Running head: PROBABILISTIC REASONING IN YOUNG INFANTS

The emergence of probabilistic reasoning in very young infants:

Evidence from 4.5- and 6-month-olds

Stephanie Denison

Christie Reed

Fei Xu

University of California, Berkeley

In press, *Developmental Psychology*

Word Count: 4480

How do people make rich inferences from such sparse data? Recent research has explored this inferential ability by investigating probabilistic reasoning in infancy. For example, 8- and 11-month-old infants can make inferences from samples to populations and vice versa (Denison & Xu, 2010a; Xu & Denison, 2009; Xu & Garcia, 2008). The current experiment investigates the developmental origins of this probabilistic inference mechanism with 4.5- and 6-month-old infants. Infants were shown 2 large boxes, 1 containing a ratio of 4 pink to 1 yellow balls, the other containing the opposite ratio. The experimenter sampled from, e.g., the mostly pink box, and removed a sample of either 4 pink and 1 yellow balls or 4 yellow and 1 pink balls on alternating trials. Six- but not 4.5-month-olds looked longer at the 4 yellow and 1 pink sample (the improbable outcome) than the 4 pink and 1 yellow sample (the probable outcome).

Human learners are capable of making large inductive leaps in the face of data that are limited and often stochastic. It is an important and ubiquitous ability. For example, imagine a person from our hunter-gatherer ancestry trying to determine which types of trees produce berries that are good for eating. Let's say they sample approximately five berries from a couple of trees and find that one tree produces four good tasting berries and the other only produces one or two. They may make the inference that berries from the former tree tend to be edible and that the latter tree type should be avoided. Or, picture a toddler attempting to learn her first words. She hears the word "doggie" in the presence of her family dog a few times, and she quickly generalizes the word to other dogs but not a stray cat or her pet hamster.

What are the cognitive mechanisms that allow human learners to make such rapid and often highly accurate inductive inferences? Recent research in cognitive development has focused on the origins of probabilistic inference in infancy as a possible starting point. First, 12-month-old infants can make inferences from populations to samples when reasoning about single-event probability (Denison & Xu, 2010b; Teglas, Girotto, Gonzalez & Bonatti, 2007; Teglas et al., 2011). Second, 8-month-old infants are capable of making inferences from small samples to large populations and vice versa (Xu & Garcia, 2008). Third, infants as young as 11 months take into account the implications of sampling conditions (e.g., random vs. non-random sampling) and object properties (e.g., solidity and cohesion) when making these inferences (Denison & Xu, 2010a; Gweon, Tenenbaum & Schulz, 2010; Teglas et al. 2007; Xu & Denison, 2009).

In the current experiment, we explore the age at which infants begin to make inferences from samples to populations, looking for the first time at infants younger than 8 months. We ask whether very young infants can make basic probabilistic inferences using a variant of the paradigm first introduced by Xu and Garcia (2008). In their experiments, a looking-time paradigm was employed to reveal whether 8-month-old infants have an intuitive ability to make generalizations from samples to populations. In Experiment 1, infants were shown samples being drawn from a large covered box and, on alternating trials, the experimenter either removed 4 red and 1 white balls or 4 white and 1 red balls. Then the experimenter revealed the population of balls – a 9:1 ratio of red to white balls. Infants looked longer at the 4 white and 1 red ball sample (the improbable outcome) than the 4 red and 1 white ball sample (the probable outcome; see Figure 1).

Although this might suggest that infants have a rudimentary ability to reason about probability, the authors note two possible interpretations of this looking pattern: The first, which we will call the "probabilistic account", suggests that infants looked longer at the 4 white and 1 red balls sample because they understand the predictive relationship between samples and populations and thus they considered it to be a relatively improbable sample. The second, termed here the "perceptual mismatch account", suggests that infants simply prefer to look at displays wherein the population box and sample container contrast in perceptual appearance. That is, infants simply looked longer at trials displaying the less probable sample because it created a perceptual mismatch between the two displays present on stage (see the outcomes in Figure 1). This account represents a lower-level interpretation of infant performance, as it predicts an identical looking pattern as the probabilistic account but does not require an understanding of the relationship between the sample and population.[1]

