Intuitive statistics by 8-month-old infants

Fei Xu[*] **and Vashti Garcia**

Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4

Edited by Susan E. Carey, Harvard University, Cambridge, MA, and approved February 6, 2008 (received for review May 11, 2007)

Human learners make inductive inferences based on small amounts of data: we generalize from samples to populations and vice versa. The academic discipline of statistics formalizes these intuitive statistical inferences. What is the origin of this ability? We report six experiments investigating whether 8-month-old infants are "intuitive statisticians." Our results showed that, given a sample, the infants were able to make inferences about the population from which the sample had been drawn. Conversely, given information about the entire population of relatively small size, the infants were able to make predictions about the sample. Our findings provide evidence that infants possess a powerful mechanism for inductive learning, either using heuristics or basic principles of probability. This ability to make inferences based on samples or information about the population develops early and in the absence of schooling or explicit teaching. Human infants may be rational learners from very early in development.

infant cognition | probability | statistical inference

One hallmark of human learning is that human learners are able to make inductive inferences given a small amount of data (1–3). Our hunter–gatherer ancestors may have tasted a few berries on a tree and then decided that all berries from the same kind of tree are edible. They may have encountered a few friendly people from a neighboring tribe and made the inference that people in that tribe are likely to be friendly in general. Once such generalizations are made, the inferences may go in the other direction as well. This type of statistical inference (going from samples to populations, and from populations to samples) is present in virtually every domain of learning, be it foraging, social interaction, visual perception, word learning, or causal reasoning (4–10). Inductive learning in general requires some understanding of intuitive statistics, perhaps a simpler version of what scientists do in laboratory experiments or field studies.

What is the origin of the ability to make inductive inferences based on a small amount of data? Our experiments ask whether 8-month-old infants are "intuitive statisticians": When they are given a random sample, can they make predictions about the overall population? Conversely, when given some information about an entire population of relatively small size, can they make predictions about a random sample from that population? In two series of experiments we tested whether infants were able to make inductive inferences of this kind, using the violation-of-expectancy looking time methodology (11, 12).

Current Research

The first series of experiments asked whether 8-month-old infants could use the information in a sample to make inferences about a larger population. In Exp. 1, 8-month-old infants watched some events unfold on a puppet stage. Each infant was first given a set of six ping-pong balls in a small container to play with for a few seconds; half of the ping-pong balls were red, half were white. Then the infant was shown four familiarization trials. On each trial, a large box was brought onto the stage. The experimenter opened the front panel of the box and drew the infant's attention to the box. The box contained either mostly red ping-pong balls and a few white ping-pong balls or mostly white ping-pong balls and a few red ping-pong balls. The experimenter showed the infants these two displays alternately; thus the infants were equally familiarized with each display. Then the test trials began (see Fig. 1 for a schematic representation of the test events). On each test trial, the same box was brought onto the stage, its content not known to the infants. The experimenter shook the box for a few seconds, closed her eyes, reached into the top opening, and pulled out a ping-pong ball. She then placed it into a transparent sample display container next to the large box. A total of five ping-pong balls were drawn from the box, one at a time. In half of the test trials, a sample of four red and one white ping-pong balls were drawn. In the other half of the test trials, a sample of one red and four white ping-pong balls were drawn. After the five ping-pong balls were placed in the sample display container, the experimenter opened the front panel of the box to reveal its content. The infant's looking time was recorded. The experimenter then cleared the stage and started the next test trial until a total of eight test trials were completed. Only one outcome display was shown for each infant, either the mostly white or the mostly red one. On alternate test trials, the infants were shown the two samples (four red and one white or one red and four white). For an infant who saw the mostly red outcome display when the box was opened, the four red and one white sample was more probable and therefore expected, whereas the four white and one red ball sample was much less probable and therefore unexpected,[†] assuming each set was a random sample from the box. For an infant who saw the mostly white outcome display, the converse was true.

The infants looked reliably longer at the unexpected outcome ($M = 9.9$s) than the expected outcome [$M = 7.5$ s; $F(1,19) = 9.422$, $P < 0.01$; Fig. 2]. It appears that infants were able to predict the content of the box from which the samples had been drawn.

Exp. 2 replicated the results of Exp. 1 with a different group of infants. Again, the infants looked reliably longer at the unexpected outcome ($M = 7.6$ s) than the expected outcome [$M = 5.9$ s, $F(1,19) = 5.956$, $P < 0.05$; Fig. 2].

