2006) during their first 6 months of life. Our results provide further evidence that infants can form an abstract representation of the statistical properties of a population that is no longer in sight and integrate this information with statistical properties of the two samples.

We also assessed whether infants' looking times scaled according to the magnitude of the relative likelihood. We expected a monotonic change in the infants' looking preferences for the unlikely event compared to the likely event as the relative likelihood changed. Our results indicate that infants looked differently at the two samples and their looking pattern changed monotonically as a function of the relative likelihood. Recently, Aslin (2012) noted that scaling, the fact that the dependent variable changes in a continuous manner in response to similar changes in the stimuli, is important to ensure that behavior reflects the processes under investigation. In this case, scaling makes it highly likely that infants' responses were affected by relative likelihood and not by low-level perceptual features, such as color, luminance, and contrasts (Aslin, 2012), which are also present in the task.

A closer look at the data revealed that all age groups looked longer at the unlikely sample when the relative likelihood was large (i.e., 625), which replicates previous findings (e.g., Xu & Garcia, 2008). However, as the relative likelihood between the samples decreased, the looking pattern altered. Although 6- and 12-month-old infants were indifferent to the two subsets at low levels of relative likelihood (i.e., 25 and 9), 18-month-olds exhibited a change in their looking preference: They looked significantly longer at the likely subset than the unlikely subset. Notably, the change in the looking pattern was qualitatively the same in all age groups. This finding was unexpected, though interesting, as we hypothesized that all age groups would perform at chance level when it is harder to differentiate between the likely and unlikely events (e.g., a relative likelihood of 25 and 9). Why is there a shift in looking behavior as the magnitude of the relative likelihood decreases? One possible explanation is that because of the increased complexity of the information, infants preferred to allocate their attention to the subset that looks familiar. The increased difficulty to dissociate the unlikely sample from the likely one or the other way around became more challenging for infants, as the relative likelihood between the two samples decreased, imposing more processing demands on infants while they responded to the likelihood information. In other words, when there is greater uncertainty about the relative likelihood of events, infants preferred to look at the subset that represents a previously observed set. This interpretation is in line with studies showing that infants pay more attention to familiar items as the complexity of the stimulus increases (Mather, 2013). In addition, this pattern of responses might be explained in the light of findings demonstrating that the probability of infants' looking away from a stimulus shows a u-shaped pattern as a function of the complexity: Infants look away more, both when the stimulus is too easy and when it is too difficult to process (Kidd, Piantadosi, & Aslin, 2012). Therefore, it might be that because of the increased level of difficulty to encode the relative likelihood between the likely and the unlikely samples when the difference between the two samples was large, infants preferred to look at the subset longer that looked familiar.

What are possible cognitive processes the infants could be engaged in when looking at the two samples to produce our observed pattern of data? As noted in the introduction, we aimed to manipulate the relative difference in likelihoods between the two samples. One possibility is that infants are indeed engaged in estimating the relative difference in likelihood between the two samples. Other interpretations of the current results are, however, also possible, making us cautious to draw firm conclusions about the nature of the information processing that resulted in the reported looking patterns. It could be that infants have a direct mapping between the likelihood of an event and an internal likelihood scale, and that their looking behavior is a direct function of how strongly this internal scale is activated. Under such a process infants do not need to make a direct comparison between the two samples. Rather they respond to each event separately with the looking time to that event being proportional to the likelihood of the event. Another possibility is a process that compares the magnitude of the two likelihoods of the events. This process could make a comparison between the likelihoods in either absolute or relative terms. Note that the latter of these two would be similar to several psychophysical processes where people can tell two stimuli apart based on the relative magnitude of their intensity (e.g., A is brighter than B) without having an accurate estimate of the intensity of each stimulus on its own (e.g., A shines with 300 lux). Although the design of the present study cannot tease these possibilities apart, it will be an interesting a venue for future studies to address the mechanisms explaining how infants' looking responses scale to differences in likelihoods.

It is also possible that infants did not respond to the likelihood information, per se, but rather responded to some more low-level perceptual feature of the task. One such possibility is that infants respond differentially to the relative likelihood of the two samples only when the two samples both contain a singleton. More specifically, for the relative likelihoods of 625 and 81, there is only one item in the subsets of the small sample set that stands out among the others (e.g., five blue balls and one yellow ball and vice versa). Thus, it could be argued that the unlikely subset is marked easily by infants and draws their attention because it is more salient. Indeed, what we observed was that the largest relative likelihood could be discriminated by even the youngest age group. However, the second largest relative likelihood elicited significantly longer looking times for the unlikely subset only in the oldest age group. If infants' responses were only driven by saliency, we would expect all age groups to look longer at the unlikely subset for the relative likelihoods of 625 and 81, not only the 18-month-olds. Moreover, if infants were looking longer at the unlikely subset due to perceptual features, we would expect these to dominate the responses of the younger age groups in particular (Althaus & Mareschal, 2012).

