Cognition 112 (2009) 97–104

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

# Statistical inference and sensitivity to sampling in 11-month-old infants

Fei Xu *, Stephanie Denison *

*Department of Psychology, 2136 West Mall, University of British Columbia, Vancouver, B.C., Canada V6K 1Z4*

ARTICLE INFO

*Article history:*
Received 7 July 2008
Revised 13 April 2009
Accepted 15 April 2009

*Keywords:*
Statistical inference
Infant learning
Random sampling

ABSTRACT

Research on initial conceptual knowledge and research on early statistical learning mechanisms have been, for the most part, two separate enterprises. We report a study with 11-month-old infants investigating whether they are sensitive to sampling conditions and whether they can integrate intentional information in a statistical inference task. Previous studies found that infants were able to make inferences from samples to populations, and vice versa [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]. We found that when employing this statistical inference mechanism, infants are sensitive to whether a sample was randomly drawn from a population or not, and they take into account intentional information (e.g., explicitly expressed preference, visual access) when computing the relationship between samples and populations. Our results suggest that domain-specific knowledge is integrated with statistical inference mechanisms early in development.



## 1. Introduction

In the last few decades, much research has focused on characterizing the initial knowledge state and the nature of early learning mechanisms in infancy. Two lines of research have helped advance our understanding of the initial conceptual state of the human infant and the learning mechanisms available to them early on. On the one hand, research has shown that young infants have systems of knowledge for reasoning about objects, agents, number, and causality (e.g., Baillargeon, 2004; Carey, in press; Gergely, Nádasdy, Csibra, & Bíró 1995; Leslie & Keeble, 1987; Spelke, 1990; Xu & Spelke, 2000); at the same time, research has also shown that infants possess powerful learning mechanisms that compute various statistics over the input they receive from the environment (e.g., Aslin, Saffran, & Newport, 1998; Gerken, 2006; Gomez, 2002; Gopnik et al., 2004; Marcus, Vijayan, Bandi Rao, & Vishton 1999; Saffran, Aslin, & Newport, 1996; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson 2002; Sobel & Kirkham, 2006; Xu & Garcia, 2008). With rare exceptions, these two lines of research have had their separate lives. The research uncovering sophisticated initial knowledge has focused on getting the ages of success to be younger and younger, and the research on learning has focused on showing how infants can compute more and more complex statistics given the input.

With adults and young children, some studies have probed how domain-specific knowledge interacts with statistical learning mechanisms. Newport and Aslin (2004), for example, showed that adults learned non-adjacent dependencies in a word segmentation task only when given linguistically appropriate units such as segments. Schulz and Gopnik (2004) found that 4-year-old children were capable of overriding domain knowledge with statistical evidence. With infants, there has not been much research investigating how domain knowledge interacts with input statistics.

What is the relationship between initial knowledge and statistical learning mechanisms in infancy? One possibility is that early in development systems of conceptual knowledge about objects, people, and number unfold according

* Corresponding authors. Tel.: +1 604 822 5972; fax: +1 604 822 6923.
*E-mail addresses:* fei@psych.ubc.ca (F. Xu), steph@psych.ubc.ca (S. Denison).

0010-0277/$ - see front matter Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.04.006

to some maturational timetable and their development is distinct from the development of early learning mechanisms; another possibility is that even in infants, initial conceptual knowledge interacts with learning mechanisms in systematic and interesting ways.

The current study investigates whether infants can integrate domain-specific knowledge with statistical inference mechanisms. In particular we ask whether infants are able to incorporate their knowledge about agents (e.g., visual access, goal or preference) in a statistical inference task.

Recent studies suggest that infants may be able to use the statistical information in a small sample to make inferences about a larger population, and vice versa (Xu & Garcia, 2008). In these studies, 8-month-old infants were shown a sample drawn from a closed box full of red and white Ping-pong balls. On alternating trials, a sample of four red and one white balls or a sample of one red and four white balls was drawn. Then the box was opened to reveal a population of mostly red balls with a few white ones mixed in. Infants looked longer at the unexpected outcome of one red and four white balls than the expected outcome of four red and one white balls. Xu and Garcia (2008) suggested that these experiments provide evidence that infants can make inferences from samples to populations based on the statistical information in the input, perhaps employing basic principles of probability (see also Teglas, Girotto, Gonzalez, & Bonatti (2007), for evidence that infants can compute single-event probabilities).