To distinguish between these two interpretations, Xu and Garcia (2008, Expt. 3) designed a control experiment in which the 4:1 sample was no longer drawn from the population. Eight-month-olds participated in a procedure that was equivalent to the one just described except that the relationship between the sample and population was eliminated: the samples were drawn from the experimenter's pocket rather than from the box. This resulted in identical test trial displays to those in Experiment 1 but in this case, infants had no reason to expect a relationship between the contents of the small container

---

[1] Adults viewed the Expt. 1 displays and rated the improbable outcome as "unexpected" and the probable outcome as "expected." They did not note perceptual mismatches or probability in their explanations. This suggests that computations of probabilities may generally be largely implicit and inaccessible to conscious thought.

and the population box. Eight-month-old infants looked about equally when the mostly red box was displayed with the 4 red and 1 white balls sample (the perceptual match) and the 4 white and 1 red balls sample (the perceptual mismatch). Therefore, when the relationship between the box and container was eliminated, neither display violated infants' expectations. This provides evidence in favor of the probabilistic account of infants' performance in Experiment 1, i.e., that infants were reacting to the relative improbability of the sample and not the perceptual mismatch.

After obtaining evidence that 8-month-olds can reason about samples and populations, Xu and Garcia began to explore whether even younger infants possess similar intuitions. It seems plausible that younger infants could succeed at a version of this task given evidence revealing sensitivity to statistical input from newborns to 6-month-olds in domains such as phoneme discrimination and visual pattern learning (e.g., Bulf, Johnson, & Valenza, 2011; Kirkham, Slemmer & Johnson, 2002; Maye, Werker, & Gerken, 2002). Thus, they tested a group of 6-month-old infants using the same procedure as Xu and Garcia (2008). The findings were inconclusive. Infants performed as expected in the replication of Experiment 1, looking longer at trials in which the experimenter sampled 4 white and 1 red balls than 4 red and 1 white balls from the mostly red population. However, infants continued to follow this looking pattern in the control experiment during which the experimenter drew from her pocket (Xu & Garcia, unpublished data).

Although this pattern of findings does not support the probabilistic account, it also cannot definitively rule it out. It is possible that younger infants appreciate the relationship between samples and populations but also look longer at the perceptual mismatch in the control task because they continue to react to the perceptual features of the mismatch when the sample was drawn from the experimenter's pocket. Unfortunately the experimental design cannot tease apart these two interpretations.

In the current experiment, we use a design appropriate for testing probabilistic inference in younger infants wherein the perceptual mismatch is eliminated but the displays remain easy to process. We equated the overall quantity of each ball color present in the population boxes during test trials by keeping two complementary boxes on display throughout (see Figure 2). Each test trial began with the two covered population boxes and a small transparent container to hold a sample on stage. The experimenter drew the infants' attention to each box and drew a sample of, e.g., 4 pink and 1 yellow balls from the box on the right and placed it in the container. Then the experimenter revealed that the box from which the sample was drawn had a 4:1 ratio of pink to yellow balls, and the other box had the opposite ratio. The trials alternated between a 4 pink and 1 yellow sample (the more probable sample) and a 4 yellow and 1 pink sample (the less probable sample). If infants are only sensitive to perceptual mismatches and not sampling, they should look equally at all test trials, as the large boxes on display have equal amounts of each color and the sample therefore creates a slight but equal mismatch across every trial. If, on the other hand, infants are sensitive to the relationship between the sample and population, they will look longer on trials where the less probable sample is drawn from the relevant population box.