Exp. 3 was designed to test the alternative interpretation that infants may have simply preferred to look at the unexpected outcome in Exps. 1 and 2 because of the mismatch between the sample and the population. Another group of 8-month-old infants was tested. The experimental procedure was the same as before, with the following critical difference: Instead of drawing the sample of ping-pong balls from the box, the experimenter pulled them out of her pocket and placed them in the small container beside the box. It was clear to the infants that the

---

Author contributions: F.X. designed research; V.G. performed research; F.X. analyzed data; and F.X. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[*] To whom correspondence should be addressed. E-mail: fei@psych.ubc.ca.

[†] The box contained 70 red balls and five white ones (or five red balls and 70 white ones). The probability of drawing a random sample of four red and one white balls in a particular order is $70/75 \times 69/74 \times 68/73 \times 67/72 \times 5/71 = 0.0531$. The probability of drawing a random sample of one red and four white balls in a particular order is $70/75 \times 5/74 \times 4/73 \times 3/72 \times 2/71 = 0.000004056$. Alternatively, the probability of drawing a sample of four red and one white balls irrespective of order is $0.0531 \times 5 = 0.2655$, and the probability of drawing a sample of one red and four white balls irrespective of order is $0.000004056 \times 5 = 0.0000228$. We do not know which of these two calculations (if either) underpin the infant's performance in our tasks, although the latter may seem more plausible.

www.pnas.org/cgi/doi/10.1073/pnas.0704450105

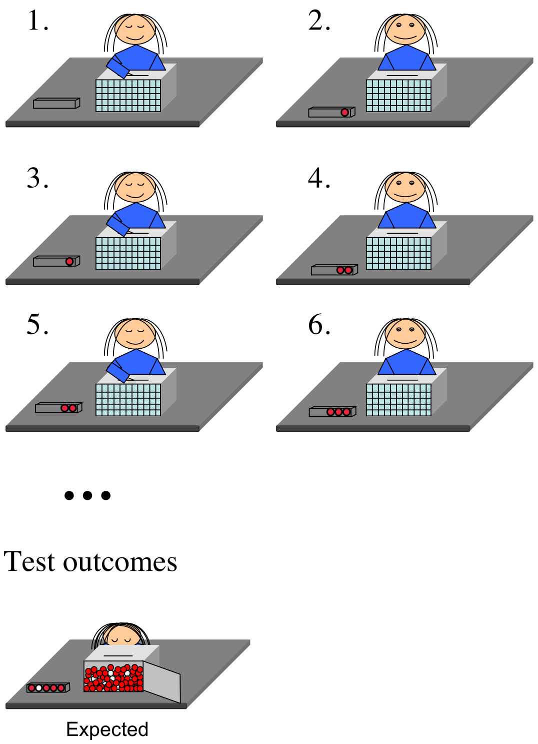

Fig. 1. Schematic representation of the test events in Exp. 1. (Images 1, 3, and 5) The experimenter shook the box for a few seconds, closed her eyes, reached into the top opening, and pulled out a ping-pong ball. (Images 2, 4, and 6) She then placed the ball into a transparent sample display container next to the large box. Test outcomes are shown at the bottom.

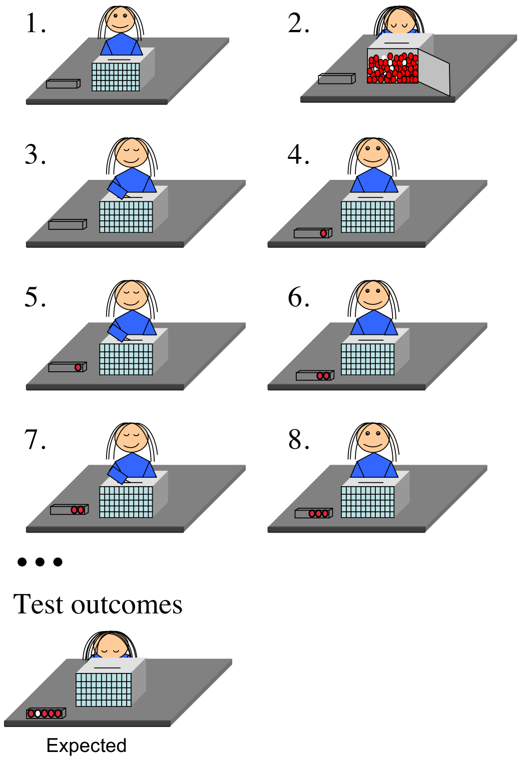

Fig. 3. Schematic representation of the test events in Exp. 4. (Image 1) The box was brought onto the stage. (Image 2) The experimenter opened the front panel of the box and let the infant look at its content for 5 s. (Images 3, 5, and 7) The experimenter closed the front panel of the box, shook the box for a few seconds, closed her eyes, and reached into the top opening. She pulled out a ping-pong ball. (Images 4, 6, and 8) She placed the ball in the transparent sample display container next to the large box. Test outcomes are shown at the bottom.