Several important questions remain open from these studies. First, in order to make use of the statistical information in the input, the learner needs to assume that the sample is randomly drawn from the population. If the learner had reasons to believe that the sample was not drawn randomly from the population, the inference would not be valid. Are learners sensitive to the random sampling assumption? This is an important question in understanding the nature of the inference mechanism. Second, is this statistical inference mechanism a low-level, automatic mechanism or is it capable of integrating other sources of knowledge? Suppose the experimenter looked into the box while pulling out the Ping-pong balls, would the infants assume that since the experimenter had visual access to the content of the box, she may not be pulling out a random sample? The current study addresses these two questions by asking if infants can integrate intentional information (e.g., visual access, explicitly expressed preference) in this statistical inference task.

Three conditions were included in our study: a random sampling condition, a non-random sampling condition, and a blindfold condition. In the random sampling condition, the procedure was similar to that of the previous studies (Xu & Garcia, 2008) except that the two samples were either five red Ping-pong balls or five white Ping-pong balls. We predicted the same results as before. In the non-random sampling condition, the experimenter expressed a preference for say, red balls, by repeatedly selecting only red balls from a container of three red and three white balls at the beginning of the experiment. During the sampling process, she looked into the large box while pulling out the Ping-pong balls, suggesting that she had visual access to the content of the box and she had the means to only pull out the balls of her preferred color. If infants were sensitive to this violation of the random sampling assumption, their looking times would be predicted by the explicitly expressed preference of the experimenter, and not by the content of the box.

In the blindfold condition, the experimenter also expressed a preference for Ping-pong balls of a particular color. However, she was blindfolded during the sampling process. Although the experimenter may have preferred red balls, her lack of visual access to the content of the box while sampling would prevent her from being able to pick and choose based on her preference. If infants were able to integrate intentional information in this statistical inference task, they would treat this condition as equivalent to the random sampling condition – their looking times should be predicted by the content of the box and not by the experimenter's preference.

Eleven-month-old infants were tested. We chose 11-month-olds for two reasons. One was to provide a replication of the earlier results with 8-month-olds in Xu and Garcia's studies with another age group. The other was that this is a more difficult task than those completed by 8-month-olds. The task required infants to encode the experimenter's preference or goal for a particular color of balls, and then decide whether this preference was relevant based on the experimenter's visual access. The developmental literature suggests that it is not until the end of the first year that infants demonstrate this understanding, a pre-requisite for the current study (Brooks & Meltzoff, 2002; D'entremont and Morgan, 2006).

## 2. Experiment

### 2.1. Method

#### 2.1.1. Participants

Participants were 72 11-month-old infants (36 girls and 36 boys; mean age 10;29 [months;days], ranged from 10;15 to 11;17). Twenty-four infants (12 girls and 12 boys) were randomly assigned to each of three conditions: the random sampling condition (mean age 11;01), the non-random sampling condition (mean age 10;26), and the blindfold condition (mean age 11;02). An additional 5, 4, and 7 infants in each condition were tested but not included in the final sample due to experimenter error (5), equipment failure (3), or not completing the study due to fussiness (8). All participants were recruited from the Greater Vancouver area. Infants received a T-shirt or bib and a diploma for their participation.

#### 2.1.2. Materials

A 39 cm × 34 cm × 22 cm box constructed out of foam core, fabric, and Plexiglas was used in the study. The box was a white rectangular cube taped together with black duct tape. The inside of the box was divided into three parts – two Plexiglas containers were inserted into the front and back of the box, each containing 72 Ping-pong balls, and a center compartment was used to hold the samples to be removed from the box during test trials. When viewed from the front, the box appeared to be one large box filled with multiple layers of balls. The front and back of the box were covered with black fabric curtains (secured to the top of the box with Velcro) that could be lifted to reveal the contents of the box through the transparent windows. The "mostly white" side of the box contained 60 white and 12 red balls, and the "mostly red" side contained the opposite ratio. The top of the box had a 10 cm × 24 cm cutout covered with two pieces of overlapping spandex that allowed the experimenter to reach into the center compartment of the box.