We tested both 4.5- and 6-month-old infants in this design because we did not have strong *a priori* predictions about the age at which this mechanism comes online. Although there is ample empirical evidence of statistical learning and probabilistic reasoning in the second half of the first year, experiments conducted on infants younger than 8 months are relatively scant. The most relevant findings with young infants are from research on visual statistical learning and conditional probability computations in early infancy. Evidence of visual statistical learning of transitional probabilities in infancy has been found in 2-month-old infants, and was recently extended to newborn infants (Bulf et al., 2011; Kirkham et al., 2002). On the other hand, investigations of conditional probability computations with young infants have revealed that this comes online much later, some time between 5 and 8 months (Sobel & Kirkham, 2006, 2007). Due to the differences in findings in these related abilities with young infants, 4.5- and 6-month-olds seem appropriate age groups with which to begin an investigation of rudimentary probabilistic reasoning.

Method

Participants

Participants were 32 infants: Sixteen 6-month-olds (6 males; *M* = 6;4 [months; days], *R*=5;15 to 6;17) and sixteen 4.5-month-olds (12 males; *M* = 4;15; *R* = 4;1 to 5;0). Ten infants (six 6-month-olds; four 4.5-month-olds) were tested but excluded due to fussiness (4), inattention during sampling (2), providing looking times over 3.5 standard deviations above average (1) or parental interference (3). Infants were recruited from the San Francisco Bay Area. Socioeconomic status and ethnicities reflected the general distribution of the area and infants were required to be exposed to English a minimum of 50% of the time. Infants received a small gift for their participation.

Materials

Ping-Pong Balls. A total of 166 (83 yellow and 83 pink) ping-pong balls were used.

Boxes and Containers. A small, transparent Plexiglas container with an open top (20 cm x 4.5 cm x 4.5 cm) was used to display the samples during test trials.

Two large (31 cm x 23.5 cm x 23.5 cm) boxes were used to display the populations during the familiarization and test trials. The boxes were rectangular cubes with Plexiglas windows to show the populations of ping-pong balls and a hidden center compartment to hold the samples that were removed from the box during test trials. From the infants' perspectives, the box appeared as one single unit, filled completely with ping-pong balls. The Plexiglas display windows were covered with fabric curtains to ensure that the boxes appeared identical when the curtains were lowered. The "mostly pink" box contained 60 pink and 15 yellow balls (pink to yellow = 4:1); the "mostly yellow" box contained the opposite (pink to yellow = 1:4).

Apparatus

Testing occurred in a room divided in half by curtains spanning its width and height. The curtains had a cut-out above a puppet stage that measured 94 cm x 55 cm (width x height). The experimenter sat behind the stage with her upper body and head visible. A curtain could be lowered to conceal the experimenter between trials. A camcorder filmed the infant through a small hole in the curtain below the stage; it was connected to a TV monitor that an observer used to code looking times online using JHAB (R. Casstevens, 2007). A second camcorder recorded the experimenter's behavior.

Infants sat in a high chair approximately 70 cm from the center of the stage. The parent sat next to the infant facing the opposite direction and was instructed to avoid looking at the stage.

Design and Procedure

Calibration. To calibrate each infant's looking window, a squeaky toy was used to direct the infant's attention to the outside parameters of the stage.

Free Play Phase. After calibration, the infant was shown a small open box with 3 pink and 3 yellow ping-pong balls. He/she was encouraged to play with the balls for approximately 30 seconds[2]. This was done to demonstrate to infants that the balls were discrete objects.

Familiarization trials (4 trials). The experimenter placed the two large boxes on the stage 30 cm apart with the front curtains down. She shook the box on the right side of the stage, saying, "What's in this box?" She then shook the box on the left saying, "What's in this box?" She lifted the front covers of both boxes simultaneously, revealing the separate populations of mostly pink and mostly yellow balls, and said "Look, [baby's name], look!" She then put her head down and directed her gaze to the floor. The observer began timing upon hearing the second, "look". Trials ended when the infant looked away for 2 consecutive seconds.

---

[2] Not all infants were capable of reaching in to grab the ping-pong balls, particularly the 4.5-month-olds. Videos were re-coded to assess how many infants manipulated the balls. 11/15 six-month-olds and 2/14 four and a half-month-olds manipulated the balls spontaneously (3 videos were inconclusive, due to the Free Play box obstructing the view of the infants' hands). For all infants, the experimenter picked up at least one ball of each color and held them out to the infant, touching the balls to various parts of their bodies, including their hands, arms or noses.