Another possibility is that infants relied on some form of exemplar matching and compared exemplars in the memory representation of the population with those present in the sample. We find such a strategy unlikely because it would require infants to keep track of a (very) large number of objects. Depending on the condition, there were 72 balls in the population and either 12 or 16 balls in total in the two samples. With respect to the limit on the number of objects infants can simultaneously track, it would be very challenging for infants to track the items in the current paradigm. It should be noted that no tracking is required if infants extract information about the proportion of the two types of balls. Another possible strategy is that infants look for individually colored balls in the two smaller containers based on either a familiarity or novelty preference in response to the proportions in the large container. This "look for X-colored balls" strategy could be used for one of two versions. In both cases, infants would have had to identify the dominant color in the population and would look for this color in the samples (or the nondominant in the case of novelty preference). In both the 1:5 and 1:3 population ratio conditions, this should be quite a reasonable task for the infant to achieve. One possibility is then that infants allocate more looking time to the sample that has more of the dominant color balls (or the other way around for a novelty preference). That is, if red is the dominant color in the population, the infant will look more at the sample with more red balls. If this is the case, it is not clear why the proportion of looking time would change as a function of the relative likelihood of the two samples. In all four conditions, there is one sample with more of the dominant color balls, which predicts the same proportion of looking time to the likely sample in all conditions. Another possibility is that the looking time is divided equally among all balls of the dominant color. Thus, in the condition with 1:5 and 5:1 samples, the infant would allocate approximately 1/6 of the looking time to the unlikely sample and 5/6 to the likely sample. In the other two conditions, this would be 1/3 and 2/3 of the time instead. This also implies that the looking time to the two samples would be independent of the ratio in the population. However, data revealed that although both 6- and 12-month-olds looked significantly more to the likely sample in the 625 condition, they did not do so in the 81 condition. Thus, it is unlikely that the general pattern of results can be explained with either an exemplar matching or a "look for X-colored balls" strategy. Thus, taken together, the data indicate that it is unlikely that low-level perceptual features are the *only* factors driving infants' looking behavior. A surprising but interesting finding of the current study is that infants' responses to relative likelihood were modulated by the sample set size.

More specifically, infants' looking preferences varied as a function of relative likelihood but only when the sample set size was 6. One reason for the differences in infants' responses might be that observing the large set placed more processing demands on infants. More specifically, although there were more items in the samples in the large set condition, infants had the same amount of time to process these items as they had for the small sets. Therefore, it might be that this particular situation was too challenging for infants to compare the likelihoods of the two samples in the large sample set condition. However, although the larger sample set introduced more complexity to processing the items, we expected the infants to distinguish between events in the two conditions with the largest relative likelihood. More specifically, previous research has indicated that infants use either an approximate number system (Feigenson, Dehaene, & Spelke, 2004) or an object tracking system (Feigenson, 2005) to process sets of objects, and we expected that infants would engage one of these systems at some point during the evaluation of the samples. Both systems, however, have limitations and break down if certain conditions for the sets are not met. Accordingly, increasing the size of the samples could have led to such conditions being violated. However, we expected neither the approximate number system nor the object tracking system to impose limitations on the processing of the two samples. We did not expect the approximate number system to restrict infants' processing, as the within-sample ratio is well within the capacity of infants of this age. Similarly, the object tracking system should not impose limitations on the processing of the items because the number of balls of each color in each sample is low enough to be tracked as ensembles by an object tracking system. Although previous research has suggested that the ability to represent probabilities draws on an ability to represent numerical magnitudes (Téglás et al., 2015), it is unclear whether the limitations of an approximate number system object tracking system or possibly both underlie the differences in infants' responses to the small and large sample sets observed here.

It should also be noted that our initial data analyses showed that infants lost interest in the stimuli after approximately 3 s. Hence, we used the first 3 s as our analysis interval. This is somewhat shorter than the 5–11 s reported in, for example, Xu and Garcia (2008), or the 10–15 s, reported by Téglás et al. (2015). An explanation for this difference could be the use of a different methodology. We used neither a familiarization nor a habituation approach in which the same population was shown repeatedly to participants. It is possible that using familiarization and/or habituation creates a larger novelty effect that attracts the infant's attention to a larger degree.

One limitation of the current study is that the samplings of the balls were not performed truly at random. The positions of the balls in the population and samples were pseudorandomized for each movie and the direction from which the balls entered the scene during the presentation phase was counterbalanced across movies. Moreover, the balls moved at random, as they piled up inside the box. However, these measures might not be enough to rule out the potential influence of low-level factors, such as the spatial arrangement of the balls, which might have guided infants' expectations about the likelihood of events. As random sampling is one of the main assumptions of probabilistic inference, a follow-up study should control for this factor, for example, by adding a condition in which the box is shaken prior to sampling.