A small Plexiglas container (20 cm × 4.5 cm × 4 cm) was placed at the front left-hand corner of the stage to display the samples pulled out of the larger box during test trials. A slightly longer Plexiglas container (28.5 cm × 4.5 cm × 4 cm) was also used during the preference phase (see below). Both containers were narrow enough such that the balls lined up in a single row when placed in the containers. A small white box (17.5 cm × 17.5 cm × 8 cm) constructed out of foam core was used to hold three red and three white Ping-pong balls. A black cotton and spandex (5 cm in width) headband served as the blindfold.

### 2.1.3. Apparatus

Testing took place in a quiet room. All events were presented on a puppet stage. The experimenter sat behind the stage; her upper body and head were visible to the infant when the back curtain was lifted. The viewable area measured 94 cm × 55 cm. The observer watched the infant on a TV monitor in one corner of the testing room and recorded the infant's looking times on an iBook using MacXHAB1.4 (Pinto, 2005). The observer was blind to the order of the trials. A fan was located in the back part of the room, set to low speed, to muffle any extraneous sounds from the hallway. The stage was lit; the rest of the room was darkened during the study.

The infant sat approximately 70 cm from the stage in a high chair. Two cameras recorded the session: one focused on the stage to record the procedure, the other focused on the infant's face to record looking behavior.

### 2.1.4. Design and procedure

All infants were tested in a violation-of-expectancy looking time paradigm. In all conditions infants were first shown a small foam core box with six Ping-pong balls (three red and three white) and were allowed to play with them for about 30 s. The experimenter picked up a few of the Ping-pong balls, one at a time, and she encouraged the infant to do the same.

### 2.1.5. Random sampling condition

*2.1.5.1. Familiarization trials.* Each infant received four familiarization trials. On each trial, the experimenter placed the large box (with the front curtain closed) on the stage and a small Plexiglas container to the right of the box. She shook the box back and forth a few times, saying, "What do I have in here?" She then lifted the front cover of the box, and lowered the backdrop of the stage to conceal herself while saying "Look, [baby's name], look!" The observer began timing upon hearing the second "look." The trial ended when the infant looked away for two consecutive seconds.[1] The four trials alternated between the mostly red population and the mostly white population, in a counterbalanced order. The large box was removed after each trial. These trials were included to familiarize the infants to the objects as well as to the general procedure of the study. Also, once infants have been exposed to the two populations, mostly white or mostly red, they can use this information during test trials to generate a hypothesis as to which box the experimenter might be sampling from. The familiarization trials lasted about 4 min.

*2.1.5.2. Preference phase.* The experimenter placed on the empty stage a second Plexiglas container with three red and three white balls. All six balls were visible to the infant since they were lined up in a row in random order. To begin, she picked up three red (for half of the infants, white) balls, one at a time, and placed them into the container to her right. She pointed to the three balls while smiling and saying, "Look at these!" then returned them to the original container. Next, she repeated the same action with the three balls of the other color. In this condition, the experimenter did *not* express a preference for balls of either color and this phase was included to equate this condition with the other two conditions. The larger container and the six balls were then removed; the Preference phase lasted approximately 30 s.

*2.1.5.3. Test trials.* On each trial, the experimenter placed the large box on the stage, with its front curtain closed. She shook the box a few times, closed her eyes, turned her head away, and reached into the box. She pulled out two Ping-pong balls of one color (red or white) and placed them into the small Plexiglas container to her right. She repeated this action, pulling out three more balls of the same color. The small Plexiglas container had a total of five balls. The experimenter then lifted the front curtain of the box and lowered the backdrop while saying, "Look, [baby's name], look!" (Fig. 1) Once the backdrop was lowered, the experimenter was no longer visible. Each test outcome consisted of only the large box with the population and the small Plexiglass container with the sample. The observer began timing upon hearing the second "look", indicating the precise moment that the box content was visible to the infant. At the end of each trial, the stage was cleared. Each infant received six test trials, alternating between a five-red sample and a five-white sample. The test trials lasted approximately 8 min.

The order of the familiarization trials (mostly red first or mostly white first), the order during the preference phase (red or white first for the random sampling condition), and the order of the samples on the test trials (red or white first) were counterbalanced across infants. Half of the infants saw the mostly white population on all test trials, and the other half the mostly red population.