The boxes were presented in the same locations for all 4 familiarization trials within a single experimental session and positioning was counterbalanced across infants. Between familiarizations the boxes were removed from the stage and a back curtain was lowered. These trials were included to familiarize infants to the materials and to the general experimental procedure. Additionally, exposing infants to two contrasting populations might cue them to attend to ratios. The familiarizations lasted approximately 3 minutes**.**

Test trials (6 trials). On each test trial, the experimenter placed the two large boxes on the stage (keeping them in the same locations on all 6 trials) with the front curtains lowered. The experimenter always sampled from the box on her right. She shook each box one at a time while saying, “What’s in this box?” She then closed her eyes, turned her head away, and reached into the box on her right. She pulled out 3 ping-pong balls and placed them into the small Plexiglas container in the middle of the stage one at a time. She then closed her eyes, turned her head away, and reached into the box on her left (not pulling out any ping-pong balls) and placed her hand on top of the small Plexiglas container in the middle of the stage to mimic the sampling motions made with the box on the right. She then repeated these actions, pulling out 2 more ping-pong balls from the right hand box and mimicking this action with the left hand box. On alternating trials the sample removed from the population box was either 4 pink and 1 yellow balls or 4 yellow and 1 pink balls. Then the experimenter lifted the front covers of both boxes on the stage simultaneously and said “Look, [baby’s name], look!” She put her head down and directed her gaze toward the floor. The observer began timing upon hearing the second, “look”, and ended the trial after 2 consecutive seconds of looking away. Between trials, the stage was cleared. The test trials lasted approximately 6 minutes.

Design. The side that the population boxes (mostly pink or mostly yellow) were on and whether the infant saw the 4 pink and 1 yellow ball sample first were fully counterbalanced across infants.

Predictions

If infants are sensitive to the relationship between samples and populations (i.e., assuming random sampling, the composition of a sample is likely to reflect the overall composition of a population) they should look longer at test trials displaying outcomes that violate this expectation than outcomes that are in line with this expectation. Therefore infants who saw the experimenter sampling from the mostly pink population should look longer at trials in which 4 yellow and 1 pink balls were sampled than trials in which 4 pink and 1 yellow balls were sampled. Conversely, infants who saw the experimenter sampling from the mostly yellow population should show the opposite looking pattern.

Results

A second observer, blind to trial order, coded 50% of the infants offline. Interscorer reliability averaged 92%[3]. Preliminary analyses found no effects of gender, test trial order (probable-outcome vs. improbable-outcome first), or the population box sampled from (mostly pink or mostly yellow) for both age groups. There was also no difference in duration of looking on familiarization trials between the two age groups. Subsequent analyses collapsed over these variables.

Looking times for test trial outcomes were analyzed using a 2 x 2 ANOVA with outcome (probable vs. improbable) as the within-subjects factor and age (4.5-month-olds vs. 6-month-olds) as the between-subjects factor. A significant interaction between Outcome and Age was found, F (1, 30) =7.03, p = .013, effect size (ηp^2) = .190. There were no other significant main effects or interactions.

To break down the interaction, we conducted follow-up t-tests exploring the effect of test trial outcome (probable vs. improbable) for each age group separately (see Table 1 for mean looking times). Six-month-old infants looked reliably longer at the improbable outcome (*M*= 8.63s, *SD*= 5.05) than the probable outcome (*M*= 5.96s, *SD*= 2.81), $t(15) = 2.67$, $p = .011$, effect size (*d)* = 0.679. Twelve of sixteen infants looked longer at the improbable outcome, more infants than would be expected by

---

[3] Reliability was calculated as the proportion of total time both observers agreed that infants were looking at the displays (see Kellman & Spelke, 1983). Thus, percentage agreement = 1-((Absolute difference in time between original and second coder)/original coder). We then obtained an average across all 160 trials.

chance, binomial test, *p* = .038, 1-tailed. In contrast, 4.5-month-olds looked about equally at the improbable outcome (*M*= 6.05s, *SD*= 3.14) and the probable outcome (*M* = 7.45s, *SD*= 5.18), *t*(15) = 1.19, *p*= .250, effect size (*d)* = 0.321. Seven of sixteen infants looked longer at the improbable outcome, not different from chance, binomial test, *p* = 0.408, 1-tailed.