Finally, although previous research suggests some developmental differences in probability estimations by infants (Denison et al., 2013), our study is the first to test 6-, 12-, and 18-month-old infants with the same task in order to map out the developmental trajectory of likelihood comparisons. Although we found some quantitative differences between the age groups in their responses to relative likelihoods, the looking time patterns of different age groups were very similar at a qualitative level. Besides mapping out the scope of probability estimations across a variety of age groups in infancy, one important contribution of our findings is that they corroborate previous work in the literature showing that infants as young as 6 months of age are sensitive to the relative probability between two events (Denison et al., 2013).

In summary, we investigated whether infants of different ages distinguish between two events based on their relative likelihood and whether their responses scale to the magnitude of relative likelihoods. Using eye-tracking methodology, we showed that infants' looking behavior varied as a function of relative likelihood and is modulated by the number of items they observe. In addition, we demonstrated that younger age groups distinguish between likely and unlikely events only when the relative likelihood between the two events is sufficiently large. In contrast, older age groups also have differential responses to smaller relative likelihoods, although looking patterns were qualitatively the same in all age groups. Future studies should address the cognitive and neural systems governing infants' reactions to the graded complexity of relative likelihoods.

References

Althaus, N., & Mareschal, D. (2012). Using saliency maps to separate competing processes in infant visual cognition. *Child Development*, *83*, 1122–1128. https://doi.org/10.1111/j.1467-8624.2012.01766.x

Aslin, R. N. (2012). Infant eyes: A window on cognitive development. *Infancy*, *17*, 126–140. https://doi.org/0.1111/j.1532-7078.2011.00097.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300. https://doi.org/10.2307/2346101

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217. doi: http://dx.doi.org/10.1037/h0047470

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. https://doi.org/10.1017/S0140525X12000477

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5-and 6-month-olds. *Developmental Psychology*, *49*, 243. https://doi.org/10.1037/a0028278

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*, 335–347. https://doi.org/10.1016/j.cognition.2013.12.001

Feigenson, L. M. (2005). A double-dissociation in infants' representations of object arrays. *Cognition*, *95*, B37–B48. https://doi.org/10.1016/j.cognition.2004.07.006

Feigenson, L. M., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*, 1–19. https://doi.org/10.1080/87565640903325758

Gredebäck, G., & von Hofsten, C. (2007). Taking an action perspective on infants' object representations. *Progress in Brain Research*, *164*, 265–282. https://doi.org/10.1016/S0079-6123(07)64015-1

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 101–117. https://doi.org/10.3758/BF03192848

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*, e36399. https://doi.org/10.1371/journal.pone.0036399

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*, 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Lawson, C. A., & Rakison, D. H. (2013). Expectations about single event probabilities in the first year of life: The influence of perceptual and statistical information. *Infancy*, *18*, 961–982. https://doi.org/10.1111/infa.12014

Mather, E. (2013). Novelty, attention, and challenges for developmental psychology. *Frontiers in Psychology*, *4*, 1–4. https://doi.org/10.3389/fpsyg.2013.00491

Nyström, P., Falck-Ytter, T., & Gredebäck, G. (2016). The Time Studio Project: An open source scientific workflow system for behavioral and brain sciences. *Behavior Research Methods*, *48*, 542–552. https://doi.org/10.3758/s13428-015-0616-x

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*, 29–46. https://doi.org/10.1037/h0024722

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D.; R Core Team. (2015). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-122. Retrieved from http://CRAN.R-project.org/package=nlme

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*, 1103. https://doi.org/10.1037/0012-1649.42.6.1103

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 19156–19159. https://doi.org/10.1073/pnas.0700271104

Téglás, E., Ibanez-Lillo, A., Costa, A., & Bonatti, L. L. (2015). Numerical representations and intuitions of probabilities at 12 months. *Developmental Science*, *18*, 183–193. https://doi.org/10.1111/desc.12196

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*, 1054–1059. https://doi.org/10.1126/science.1196404

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. https://doi.org/10.1126/science.1192788

Tummeltshammer, K. S., & Kirkham, N. Z. (2013). Learning to look: Probabilistic variation and noise guide infants' eye movements. *Developmental Science*, *16*, 760–771. https://doi.org/10.1111/desc.12064

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*, 455–480. https://doi.org/10.1016/j.cogdev.2012.07.005

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science*, *18*, 175–182. https://doi.org/10.1111/desc.12208

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*, 97–104. https://doi.org/10.1016/j.cognition.2009.04.006

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 5012–5015. https://doi.org/10.1073/pnas.0704450105

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Table S1. Mean and Standard Deviation (*SD*) of Looking Times at the Population Box During the First 12 s of the Stimulus Movies (Before the Samples Are Visible)

Movie S1. An example stimulus movie for the relative likelihood of 625.