sample did not come from the box because the experimenter did not look into the top of the box and she did not reach in to pull out the ping-pong balls. The results showed that the infants looked about equally at the matched outcome (*M* = 9.1 s) and the mismatched outcome (*M* = 9.0 s; not significant).

The first series of experiments (Exps. 1–3; Fig. 2) showed that 8-month-old infants were able to use the information in a sample to make inferences about a larger population. Their looking time pattern could not be explained by any intrinsic preferences for the outcomes.

The second series of experiments asked whether 8-month-old infants were intuitive statisticians in the converse direction: Can they use the information about a population to make predictions about a sample drawn from it?

In Exp. 4, 8-month-old infants were asked to use the base rate of red and white ping-pong balls in the box to predict which of two samples was more probable. That is, they were given the base rate information of the population. The setup was the same as in the first series of experiments. As in Exp. 1, infants were first given a few ping-pong balls to play with for a few seconds, followed by four familiarization trials. Then the test trials began (see Fig. 3 for a schematic representation of the test events). On each test trial, the box was brought onto the stage. The experimenter opened the front panel of the box and let the infant look at its content for 5 s. The box contained either mostly red or mostly white ping-pong balls. Then the experimenter closed the front panel of the box, shook the box for a few seconds, closed her eyes, and reached into the top opening. She pulled out a ping-pong ball and placed it in the transparent sample display container next to the large box. She repeated this sequence until a total of five ping-pong balls had been drawn from the box. On alternate test trials, she pulled out either four red and one white ping-pong balls or one red and four white ping-pong balls. After all five balls were placed in the sample display container, the infant's looking time was recorded. The experimenter cleared the stage and started the next test trial until a total of four test trials were completed. If the infants could use the base rate

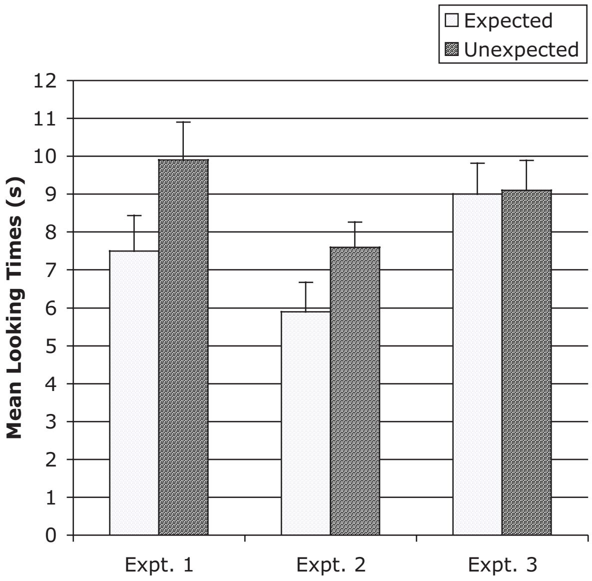

Fig. 2. Mean looking times for Exps. 1–3 with standard error.

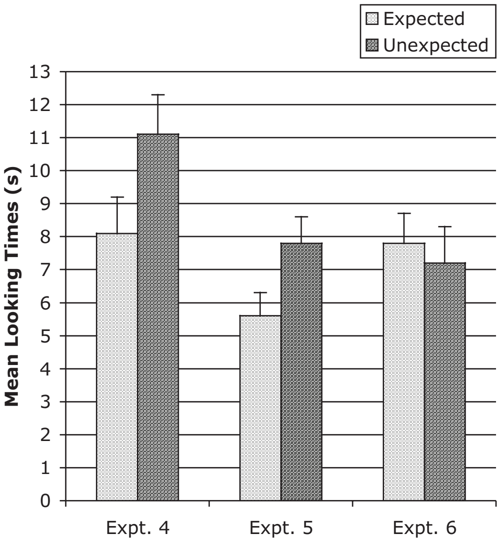

Fig. 4. Mean looking times for Exps. 4–6 with standard error.

information about the population to make predictions about the sample, they would find half of the samples much less probable and therefore unexpected. Suppose an infant was shown a box with mostly red ping-pong balls, she would expect a random sample from the box to consist of mostly red ping-pong balls. The converse was true for an infant who was shown a box with mostly white ping-pong balls.