We also coded infants' scanning behavior offline (eight 6-month-olds and fourteen 4.5-month-olds) to obtain more fine-grained information about whether infants of each age group attended to both boxes during test trials[4]. We calculated the average duration of looking to the sampled vs. the unsampled boxes during the sampling phase of the test trials, starting with the experimenter shaking the first box at the start of the trial until she finished sampling and revealed the populations for the online coder to begin timing. We performed this analysis to examine whether infants of each age group attended to the same parts of the stage during sampling. The 6-month-old infants looked approximately equally to both the sampled and unsampled boxes, *F*(1,7) = 9.22, *p* = .306, effect size = 0.148 . The 4-month-olds did not show this pattern; they spent significantly more time attending to the *unsampled* box than the sampled box, *F* (1,13) = 19.21, *p* = .001, effect size = 0.596 (See Table 1 for means).

We also coded infants' average duration of looking to each box during the display phase of the test trials, commencing when the populations were revealed (i.e., when the experimenter said, "look") and ending when the online coder stopped timing (when the infant looked away for 2 consecutive seconds). This analysis addresses the potential concern that infants may have looked significantly longer at the sampled box than the unsampled box when the populations and samples were visible, and thus were reacting to a perceptual mismatch between the sampled box and the sample, rather than estimating probability. We ran a repeated-measures ANOVA with duration of looking towards the sampled vs. the unsampled box as a within-subjects factor and age group as a between-subjects factor. There were no main effects or interactions (all *p*-values >.5). Thus, when all of the perceptual

---

[4] We coded half of the sample of 6-month-olds (8 randomly chosen infants) because the scanning behavior of these randomly selected infants showed no signs of potential differences. We coded our full sample of 4.5-month-olds with the exception of 2 infants whose data were unavailable because their videos were damaged.

information was in sight, (i.e., the populations and the sample were visible) infants of both age groups looked equally at both boxes (see Table 1 for means) [5].

Discussion

Our results suggest that 6-month-old infants can make generalizations from samples to populations. When perceptual features are equated and infants cannot react to displays based solely on perceptual mismatches, 6-month-old infants look longer at a less probable sample of, for example, 4 yellow and 1 pink balls drawn from a mostly pink box than at a more probable sample of 4 pink and 1 yellow balls. Four-month-olds did not show this pattern; they looked roughly equally at both samples. This suggests that the ability to make generalizations from samples to populations emerges at around 6 months of age.

In addition, further analyses were conducted to address the potential concern that, despite efforts to draw attention to both boxes on stage, infants only attended to the sampled box during the timed portion of the test trials, and then simply reacted to the perceptual features between that box and the sample. This was not the case. These analyses argue against the interpretation that 6-month-olds were simply reacting to perceptual mismatches: When the displays were revealed, both age groups attended to each box equally. This weakens the perceptual mismatch account of the 6-month-olds' data, as infants of both ages attended to the same perceptual features during test trials but only the 6-month-olds demonstrated increased looking on trials when improbable samples were drawn.

It is still possible that infants used a reasoning bias known as the representativeness heuristic (i.e., samples and populations should be similar in appearance, e.g., Tversky & Kahneman, 1974) to make judgments in our experiment. This is a different concern than the methodological issue raised throughout regarding perceptual features in the displays, as this assumes that infants reasoned correctly

---

[5] We ran a number of additional analyses, none of which returned statistically significant results. We ran an ANOVA to determine if infants who were "correct" (i.e., looked longer overall at improbable samples) showed different scanning behavior than infants who were "incorrect" (i.e., looked longer at probable samples). There were no interactions between age, whether or not infants' looking times were "correct", and their duration of looks to each box location (*p's* > .05). In addition, there were no effects of whether or not infants' looking times were correct and the number of times they scanned between the two population boxes and the sample.

