Integrating Physical Constraints in Statistical Inference by 11-month-old Infants

Stephanie Denison & Fei Xu

*University of California, Berkeley*

(In press, *Cognitive Science*)

Keywords: statistical inference, learning mechanisms in infancy, physical reasoning, probabilistic reasoning
Abstract

Much research on cognitive development focuses either on early-emerging domain-specific knowledge or domain-general learning mechanisms. However, little research examines how these sources of knowledge interact. Previous research suggests that young infants can make inferences from samples to populations (Xu & Garcia, 2008) and 11- to 12.5-month-old infants can integrate psychological and physical knowledge in probabilistic reasoning (Teglas, Girotto, Gonzalez, & Bonatti, 2007; Xu & Denison, 2009). Here we ask whether infants can integrate a physical constraint of immobility into a statistical inference mechanism. Results from three experiments suggest that first, infants were able to use domain-specific knowledge to override statistical information, reasoning that sometimes a physical constraint is more informative than probabilistic information. Second, we provide the first evidence that infants are capable of applying domain-specific knowledge in probabilistic reasoning by using a physical constraint to exclude one set of objects while computing probabilities over the remaining sets.
Integrating Physical Constraints in Statistical Inference by 11-month-old Infants

1. Introduction

The focus of research on cognitive development in recent years highlights early-emerging knowledge and concepts in infancy as well as the importance of learning and inferential mechanisms. However, the majority of research in the field to date has explored either early knowledge or learning mechanisms but has not investigated the interaction of the two. The present study explores whether or not young infants can apply domain-specific physical knowledge within a domain-general learning mechanism.

The literature exploring early knowledge reveals that infants and young children possess domain-specific knowledge to facilitate reasoning about agents, number, objects, and causality (e.g., Baillargeon, 2004; Gergely, Nadasdy, Csibra, & Biro, 1995; Leslie & Keeble, 1987; Spelke, 1994; Woodward, 1998; and Xu & Spelke, 2000; see Carey, 2009, for a review). One domain that has been extensively investigated in early infancy is that of physical knowledge. Evidence of infants’ knowledge in this domain is apparent at as young as 2 months of age (Baillargeon, 2008; Spelke, Breinlinger, Macomber, & Jacobson, 1992). According to the core knowledge thesis, when reasoning about physical events, infants use three basic principles—cohesion, continuity, and contact— to make inferences about the behavior of hidden objects (Spelke, 1994). A great deal of research has been conducted to explore at what ages infants begin to reason in accord with these core principles and to what extent this understanding facilitates their interpretation of more complex physical events such as occlusion and containment (see Baillargeon, 2008, for a review).
The current experiments focus on infants’ ability to reason about the immobility of objects in the context of statistical inference. Immobility is closely related to the principle of cohesion, which states that objects should move as connected, bounded units that do not spontaneously break apart or merge together with other objects (Spelke, 1994). The understanding of the principle of cohesion has been explored in infants using looking-time measures and experimental evidence suggests that by approximately 3 months of age, infants are capable of reasoning about physical objects in accord with cohesion.

Parallel to the research conducted on early domain-specific knowledge, researchers have also focused on infants’ and young children’s use of powerful learning mechanisms that allow them to compute various statistics over the input they receive from the environment. Findings suggest that infants are capable of making rational generalizations based on a relatively small amount of data across several domains, including speech perception, word learning, causal learning, preference attribution, and physical reasoning (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004; Kushnir, Xu, & Wellman, in press; Maye, Werker, & Gerken, 2002; Sobel & Kirkham, 2006, 2007; Xu, 2007; Xu & Garcia, 2008; Xu & Tenenbaum, 2007).