[1] In the random sampling and non-random sampling conditions, 12 infants received these 4 full familiarization trials, the other 12 received 4 five-second familiarizations. These groups were later pooled in data analyses, as they did not differ statistically.

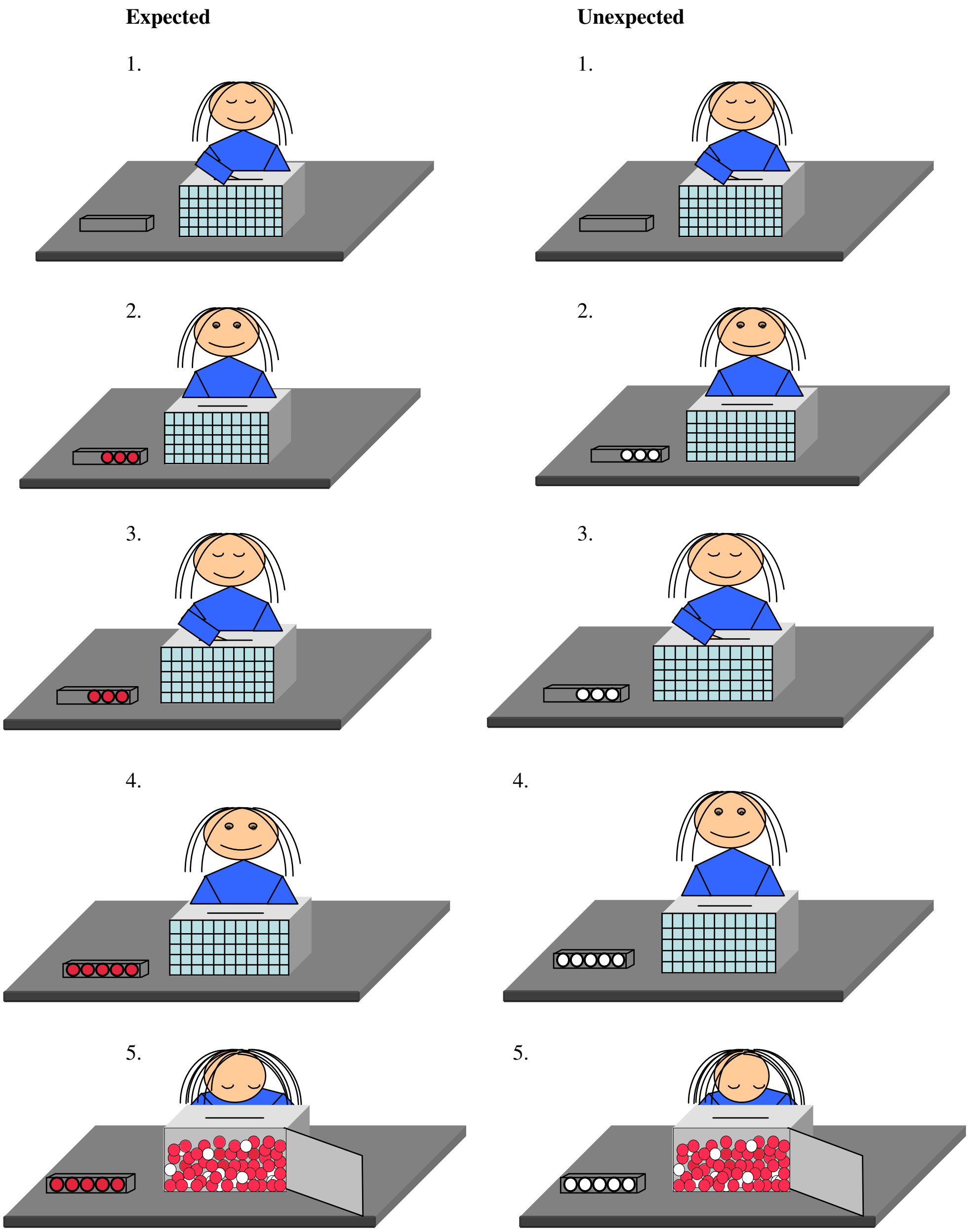


**Fig. 1.** Schematic representation of the procedure of the random sampling condition.

*2.1.6. Non-random sampling condition*

The design and procedure were the same as the random sampling condition except for the following critical differences.