about random sampling, but questions whether they used a reasoning shortcut rather than probability computations. Some evidence from older infants suggests that this is an unlikely alternative interpretation. At 8- and 11-months, infants are able to make probabilistic inferences when samples and populations do not match in appearance (Denison & Xu, 2010a; Denison & Xu, 2011). In these experiments, populations with three sets of balls were used and infants were shown that all green balls – 50% of the balls in the population – were stuck and unmovable, and they were required to compute on the remaining sets of red and yellow balls in the population. On the test trials, both the probable and improbable samples (which only contained red and yellow balls) looked quite different from the population, and infants were still able to make correct inferences. Furthermore, preliminary results in our lab suggest that even when representativeness and probability are put in direct conflict, 11-month-olds are able to reason correctly based on probabilities (Denison & Xu, 2011). Similar experiments with 6-month-olds will help us more directly rule out the interpretation that young infants simply use the representativeness heuristic.

Although we now have evidence of intuitive probabilistic reasoning in 6-month-olds, it appears that 4.5-month-olds may not share similar intuitions. Two explanations may account for the negative findings from the 4.5-month-olds. The first possibility is that these findings demonstrate a true developmental difference between 4.5- and 6-month-old infants in probabilistic reasoning. This developmental progression parallels the one found in earlier studies on using conditional probabilities in causal learning (Sobel & Kirkham, 2006, 2007). In order to make accurate probabilistic inferences, infants must track where samples are drawn from, and this may be an ability that 4.5-month-old infants lack. When we coded where infants looked during the sampling phase of our experiment, we found that, surprisingly, 4.5-month-olds spent about 70% of their time attending to the *unsampled* population box. Six-month-olds, on the other hand, spent time scanning the entire scene. Perhaps the 6-month-old infants were able to scan the scene and extract the relevant information for making generalizations whereas the 4.5-month-olds were not yet able to hone in on the most pertinent components of a scene when making probabilistic inferences. Infants at this age may not realize that it is necessary to attend to the source from which a sample is drawn in order to make accurate generalizations. Future experiments are needed to explore this possibility more directly. The current experiment represents a much-needed attempt to fill the gap in the literature on visual statistical inference in infants below 8 months of age.

The second possible explanation of the 4.5-month-olds' null results is that the experimental procedures were not suitable for use with infants of this age. The younger infants may have looked longer to the unsampled box not because they lack a conceptual understanding relevant to sampling but because they were confused or distracted by the methodology. The mimicking action on the unsampled box could have disrupted infants' performance in a number of ways: They may have thought that balls had been drawn from both boxes or they may have been distracted by the mimicking, causing them to focus on the mimicking rather than the true sampling. Future work may use a manipulation such as drawing out balls from the 'unsampled' box and returning them each time without ever putting them in the small display container. This design would still allow us to equate the actions made with both boxes, and the act of sampling and returning balls may be more familiar to infants than pretend sampling. Another possibility is to consider measures other than looking-time when testing this ability in very young infants, for example, ERP experiments have found evidence for statistical learning in newborns (Teinonen et al., 2009).

Our findings here are consistent with recent research applying probabilistic models of human cognition to experimental findings in infancy and early childhood (Schulz, Bonawitz & Griffiths, 2007; Teglas et al, 2011; Xu & Tenenbaum, 2007). One of the key goals of this enterprise is to identify the rational, inferential, and statistical learning mechanisms that exist early in life, and have the power to support conceptual development (Xu & Griffiths, 2011). Indeed, if humans' beliefs are represented probabilistically as this class of theories assume then at minimum, infants should be able to represent and compute rudimentary probabilities. Most of the current Bayesian models focus on ideal-observer analyses of human behavior at a computational level, but recently, several models have attempted to capture data from infants and young children by using algorithms that approximate full Bayesian inference (e.g., Bonawitz, Denison, Chen, Griffiths, Gopnik, 2011; Denison, Bonawitz, Gopnik, & Griffiths, 2011, under review; Teglas et al., 2011). On the empirical side, there is also new evidence suggesting that we must take into account resource constraints such as working memory load when investigating young infants' learning algorithms (e.g., Bulf et al., 2011). For the current study, a task reducing the information-processing demand may reveal earlier competence for probabilistic reasoning in infants younger than 6 months.