Three recent studies investigating early domain-general learning mechanisms show that infants are capable of rudimentary probabilistic reasoning. First, Teglas, Girotto, Gonzalez, and Bonatti (2007) found that 12.5-month-old infants could reason about single-event probability in a violation of expectancy looking-time paradigm. They showed infants a scene with three identical objects and one difference object in a lottery machine and found that infants looked longer when the single different object (less probable) exited the machine than when one of the (more probable) identical objects did. Second, Denison and Xu (in press) also found that 12- to 14-
month-old infants could make inferences about single-event probability with large set sizes. Finally, Xu and Garcia (2008) conducted a series of experiments investigating a statistical inference mechanism involving large populations and multi-object samples. They tested 8-month-old infants in a violation of expectancy looking-time paradigm. In these experiments, infants saw a sample of either four red balls and one white ball or four white balls and one red ball being drawn from a covered box on alternating trials. Then the experimenter revealed a population of, for example, mostly red Ping-pong balls. Infants looked longer at the less probable sample of four white and one red balls, indicating that this outcome violated their understanding of probability. Taken together, these findings suggest that infants can use statistical information to make inferences about the composition of populations given data obtained from small samples and vice versa.

Given that young learners possess both substantive domain-specific knowledge and powerful statistical inference mechanisms, the question arises as to whether infants can take into account domain-specific knowledge while engaging in statistical learning. It is possible that infants are not capable of incorporating domain-specific knowledge into domain-general learning mechanisms and that these systems develop separately early in life. However, it is also possible that infants’ initial conceptual knowledge interacts with input statistics in informative ways, allowing them to use both types of knowledge simultaneously. Due to the fact that most infancy research investigates either initial knowledge or statistical learning mechanisms, we do not have much evidence to suggest which of these alternatives represents the true state of affairs. However, some evidence exists with older children and adults suggesting that they can make inferences, drawing on domain-specific knowledge when computing statistics from the input
(e.g., Kushnir & Gopnik, 2007; Newport & Aslin, 2004; Schulz, Bonawitz, & Griffiths, 2007; Schulz & Gopnik, 2004).

Two recent studies also suggest that infants can apply domain-specific knowledge in statistical inference (Teglas et al., 2007; Xu & Denison, 2009). In the experiments conducted by Teglas et al. (2007, described earlier in the section), a control experiment was included which required 12.5-month-old infants to use their knowledge of solidity when deploying a probabilistic inference mechanism. They placed a barrier in the middle of the lottery machine with the three identical objects positioned above the barrier and the one different object positioned below the barrier. In this case, infants looked longer when one of the identical objects exited from the bottom of the machine than when the different object exited, suggesting that they recognized the former as an impossible event. This suggests that 12.5-month-old infants were able to override probabilistic information and instead use their knowledge of solidity to reason about the likelihood of the test events.

A second experiment investigated whether 11-month-old infants can integrate knowledge regarding agents into a statistical inference mechanism (Xu & Denison, 2009). We investigated whether infants could reason about psychological constraints, namely an agent’s goal and visual access, in the context of Xu and Garcia’s (2008) statistical inference task. The procedure was similar to Xu and Garcia (2008) except that the experimenter conveyed a goal to the infants of obtaining one color of balls over the other, say the white balls. In the “visual access” condition, the experimenter looked directly into the box while removing a sample of either five white or five red balls from the box during the test phase. In the “blindfolding” condition, the experimenter removed the same samples while wearing a blindfold. Infants in both conditions were then shown a population of, for example, mostly red balls and just a few white balls. Infants
in the “visual access” condition looked longer at the sample of five red balls than five white balls, presumably because this was inconsistent with the experimenter’s goal. In contrast, infants in the “blindfolding” condition looked longer at the sample of five white balls than five red balls, presumably because the blindfold had prevented the experimenter from drawing a sample consistent with her goal therefore she should have drawn a random sample from the box. These findings suggest that infants were able to integrate domain-specific knowledge regarding agents into a statistical inference mechanism in a meaningful way. They were able to override probabilistic information in favor of domain-specific knowledge and vice versa, under the appropriate sampling conditions.