*2.1.6.1. Preference phase.* The experimenter picked out balls of the same color both times. Her action conveyed to the infant that she always picked out a particular color. For half of the infants she chose only red balls and for the other half only white.

*2.1.6.2. Test trials.* On each trial, instead of closing her eyes and turning her head away, the experimenter looked into the top opening of the box while pulling out the sample, i.e., she had visual access to the content of the box.

*2.1.7. Blindfold condition*

The design and procedure were the same as the non-random sampling condition except for the following critical differences.

*2.1.7.1. Blindfolding.* While the infant and parent were in the reception room, the experimenter, parent and infant participated in a short peek-a-boo game with a blindfold. The experimenter began by showing the infant the blindfold and saying, "Look at this!" She then placed the blindfold over her eyes and said, "[Baby's name], where did you go?" removed the blindfold and said, "There you are!" She put the blindfold on again and said, "Where's (baby's name)?" removed the blindfold again and said, "Peek-a-boo!" She then repeated these sequences a second time. Next she said, "Should Mom try the game? Let's give the blindfold to mom." The experimenter then allowed the mom (or dad) to play the game and asked her (or him) to place the blindfold over their baby's eyes three or four times for a couple of seconds each. This part was included given that Brooks and Meltzoff (2002) found that placing the blindfold on the infant was an effective way to demonstrate that a person could not see when blindfolded.

*2.1.7.2. Familiarization trials.* Right before the familiarization trials, the experimenter showed the infant the blindfold one more time. She placed it over her eyes and said, "Remember this? I cannot see you!" She then put the blindfold around the top of her head similar to a bandana and sat down behind the stage. Then the familiarization trials began.

*2.1.7.3. Preference phase.* This was identical to that of the non-random sampling condition. The experimenter expressed a preference for a particular color balls.

*2.1.7.4. Test trials.* After placing the large box on the stage, the experimenter drew the infant's attention by saying, "Are you ready?" She made sure that the infant was watching while she pulled the blindfold down over her eyes. Therefore, despite her expressed preference for a particular color ball, she did not have visual access to the content of the box during sampling. The test trials then unfolded as in the random sampling condition.

*2.1.8. Predictions*

*2.1.8.1. Random sampling condition.* Infants' looking times on test trials should be predicted by the population in the box, as in Xu and Garcia (2008).

*2.1.8.2. Non-random sampling condition.* Infants' looking times on test trials should be predicted by the experimenter's expressed preference for one color ball, and not the content of the box. If the sample corresponds to the experimenter's preference, it would be expected; if the sample does not correspond to the experimenter's preference, it would be unexpected. In this condition, infants should ignore the content of the population, and use the preference information, as these sources of information are not always complimentary. Since the experimenter had visual access during sampling, she had the means to remove the balls of her preferred color, which may not be a random sample given the population of the box.

*2.1.8.3. Blindfold condition.* Infants' looking times should be predicted by the population in the box, as in the random sampling condition. This condition asks whether infants understand that the expressed preference would not predict the experimenter's behavior since she does not have visual access due to the blindfold. We used a blindfold as opposed to having the experimenter close her eyes since the latter may have been too subtle for these young infants. This condition also provides an important control for the non-random sampling condition by demonstrating that infants do not simply look longer at the color that the experimenter did not pick out during the preference phase due to a novelty response.

*2.1.9. Results*

Preliminary analyses found no effects of gender, content of the box (mostly white vs. mostly red), expressed preference (white or red), or order of the familiarization trials. Subsequent analyses collapsed over these variables. The average looking times for the test trials of each condition are displayed in Fig. 2. A second observer who was unaware of the order of the trials coded the data. Inter-observer reliability averaged 95%, 97%, and 92% for the three conditions, respectively.

An analysis of variance found no differences between average looking times on the familiarization trials across conditions ($F(2,45) = 1.36$, $p > .1$; $M_{\text{random-sampling}} = 13.4$ s, $M_{\text{non-random-sampling}} = 14.4$ s, $M_{\text{blindfold}} = 11.4$ s).