We presented the first experiment exploring probabilistic inference in young infants. The findings suggest that 6- but not 4.5-month-old infants can make generalizations from small samples to larger populations. Our results, in combination with recent evidence from similar experiments, provide convergent support for early competence in probabilistic reasoning in infancy (Denison & Xu, 2010a, 2010b, 2011; Teglas et al., 2007, 2011; Xu & Denison, 2009; Xu & Garcia, 2008). These findings, both from looking-time and action-based measures are particularly impressive given the extensive experimental results suggesting that adults often make faulty probabilistic inferences in a wide range of tasks. For example, Tversky and Kahneman (1974, 1981) found that adult judgments were often hindered by the incorrect application of reasoning heuristics when making probabilistic inferences. The infant findings suggest that humans do have an intuitive, implicit probabilistic reasoning mechanism. Starting at around 6 months of age, infants appear to understand something about the predictive relationship between samples and populations; by the end of the first year, infants can compute probabilities in looking-time studies and the output of these computations can guide their action.

Acknowledgments

We thank Morgan Bartholomew, Eduardo Colon and Pallavi Trikhutum for help with data collection, as well as Annie Chen for secondary coding of data. We also thank the parents and infants for their participation. This research was supported by a graduate student fellowship awarded to S. Denison by the Natural Sciences and Engineering Research Council of Canada.

References

Bonawitz, E.B., Denison, S., Chen, A., Gopnik, A., & Griffiths, T.L. (2011) A simple sequential algorithm for approximating Bayesian inference. *Proceedings of the Thirty-third Cognitive Science Society.*

Bulf, H., Johnson, S.P. & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition, 121*, 127-132. doi: 10.1016/j.cognition.2011.06.010

Casstevens, R.M. (2007). jHab: Java Habituation Software (Version 1.0.0) [Computer software]. Chevy Chase, MD.

Denison, S., Bonawitz, E.B., Gopnik, A., & Griffiths, T.L. (2010). Preschoolers sample from probability distributions. *Proceedings of the Thirty-second Cognitive Science Society.*

Denison, S., Bonawitz, E.B., Gopnik, A., & Griffiths, T. (in review) Rational variability in children's causal inferences: The Sampling Hypothesis.

Denison, S. & Xu, F. (April 2011). Eight- and eleven-month-old infants integrate deterministic and probabilistic physical constraints in statistical inference, *Society for Research in Child Development Conference*, Montreal, Quebec.

Denison, S. & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science, 34*, 885-908. doi:10.1111/j.1551-6709.2010.01111.x

Denison, S. & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single- event probability with large set sizes. *Developmental Science, 13*, 798-803. doi:10.1111/j.1467-7687.2009.00943.x

Gweon, H., Tenenbaum, J.B., & Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107(20),* 9066-9071. doi:10.1073/pnas.1003095107

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*, 483-524. doi:10.1016/0010-0285(83)90017-8

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition, 83*, B35-B42. doi:10.1016/S0010-0277(02)00004-5

Kuhl, P K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9,* F13-F21. doi: 10.1111/j.1467-7687.2006.00468.x

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82(3),* B101-B111. doi:10.1016/S0010-0277(01)00157-3

Schulz, L., Bonawitz, E.B., & Griffiths, T.L. (2007). Can being scared give you a tummyache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology*, *43(5),* 1124-1139. doi: 10.1037/0012-1649.43.5.1124

Sobel, D.M. & Kirkham, N.Z. (2006) Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, 42, 1103-1115. doi: 10.1037/0012-1649.42.6.1103