Infants looked reliably longer at the unexpected sample (M = 11.1 s) than the expected sample [M = 8.1 s, $F(1,19) = 4.856, P < 0.05$; Fig. 4]. That is, the infants had used the base rate information provided at the beginning of each test trial to predict which of two samples was more probable.

Exp. 5 replicated the results of Exp. 4 with a different group of infants. Again, the infants looked reliably longer at the unexpected sample ($M = 7.8$ s) than the expected sample [$M = 5.6$ s, $F(1,19) = 4.706$, $P < 0.05$; Fig. 4].

Exp. 6 was a control study similar to Exp. 3, testing the alternative interpretation that infants may have simply preferred to look at the unexpected outcome in Exps. 4 and 5 because of the mismatch between the sample and the population, even though the population (i.e., the content of the box) was not shown at the end of the test trials. Another group of 8-month-old infants was tested. The procedure was the same as that of Exp. 4 except that the ping-pong balls were pulled out of the experimenter's pocket and not out of the box. Infants looked about equally at the matched outcome ($M = 7.8$ s) and the mismatched outcome ($M = 7.2$ s; not significant).

The second series of experiments (Exps. 4–6; Fig. 4) showed that 8-month-old infants can use base rate information about the population to make inferences about which of two samples is more probable. Their looking time pattern could not be explained by any intrinsic preferences for the test outcomes.

The infants' performance in these studies is impressive given the short familiarization on each trial. The infants made a connection between the sample and the population quickly, and the looking time patterns suggested that they shared adults' intuition about how random samples should be related to larger populations. It is even more impressive that the infants were able to succeed in the second series of experiments when the memory demand was high; they remembered the content of the box and made predictions about the samples drawn from it. Again, the looking time patterns suggested that the infants shared adults' intuitions about how to use information about the population to make predictions about the samples.

Previous statistical learning studies with infants have focused mostly on how infants can use frequency or transitional probability information to find larger, meaningful units from the input, e.g., segmentation of the speech stream into potential words or sequences of visual shapes (13–15). Our experiments focus on how infants can make use of a small amount of data to make inductive inferences about larger populations, and conversely, to make inferences from populations to samples, much like the kind of inference tasks all human learners face regularly in everyday life.

Several questions remain open. First, what is the computation that underlies the infants' performance in these experiments? At least two possibilities are consistent with the current findings. One possibility is that infants had computed probabilities of the samples and used them to predict which of the outcomes was more probable. Another possibility is that infants had expected the sample to be related to the distribution of the population in a qualitative way. This may be a version of the representativeness heuristic (16). One way to distinguish these two alternatives is to contrast a 4:1 sample and a 5:0 sample; the former provides a better distributional match, whereas the latter is more probable. A recently published paper (17) using a different methodology with 12-month-old infants suggests that by the end of the first year infants may be able to estimate probabilities in making predictions about future events. It remains to be seen whether younger infants are also able to do so.

Second, the current studies did not distinguish between the infants viewing the display as consisting of red and white individual objects (a discrete variable) or areas of red and white color patches (a continuous variable). The results may be described as infants being sensitive to sampling discrete objects or sampling from a continuous region. The continuous variable alternative seems unlikely for two reasons. One is that the infants were given some red and white ping-pong balls to play with before the experiment, so unambiguous evidence was provided that objects were to be presented. The other is that the literature on number discrimination in infants suggests that with large numerosities (>3 or 4), infants were not particularly sensitive to continuous variables such as overall area, brightness, and density; instead they were able to respond in various tasks based on numerosity (18). Nonetheless, further empirical work may address this issue more directly by using balls of different sizes.

Third, one important assumption in a statistical inference task, using probability or heuristics, is the assumption of random sampling. It is only under conditions of random sampling that the inference is warranted. If the learner has evidence that she is not receiving a random sample from the population, she can no longer use the statistical information in the sample to make guesses about the overall population.