An omnibus ANOVA examined the effect of outcome (expected vs. unexpected, according to the content of the box) and condition (random, non-random, and blindfold). This revealed a significant effect of condition, $F(2,69) = 9.24$, $p < .01$, effect size ($\eta p^2$) = .21, and a significant interaction, $F(2,69) = 6.79$, $p < .01$, effect size ($\eta p^2$) = .16.

Tukey post-hoc comparisons indicated that there was a significant difference in overall looking time on test trials between the non-random sampling condition ($M = 11.96$) and the random sampling condition ($M = 9.44$), $p < .05$, and between the non-random sampling condition and blindfolding condition ($M = 7.72$), $p < .001$. This may be due to the fact that the processing demands for the non-random sampling condition were higher than in the other two conditions. In the random sampling condition, infants

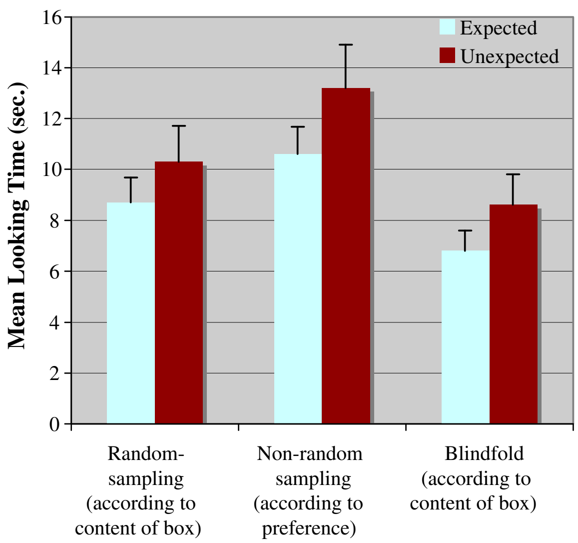


**Fig. 2.** Mean looking times with standard error.

need only reason about the relationship between the sample and the population. In the blindfold condition, it is possible that once infants saw the blindfold on the experimenter at the beginning of the test trials, they realized that despite her expressed preference, she had no means for picking out only the preferred color. The task is now the same as in the random sampling condition. In the non-random sampling condition, however, infants must remember that the experimenter only picked out balls of one color. They were then shown a sample, which sometimes contained only the opposite color balls, and exposed to the population. It may be the case that the infants' first reaction was to compare the sample with the population, then they had to inhibit this initial reaction in order to take into account the experimenter's expressed preference, resulting in overall longer looking times. In other words, the longer looking times in the non-random sampling condition may be taken as evidence that more processing was needed due to the conflict presented between the population, the sample, and the expressed preference.

Given the interaction between outcome and condition, ANOVAs were performed to analyze each condition separately. Specific patterns of looking times were predicted, thus these were planned comparisons.

#### 2.1.10. Random sampling condition

An ANOVA examined the effects of test trial order (expected first or unexpected first), trial pair (1–3), and outcome (expected vs. unexpected, *according to the content of the box*). There was a main effect of outcome, $F(1, 22) = 4.329$, $p < .05$, effect size ($\eta p^2$) = .16. Infants looked reliably longer at the unexpected outcome ($M$ = 10.3 s, SD = 7.0) than the expected outcome ($M$ = 8.7 s, SD = 4.8). There was also a main effect of trial pair, $F(2, 44) = 3.916$, $p < .05$; average looking time decreased over time. Sixteen of twenty-four infants looked longer at the unexpected outcome, Wilcoxon signed-ranks test: $z = -1.97$, $p < .05$. There were no other main effects or interactions.

#### 2.1.11. Non-random sampling condition

An ANOVA examined the effects of test trial order, trial pair, and outcome (expected vs. unexpected, *according to the experimenter's expressed preference*). There was a main effect of outcome, $F(1, 22) = 4.947$, $p < .05$, effect size ($\eta p^2$) = .17. Infants looked reliably longer at the unexpected outcome ($M$ = 13.2 s, SD = 8.4) than the expected outcome ($M$ = 10.7 s, SD = 5.3). There was also a main effect of trial pair, $F(2, 44) = 14.143$, $p < .0001$; average looking time decreased over time. Twelve of twenty-four infants looked longer at the unexpected outcome, Wilcoxon signed-ranks test: $z = -1.54$, $p < .05$. There were no other main effects or interactions.