Sobel, D.M. & Kirkham, N.Z. (2007). Bayes nets and Babies: Infants' developing representations of causal knowledge. *Developmental Science*, 10, 298-306. doi: 10.1111/j.1467-7687.2007.00589.x

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10-21. doi: 10.1186/1471-2202-10-21

Teglas, E., Girotto, V., Gonzalez, M., Bonatti, 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. doi:10.1073/pnas.0700271104

Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., and Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332,* 1054 – 1058. doi: 10.1126/science.1196404

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science 185(4157),* 1124-1131. doi:10.1126/science.185.4157.1124

Tversky, A.; Kahneman, D. (1981). "The framing of decisions and the psychology of choice". *Science 211(4481),* 453–458. doi: 10.1126/science.7455683

Xu, F. & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112,* 97-104. doi: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. doi:10.1073/pnas.0704450105

Xu, F. & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10(3),* 288-297. doi:10.1037/0033-295X.114.2.245

Figure 1: The sequence of a test trial in Expt. 1 (Xu & Garcia, 2008). The experimenter shakes the box, closes her eyes and draws out balls from the closed box. She then reveals the population.

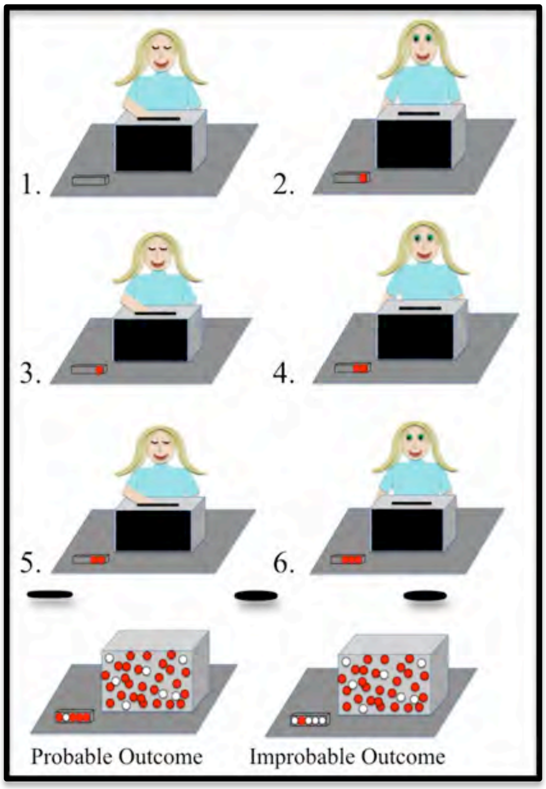

Figure 2: The two possible outcomes in the current experiment. The population boxes displayed simultaneously ensure equal amounts of pink and yellow are displayed.

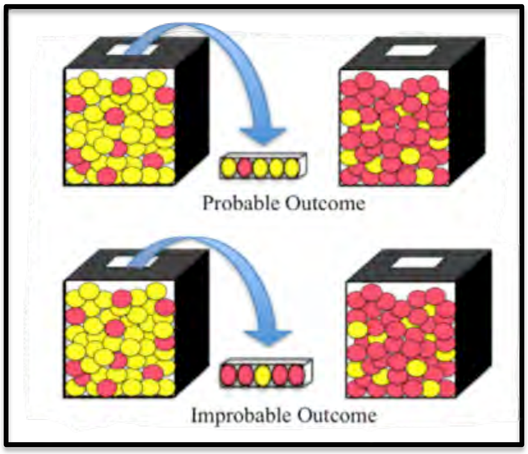

Table 1: Mean Looking Times in Seconds to Each Box by Age Group

Duration of Looking (in sec) to Boxes During Sampling		
Display Phase	Sampled Box (*SE*)	Unsampled Box (*SE*)
4-Month-Olds (N=14)	2.35 (1.57)	2.02 (1.65)
6-Month-Olds (N=8)	4.92 (2.08)	4.93 (2.19)
Sampling Phase		
4-Month-Olds (N=14)	1.96 (.50)	5.13 (0.88)
6-Month-Olds (N=8)	3.41 (0.66)	4.49 (1.17)