Recent studies in our laboratory suggest that 11-month-old infants are sensitive to sampling conditions. Infants were randomly assigned to three conditions. One condition was similar to that of Exp. 1, and we replicated the basic results with this age group. The second condition was one in which the infants were given evidence that the random sampling assumption had been violated. The experimenter first expressed a preference for one type of ping-pong balls, say red ones, by showing the infant a small container with both types of ping-pong balls (red and white) and selectively picking up only the red ones and placing them in another container. On the test trials, the experimenter looked into the box while pulling out the ping-pong balls, i.e., she had visual access to the content of the box. If infants were sensitive to the fact that the random sampling assumption had been violated, their looking times should no longer be predicted by which sample was more probable given the content of the box. Instead their looking times should be predicted by whether the sample was consistent with the experimenter's preference or not. That was exactly what we found; infants looked longer when the experimenter pulled out a sample that was inconsistent with her expressed preference, regardless of the content of the box. In the third condition, the experimenter expressed a preference but she was blindfolded during the sampling process. The infants were able to integrate these two sources of information, and their looking times were once again predicted by the content of the box (F. Xu and S. Denison, unpublished data). These results suggest that infants engaged in a rather sophisticated form of statistical inference in this task, and their looking time patterns on the test trials were not simply a matter of matching the sample to the overall population in terms of distribution.

The present studies provide evidence that early in development infants are able to use a powerful statistical inference mechanism for inductive learning. They can make generalizations about a population based on a sample, and conversely, they can make predictions about a sample given information about a population. This ability for performing intuitive statistics develops early and in the absence of schooling or explicit teaching. It may be the roots of later acquisition of statistical principles, in both the course of developing an understanding of scientific inquiry and learning about probabilistic reasoning and statistics. This inference mechanism is likely to be present in many domains, and it allows human learners to acquire knowledge and skills rapidly and accurately. It remains to be seen how general this inductive learning ability is and whether we share it with nonhuman animals (19, 20).

These findings bear on two debates on the origins of human knowledge and reasoning. First, we have demonstrated that 8-month-old infants are able to make inferences from samples to populations, and vice versa, suggesting that such abilities are not entirely the result of formal education. Human beings may be rational learners from very early in development. Second, some cognitive scientists have suggested that "children are scientists" in how they represent clusters of concepts and how their knowledge structure changes over time (21–23). One way to test this claim is to investigate whether children's learning mechanisms are qualitatively similar to those inference mechanisms used by scientists. Here, we provide some evidence that early in development infants are intuitive statisticians and the statistical inference mechanisms they use may be qualitatively similar to the mechanisms used in scientific inquiry.

Methods

Infant Experiment. ***Subjects.*** All infants were recruited from the Vancouver area. They were all full term, and their ages ranged from 7 months, 15 days to 8 months, 15 days. Half were boys, and half were girls.

General procedure. While sitting in a high chair all infants watched the events unfold on a puppet stage. The parents sat next to the infants and faced away from the stage. They were instructed not to look at the displays during the study. An experimenter wearing a light blue cape sat behind the stage. The stage was lit, and the rest of the room was dark. Each infant was first given three white and three red ping-pong balls to play with for a few seconds. The experimenter picked up each ping-pong ball and handed it to the infant. The ping-pong balls were taken away, and the familiarization trials began. Each infant was shown a set of four familiarization trials and eight (or four) test trials. Infants' looking times were recorded. Each trial ended when the infant looked away for 2 consecutive s. A video camera below the stage focused on the infant's face and recorded the entire session. An observer sat in a corner of the testing room, watched the infant on a TV monitor, and recorded the looking times by depressing a computer key. A computer program was used to record the looking times. The observer had no knowledge of the order of the trials. A second observer coded the data from the video, and interobserver reliability averaged 92%.

The box was divided into three compartments. The front compartment was filled with 70 red ping-pong balls and five white ones; the back compartment was filled with 70 white ping-pong balls and five red ones. These compartments were separated by two pieces of white cardboard to create a middle compartment where the samples were placed. The front and back of the box were decorated identically. Thus the infants could not tell whether the box contained mostly red or mostly white ping-pong balls from inspecting the front of the box. The middle compartment was created to ensure that the correct sample was drawn each time. When the sample was drawn from the box, it appeared to be drawn from a big box filled with ping-pong balls.

Adult Rating Study. ***Subjects.*** Sixteen adults (mean age 26.7, ranging from 19 to 51 years; eight males and eight females) rated the test events shown to infants in Exp. 1; 16 adults (mean age 23.1, ranging from 18 to 54 years; eight males and eight females) rated the events shown to infants in Exp. 4. None of the adult participants were aware of the purpose or the design of the experiments.