A second ANOVA examined the effects of test trial order, trial pair, and outcome (expected vs. unexpected, *according to the content of the box*). There was no main effect of outcome, $F < 1$, n.s. Infants looked about equally at the unexpected outcome ($M$ = 11.1 s) and the expected outcome ($M$ = 12.8 s). There were no other interactions.[2]

#### 2.1.12. Blindfold condition

An ANOVA examined the effects of test trial order, trial pair, and outcome (expected vs. unexpected, *according to the content of the box*). There was a main effect of outcome, $F(1, 22) = 5.177$, $p < .05$, effect size ($\eta p^2$) = .19. Infants looked reliably longer at the unexpected outcome ($M$ = 8.6 s, SD = 5.9) than the expected outcome ($M$ = 6.8s, SD = 3.9). There was also a main effect of trial pair, $F(2, 44) = 7.825$, $p < .001$; average looking time decreased over time. Nineteen of twenty-four infants looked longer at the unexpected outcome, Wilcoxon signed-ranks test: $z = -2.34$, $p < .05$. There were no other main effects or interactions.

A second ANOVA examined the effects of test trial order, trial pair, and outcome (expected vs. unexpected, *according to the experimenter's expressed preference*). There was no main effect of outcome, $F < 1$, n.s. Infants looked about equally at the unexpected outcome ($M$ = 7.73 s) and the expected outcome ($M$ = 7.72 s). There were no other main effects or interactions.

## 3. Discussion

The results of this study provide further evidence for a sophisticated statistical inference mechanism in infants. The results from the random sampling condition replicated those of Xu and Garcia (2008) with another age group. The results from the non-random sampling condition showed that 11-month-old infants were sensitive to whether a sample had been drawn randomly from a population – when the sample had not been drawn randomly, they were able to discard the statistical information in the input. Instead, infants were able to use intentional information (in this case the experimenter's expressed preference and visual access during the sampling process) as the basis for their inferences. Even more impressively, in the blindfold condition, the infants were able to integrate multiple sources of information in deciding whether to employ the statistical inference mechanism. The findings from this condition suggest that infants could override the random sampling assumption, but only under appropriate circumstances – even though they had encoded the experimenter's preference (as was demonstrated in the non-random sampling condition), they were able to discard that information given the lack of visual access during the sampling process. When the infants saw the blindfold on the experimenter on the test trials, they learned that she had no visual access to the content of the box. The infants then reverted back to assuming random sampling

[2] Due to the fact that the population in the box (mostly red or mostly white) and the experimenter's preference (red or white) were fully crossed, the population of balls in the box and the experimenter's preference were consistent for half of the infants and inconsistent for the other half of the infants. One might ask whether the infants in the consistent half performed better on this task than those in the inconsistent half. An additional ANOVA examined the effects of consistency and outcome, and found no interaction between these two factors, $F(1,22) = 0.018$, $p = .89$.

and made inferences about the content of the box based on the statistical information in the sample. Future studies will investigate in more detail the nature of the statistical inference mechanisms as well as how they are able to integrate domain-specific knowledge.

It is important to note that we do not intend to make any strong claims about infants' abilities to reason about others' preferences or goals. In the non-random sampling condition, infants may have attributed a preference or goal to the experimenter. Alternatively, the infants may have simply encoded that the agent, for whatever reason, always picked out balls of a particular color. Whether or not the infants viewed this as a preference or goal is interesting but beyond the scope of the current study. Our main concern is to demonstrate that infants can integrate other sources of information into the statistical inference mechanism. Future studies may investigate this question by asking whether the preferences expressed in this experiment – conveyed by the experimenter choosing one color balls on two different occasions – will lead to the prediction that the experimenter will reach for that color ball in a goal-directed action paradigm (e.g., Woodward, 1998).