Although the experiments conducted by Teglas et al. (2007) and Xu and Denison (2009) provide some evidence that infants can integrate domain-specific knowledge and statistical inference mechanisms, the findings leave open a number of questions. First, given that infants could reason about solidity in the Teglas et al. experiment, we asked whether infants could reason about a different physical constraint (i.e., immobility) using the paradigm of Xu and Garcia (2008). Second, and more importantly, both Teglas et al. and Xu and Denison (2009) required infants to reason about a constraint when deploying the statistical inference mechanisms only to the extent that infants had to determine which of the two sources of information they should appeal to when faced with a conflict between substantive domain knowledge and probabilistic information. Therefore, we ask whether infants can integrate initial knowledge into a statistical inference mechanism in a situation that does not allow them to completely disregard one type of knowledge but instead requires them to use both domain-specific knowledge and statistical information.
The current experiments investigate whether 11-month-old infants can integrate a physical constraint, namely immobility of objects, into a statistical inference mechanism. In Experiment 1 we ask whether infants can detect a physical constraint using the same methodology as Xu and Garcia (2008) and Xu and Denison (2009). We test the hypothesis that, if infants are sensitive to a violation of immobility, they will use this information to override the probabilistic information available to them when making inferences from samples to populations. Experiment 2 serves as a replication of earlier findings, as well as a control for an alternative interpretation of the looking-time pattern found in Experiment 1. In Experiment 3, we ask whether infants can use knowledge regarding physical constraints when deploying the statistical inference mechanism. In this experiment, we show infants a box with three sets of Ping-pong balls in a 5:4:1 color ratio and ask whether they can exclude one of these sets of balls and then compute probabilities over the remaining sets.

2. Experiment 1

Method

Participants

Participants were sixteen 11-month-old infants (8 girls and 8 boys; mean age = 10;26 [months;days], range = 10;16 - 11;14). An additional 3 infants were tested but not included in the final sample due to experimenter error (1) or not completing the study due to fussiness (2). Infants were recruited via phone from the greater Vancouver area. In all experiments, parents filled out a consent form at the beginning of the experiment and infants received a T-shirt or bib and a diploma for their participation.
Materials

Ping-pong Balls. A total of 236 (118 green and 118 red) Ping-pong balls were used. The balls were separated into two sets. One set included green balls with six white Velcro strips glued to them (strips were approximately .8cm x 1.5cm) so that there was one strip on each side of each ball; the red balls had no Velcro. The other set contained red balls with Velcro and green balls with no Velcro. The Velcro signified that one of the sets of balls had the property of being physically constrained or immovable and each infant only saw one of the two sets of balls. For the remainder of the Method section, we will describe the condition in which the red balls had Velcro and were therefore immovable or stuck inside the box.

Boxes and Containers. A small white box (17.5cm x 17.5cm x 8cm) without a top, constructed out of foam core was used to hold three green and three red Ping-pong balls; the red balls were glued to the box.\footnote{Balls with Velcro (i.e., immobile or constrained balls) were actually glued to the insides of the boxes and containers rather than simply Velcroed. This was done because simply Velcroing the balls to the containers was not sufficient for preventing the balls from being removed from the surfaces when the infants pulled on the balls at the beginning of the experiment and when the experimenter shook the boxes during trials.} Infants were allowed to play with the balls in the box during the Free Play Phase at the beginning of the experiment.

A small Plexiglas container (20cm x 4.5cm x 4cm) was placed at the front left-hand corner of the stage to display the five-ball samples pulled out of the box during test trials.

A slightly longer Plexiglas container (28.5cm x 4.5cm x 4cm) was used during Demonstration Phase 1 (see Procedure section below). This container held six balls, three red (with Velcro) and three green (without Velcro). The red balls were glued to the inside bottom surface of the container. Both of the Plexiglas containers were narrow enough such that the Ping-pong balls lined up in a single row when placed inside.
A small white box made from foam core, fabric and Plexiglas (27cm x 16cm x 13cm) was used in Demonstration Phase 2 (see Procedure section below). Purple fabric (secured to the top of the box by Velcro) covered the front of the box and the fabric could be lifted to reveal the front Plexiglas window. The top of the box had a 13cm x 8cm cutout for the experimenter to reach into to access the balls in the box. This box contained 24 Ping-pong balls, 12 green and 12 red. Each red ball was glued to one of the six inside surfaces of the box.