Procedure. Each adult was given a rating sheet with instructions. They were told that the video clips were filmed for a study being done with young children, and we wanted to get their reactions. There were no tricks. Adults were asked to watch each event and rate the test outcome on a 1 to 7 scale (1 being "not at all unexpected" and 7 being "very unexpected"). They were also asked an open-ended question ("what do you think this study is about") at the end of the study.

Results. For the test events in Exp. 1, the average rating for the expected outcome was 2.3 and the average rating for the unexpected outcome was 4.5 [$t(15) = -4.1$, $P < 0.001$]. For the test events in Exp. 4, the average rating for the expected outcome was 2.5 and the average rating for the unexpected outcome was 4.7 [$t(15) = -3.2$, $P < 0.005$]. One adult mentioned probability in answering the open-ended final question; everyone else simply said "I don't know" or the equivalent.

For naïve adults, the test events in Exps. 1 and 4 elicited clear responses in terms of how expected or unexpected they were. Interestingly, adults did not seem to have guessed the purpose of the experiments upon conscious reflection.

ACKNOWLEDGMENTS. We thank members of the University of British Columbia Baby Cognition Laboratory for assistance in data collection; Stephanie Denison, Kathryn Dewar, Anjula Joshi, Laura Kerlin, Mijke Rhemtulla, and Henny Yeung for helpful discussion, and two anonymous reviewers for constructive comments on an earlier version of the manuscript. This research was supported by grants from the National Science Foundation and the Natural Sciences and Engineering Research Council of Canada (to F.X.).

1. Hume D (1739/1978) *A Treatise of Human Nature* (Clarendon, Oxford).
2. Quine WVO (1960) *Word and Object* (MIT Press, Cambridge, MA).
3. Goodman N (1955) *Fact, Fiction, and Forecast* (Harvard Univ Press, Cambridge, MA).
4. Nisbett RE, Krantz DH, Jepson C, Kunda Z (1983) The use of statistical heuristics in everyday inductive reasoning. *Psychol Rev* 90:339–363.
5. Markman E (1989) *Categorization and Naming in Children* (MIT Press, Cambridge, MA).
6. Xu F, Tenenbaum JB (2007) Word learning as Bayesian inference. *Psychol Rev* 114:245–272.
7. Yuille A, Kersten D (2006) Vision as Bayesian inference: Analysis by synthesis? *Trends Cogn Sci* 10:301–308.
8. Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn Sci* 10:309–318.
9. Gigerenzer G (1989) *The Empire of Chance: How Probability Changed Science and Everyday Life* (Cambridge Univ Press, Cambridge, UK).
10. Oaksford M, Chater N (1998) *Rational Models of Cognition* (Oxford Univ Press, Oxford).
11. Onishi KH, Baillargeon R (2005) Do 15-month-old infants understand false beliefs? *Science* 308:255–258.
12. Spelke ES, Breinlinger K, Macomber J, Jacobson K (1992) Origins of knowledge. *Psychol Rev* 99:605–632.
13. Aslin R, Saffran JR, Newport EL (1998) Computation of conditional probability statistics by human infants. *Psychol Sci* 9:321–324.
14. Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
15. Kirkham NZ, Slemmer JA, Johnson SP (2002) Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition* 83:B35–B42.
16. Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–1131.
17. Teglas E, Girotto V, Gonzalez M, Bonatti L (2007) Intuitions about probabilities shapes expectations about the future at 12 months and beyond. *Proc Natl Acad Sci USA* 104:19156–19159.
18. Xu F, Spelke ES (2000) Large number discrimination in 6-month-old infants. *Cognition* 74:B1–B11.
19. Gallistel R (1990) *The Organization of Learning* (MIT Press, Cambridge, MA).
20. Hauser MD, Spalding B (2006) Wild rhesus monkeys generate causal inferences about possible and impossible physical transformations in the absence of experience. *Proc Natl Acad Sci USA* 103:7181–7185.
21. Kuhn T (1962) *The Structure of Scientific Revolutions* (Univ of Chicago Press, Chicago).
22. Carey S (1985) *Conceptual Change in Childhood* (MIT Press, Cambridge, MA).
23. Gopnik A, Meltzoff A (1996) *Words, Thoughts, and Theories* (MIT Press, Cambridge, MA).