We have provided evidence that infants can integrate psychological constraints in employing their statistical inference mechanism. What about physical constraints? A wealth of evidence suggests that infants as young as 2–4 months of age already have a set of principles for physical reasoning (e.g., Baillargeon, 2004; Spelke, Breinlinger, Macomber, & Jacobson 1992). In some recent studies, we have begun to explore this question (Denison & Xu, in preparation, 2008). In these experiments, 11-month-old infants were shown that green balls were 'stuck' to the box (with Velcro), they were then put in a statistical inference task with a box of Ping-pong balls of three colors, green, red, and white. The question of interest was whether the infants were able to exclude the green balls and compute on the red and white balls based on the statistical properties of that sub-population. Our findings suggest that they can. Similarly, Teglas et al. (2007) found that when solidity – a principle that guides infant's physical reasoning as early as 4 months of age – comes into conflict with probability computations, infants disregard probability information and reason according to solidity. Thus the statistical inference mechanism is capable of integrating both psychological and physical knowledge by 11 months of age.

Will younger infants behave similarly in these studies? Perhaps a simpler experimental paradigm can be developed to test younger infants. Since the existing evidence suggests an earlier developing physical reasoning system than the psychological reasoning system, it may be the case that younger infants will be more apt in integrating physical constraints than psychological ones.

The study reported here, as well as Teglas et al. (2007) and Denison and Denison and Xu (2008, in preparation), gives a tentative answer to the question raised in the introduction: by the end of the first year, infants are not only sophisticated statistical learners, but are also capable of integrating domain-specific knowledge – from both the physical and the psychological realm – in their computations. This answer blurs the line between the research enterprise on statistical learning and the research enterprise focusing on core knowledge. It calls into question the claim that the learning mechanisms infants have are automatic and completely bottom–up; it also calls into question the claim that principles of physical and psychological reasoning develop in a vacuum – the inferential mechanisms studied here and elsewhere may be partially responsible for the acquisition of these principles. The approach we have taken here, where domain-general statistical inference mechanisms compute over domain-specific knowledge, is in complete agreement with the recent emergence of the rational constructivist view (Xu, 2007; Xu, Dewar, & Perfors, 2009) and Bayesian models of cognition and cognitive development (e.g., Chater, Tenenbaum, & Yuille 2006; Kemp, Perfors, & Tenenbaum 2007; Schulz, Bonawitz, & Griffiths 2007; Sobel, Tenenbaum, & Gopnik 2004; Tenenbaum, Griffiths, & Kemp 2006; Xu & Tenenbaum, 2007; Xu et al. 2009). Our aim is not to separate the study of learning mechanisms from the study of content knowledge. Instead, our aim is to understand how learning mechanisms integrate prior constraints and input statistics, and to develop precise computational models of these interactions.

## Acknowledgments

We thank members of the UBC Baby Cognition Lab for help with data collection, and members of the UBC Baby Cognition Lab and the MIT Computational Cognitive Science Lab for helpful discussions. We also thank the infants and parents who participated in our study. This research was supported by a grant from the Natural Science and Engineering Research Council of Canada (NSERC) to FX.

## References

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science, 3*, 89–94.

Brooks, R., & Meltzoff, A. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology, 38*(6), 958–966.

Carey, S. (in press). The origins of concepts. Oxford University Press.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences, 10*, 287–291.

Denison, S., & Xu, F. (2008) Sensitivity to physical and psychological constraints in infants' statistical inference. In *Poster presented at the 16th Biennial International Conference on Infant Studies, March 27–29, Vancouver, Canada.*

Denison, S., & Xu, F. (in preparation) Integrating physical constraints in statistical inference by 11-month-old infants.

D'entremont, B., & Morgan, R. (2006). Experience with visual barriers and its effects on subsequent gaze-following in 12–13-month-olds. *British Journal of Developmental Psychology, 24*, 465–475.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, B67–B74.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review, 111*, 1–31.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307–321.

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

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.

Pinto, J. (2005) Xhab1.4. Stanford University.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–2928.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared make your tummyache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology, 43*(5), 1124–1139.

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40*, 162–176.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*, 13–19.

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

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*, 303–333.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 29–56.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*, 605–632.

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.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*, 309–318.

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.

Xu, F. (2007). Rational statistical inference and cognitive development. In P. Carruthers, S. Laurence, & S. Stich (Eds.). *The innate mind: Foundations and the future* (Vol. 3, pp. 199–215). Oxford University Press.

Xu, F., Dewar, K., & Perfors, A. (2009). Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism. In B. M. Hood & L. Santos (Eds.), *The origins of object knowledge*. Oxford University Press.

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.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1–B11.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.