A large, 39cm x 34cm x 22cm box constructed out of foam core, fabric, and Plexiglas was used to display the large population outcomes on the familiarization and test trials. The box was a white rectangular cube lined with black duct tape around all edges. The inside of the box was divided into three parts with two Plexiglas containers inserted into the front and back of the box, each containing 72 Ping-pong balls, and a hidden center compartment used to hold the samples to be removed from the box during test trials. When viewed from the front or the back, the box appeared to be one large box filled with Ping-pong 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 Plexiglas window. The “mostly red” side of the box contained 60 red and 12 green balls (red : green = 5:1), the “mostly green” side contained the opposite ratio (red : green = 1:5); the red balls were covered with Velcro strips. The top of the box had a 10cm x 24cm cutout covered with two pieces of overlapping spandex that allowed the experimenter to reach into the center compartment of the box.

**Apparatus**

Testing took place in a quiet room. The room was divided in half by two curtains spanning the width and height of the room. All events were presented on a stage. There was a black curtain on the back of the stage, attached at the top of the stage by Velcro. The
experimenter sat behind the stage. When the back curtain was lifted, her upper body and head were visible to the infant and when it was pulled down, the experimenter was no longer visible. The viewable area of the stage measured 94cm x 55cm (width x height). An observer watched the infant on a TV monitor in one corner of the testing room and recorded each infant’s looking times on a Macintosh iBook using MacXHAB 1.4 (J. Pinto, 2005). The observer was blind to the order of the trials. An Optimus fan set at low speed was located in the back part of the room to muffle sounds from the hallway. The stage was lit; the rest of the room was darkened during the study.

The infant sat approximately 70cm from the stage in a high chair and the parent sat next to the infant facing away from the stage. The parent was instructed not to influence the infant’s looking behavior in any way and to face away from the stage throughout the experiment. Two cameras recorded the session, one focused on the stage to record the procedure, the other focused on the infant’s face to record her looking behavior.

Procedure

All infants were tested in a violation-of-expectancy looking time paradigm. The experiment began with the experimenter calibrating the observer to each infant’s looking behavior. Using a small squeaky toy, the experimenter oriented the infant to the outer limits of the viewable area of the stage while the observer watched the infants’ eyes on the TV screen. This allowed the observer to know when the infant was looking at the stage and when he/she was looking away from the stage.

Free Play Phase. Infants were first shown the small foam core box with three red (with Velcro) and three green (no Velcro) Ping-pong balls and were permitted to play with them for about 1 minute. The experimenter shook the balls around in the box and she encouraged the
infants to try to pick up some balls of each color, giving them the opportunity to notice that the green balls were easily removed from the box and the red balls were stuck.

_Demonstration Phase 1._ The experimenter went behind the stage and placed the small transparent container on the front of the stage where it remained throughout the experiment. The experimenter then brought out the longer transparent container, which held three red balls (with Velcro) and three green balls in a random order. Next the experimenter drew attention to each of the six balls, one color first, then the other color. She did this by picking up each of the green balls while saying, “Look at this one! See this one?” She then grasped each red ball (which did not move) as if she was trying to pick it up and said, “Look at this one; it’s stuck!” The demonstration lasted for approximately 30 seconds after which the experimenter cleared the container from the stage. The purpose of this phase was to demonstrate that there were two colors of balls in the experiment and that the balls with Velcro were immovable.

_Familiarization trials._ Each infant received four familiarization trials. On each trial, the experimenter placed the large box on the stage with the front curtain closed. She shook the box back and forth a few times, saying, “What’s in the box?” Then, with one hand, she lifted the front cover of the box and with the other, she simultaneously lowered the backdrop of the stage while saying “Look, [baby’s name], look!” The observer began timing upon hearing the second, “look”, as this was precisely the moment that the experimenter lifted the cover on the box to reveal the population of either mostly red (with Velcro) or mostly green (without Velcro) balls. Once the backdrop was lowered, the experimenter was no longer visible to the infant. The trial ended when the infant looked away for 2 consecutive seconds. The four trials alternated between the mostly green population and the mostly red population. The large box was removed after
each trial and the black curtain was lowered between trials. The Familiarization phase lasted approximately 4 minutes.

These trials were included to familiarize infants to the large box and the objects, as well as to the general procedure of the study. Also, once infants have been exposed to two populations, one of mostly green balls, the other of mostly red, they can use this information during test trials to generate a hypothesis as to which box the experimenter might be sampling from before the population is revealed.

Demonstration Phase 2. The experimenter brought the small white box onto the stage. She lifted the front curtain so the infants could see the balls in the box, flipped the box upside down and then right side up, turned the box side to side and shook it and placed it back on the stage. She reached into the box, her hand visible to the infant through the transparent front window. Next, she picked up one green ball and lifted it to the top of the box but did not bring it outside of the box while saying, “See this one? Look at this!” Then she grasped one of the red balls (which was glued to the inside surface of the box) and said, “Look at this one. It’s stuck!” The experimenter lowered the front curtain on the box. She then lifted the curtain again and said, “Should we do that one more time?” and repeated the entire sequence. Following the second sequence, she removed the box from the stage. This phase lasted approximately 45 seconds.

This phase was included so that infants could see that balls with Velcro were stuck inside the box and did not move even when the box was shaken vigorously. This phase also allowed infants to experience a sampling procedure that is totally transparent: they see the experimenter’s hand reach into the box to pick up the Ping-pong balls. This should make the occluded sampling during test trials less mysterious.
Test trials. Each infant received six test trials (see Fig. 1 for a schematic representation of the test trial procedure). 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 three Ping-pong balls of one color (red or green) and placed them into the small Plexiglas container to her right one at a time. She then repeated this action, pulling out two more balls of the same color. The small Plexiglas container held 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!” Just as in the Familiarization trials, the observer began timing upon hearing the second “look”, indicating the exact moment that the population was visible to the infant. The trial ended when the infant looked away for 2 consecutive seconds. Each infant saw one of the two populations (mostly green or mostly red) on all six test trials. The samples alternated between five mobile balls (e.g., 5 red balls) and five immobile balls (e.g., 5 green balls) on each trial, resulting in three sets of two trials. At the end of each trial, the stage was cleared. Test trials lasted approximately 7 minutes.

Design

The population of the box (mostly green or mostly red) used for test trials and the color of balls with Velcro were fully crossed between infants. Half of the infants saw the mostly red population on all test trials, and the other half saw the mostly green population. Half of the infants who saw the mostly green box on test trials saw green balls with Velcro strips and the other half saw red balls with Velcro strips. The same was true for infants who saw the mostly red box. The order of the familiarization trials (mostly red first or mostly green first), the order during the demonstration phases (mobile-object samples first or immobile-object samples first),
and the order of the samples on the test trials (red or green first) were counterbalanced across infants.

Predictions

We predicted that the 8 infants who saw red balls with Velcro should look longer at the five red ball sample than the five green ball sample. That is, seeing five balls that were stuck being removed from the box should violate their expectations, as this is an impossible event. For the 8 infants who saw green balls with Velcro the opposite prediction holds. Therefore, regardless of the population of the box as mostly red or mostly green, if infants are sensitive to the physical constraint, they should look longer when the sample consists of five balls with Velcro than when the sample consists of five balls without Velcro.

If, however, infants are unable to recognize and integrate the physical constraint, the proportion of balls in the box should predict their looking behavior. That is, infants should look longer at a sample that is less probable than one that is more probable given the population of Ping-pong balls in the box. For example, for the 8 infants who saw the mostly red box on test trials, a sample of five green balls is less probable than a sample of five red balls, thus infants should look longer at the five green ball sample.

Results

Preliminary analyses found no effects of gender, content of the box (mostly red or mostly green), color of balls with Velcro, or order of the familiarization trials. Subsequent analyses collapsed over these variables. A second observer who was unaware of the order of the trials timed all familiarization and test trials. Inter-observer reliability averaged 94%.

An alpha level of .05 was used for all analyses. An analysis of variance (ANOVA) examined the effects of test trial order (mobile objects sample outcome first or immobile objects
sample outcome first), trial pair (1, 2, and 3), and outcome (mobile vs. immobile objects samples). There was a main effect of outcome, $F(1,15) = 15.12, p = .001$, effect size ($\eta^2$) = .33. Infants looked reliably longer at the outcomes with samples containing immobile objects ($M = 9.08s, SD = 3.91$) than outcomes with samples containing mobile objects ($M = 6.47s, SD = 2.40$) (see Fig. 2). There were no other main effects or interactions. Thirteen of sixteen infants looked longer at the immobile objects outcomes, Wilcoxon signed-ranks test: $z = -3.10, p = .001$.

However, there exists an alternative interpretation of the data. Due to the counterbalancing of the experiment, half of the infants saw a mostly red box on test trials and half of the infants saw a mostly green box on test trials. Within this, half of the infants saw green balls with Velcro (i.e., green balls were stuck) and half of the infants saw red balls with Velcro. This counterbalancing results in the creation of two groups: One group of infants, termed the Consistent group ($n = 8$), saw a probabilistic cue and a constraint cue that should predict the same looking time pattern (e.g., they saw a mostly green box and the red balls had Velcro) thus a sample of five red balls is unexpected both in terms of probability and constraints. The other group of infants, termed the Inconsistent group ($n = 8$) saw a probabilistic cue and a constraint cue that should predict opposite looking time patterns (e.g., they saw a mostly red box and the red balls had Velcro) thus a sample of five red balls is expected in terms of probability but it is unexpected in terms of the physical constraint—the cues were in conflict. The initial analysis does not control for the possibility that infants were insensitive to the physical constraint and that the infants in the Consistent group drove the entire main effect of outcome. To control for this possibility, we re-analyzed the data with the trials coded according to the probability of obtaining the samples from the box if all balls were free to move (i.e., assuming that infants were reasoning about the proportions of the samples and the box and not the physical constraint). For example, if
infants saw a mostly red box, the five red ball sample is more probable than the five green ball sample regardless of which balls had Velcro, if the physical constraint is ignored.

Therefore, we conducted an ANOVA examining the within-subjects effects of trial pair (1, 2, and 3), and outcome (majority-color sample vs. minority-color sample) and the between subjects factor of consistency (Consistent group versus Inconsistent group). We found no main effects of outcome (majority-color sample vs. minority-color sample), trial pair, or consistency. This suggests that infants were not reasoning solely in accord with the probabilistic constraint. More importantly, we found an interaction between outcome and consistency, $F(1,14) = 16.31, p = .001$, effect size ($\eta^2_p$) = .54 (see Fig. 3). Separating the Consistent versus the Inconsistent groups, we found that in the Consistent group, infants looked reliably longer at the minority-color sample outcome ($M = 9.54$ sec, $SD = 1.30$) rather than the majority-color sample outcome ($M = 5.98$ sec, $SD = 1.03$) trials, $F(1,7) = 10.58, p = .014$, effect size ($\eta^2_p$) = .60. Seven of eight infants looked longer at the minority-color sample outcomes, Wilcoxon signed-ranks test: $z = -2.38, p = .017$. In this group it is impossible to discern whether infants were using the probabilistic cue or the constraint cue. In the Inconsistent group, infants looked reliably longer at the majority-color sample outcome ($M = 8.63$ sec, $SD = 1.30$) versus the minority-color sample outcome ($M = 6.97$, $SD = 1.03$) trials, $F(1,7) = 5.83, p = .046$, effect size ($\eta^2_p$) = .45. Six of eight infants looked longer at the majority-color sample outcome, Wilcoxon signed-ranks test: $z = -1.96, p = .05$. Therefore, infants in the Inconsistent group looked longer at a sample of, for example, five red balls from a mostly red box, suggesting that their expectations were violated when balls that should have been stuck inside the box were removed, regardless of the fact that the probability of obtaining that sample from the population (if the constraint is ignored) was
very low. This suggests that the physical constraint overrides the probability calculation in this situation.

Discussion

When infants were given evidence of a physical constraint violation—in this case a violation of immobility—they looked longer at a sample of five balls that violated immobility than at a sample of five balls that did not violate this constraint. These results suggest that infants were sensitive to the physical constraint in this task. The results from infants in the Inconsistent group (i.e., infants who saw probabilistic information that conflicted with the physical constraint) are particularly striking. These infants were given two conflicting sources of information: First, they were shown that balls with Velcro had the property of being stuck to the inside surfaces of the box. Second, they were shown that a sample of five balls was drawn from a population of Ping-pong balls that contained a ratio of 5 balls with Velcro to 1 ball without. Therefore, the probabilistic information in the population should suggest that a sample of balls with Velcro is more probable than a sample of balls without Velcro. However, the physical constraint information should suggest that the sample of balls without Velcro is much more likely. It appears that infants properly combined these two sources of information, overriding probabilistic information in favor of substantive domain-specific knowledge.

An alternative interpretation of infants’ looking time patterns is possible, however. Because the unexpected outcome always involved five balls covered with Velcro, infants may have simply been looking longer at samples of balls with Velcro than balls without Velcro on test trials. Therefore, in Experiment 2, we attempt to control for this possibility. Experiment 2 also provides a replication of Xu & Garcia (2008) and Xu & Denison (2009).

3. Experiment 2
Method

Participants

Participants were sixteen 11-month-old infants (8 girls and 8 boys; mean age = 11;1 [months;days], range = 10;16 to 11;14). An additional 4 infants were tested but not included in the final sample due to experimenter error (2) or not completing the study due to fussiness (2).

Materials

All materials were the same as those used in Experiment 1 with one critical exception: Balls with Velcro were not glued to any boxes or containers.

Procedure

The procedure was the same as in Experiment 1 except that the balls with Velcro were not demonstrated to be stuck. Infants saw that all balls moved freely.

Free Play Period. At the beginning of the experiment, the experimenter showed infants the small white box with three red (with Velcro) and three green (without Velcro) balls for about 1 minute. She allowed the infants to play with the balls and encouraged them to pick the balls up so that the infants could see that the balls were discrete objects.

Demonstration Phase 1. The experimenter brought out the long Plexiglas container with three red balls and three green balls and picked up each ball, one color at a time, while saying, “Look at this one! See this?” This phase lasted for approximately 30 seconds, after which experimenter cleared the container from the stage.

Familiarization trials. Four familiarization trials were conducted in exactly the same manner as in Experiment 1. These trials lasted approximately 4 minutes.

Demonstration Phase 2. This phase was identical to Experiment 1 except that the balls with Velcro were movable inside the box. The experimenter simply lifted one ball of each color
while saying, “See this one? Look at this!” She completed the sequence twice. This phase lasted approximately 45 seconds.

*Test trials.* The test trials were identical to those in Experiment 1. In this experiment, test trials again alternated between five red and five green ball samples; however, these are now referred to as the minority-color objects samples and the majority-color objects samples.

*Design*

The Design was identical to Experiment 1.

*Predictions*

We predicted that infants should look longer at a sample that is less probable than one that is more probable given the population of Ping-pong balls in the box. For the 8 infants who saw the mostly red box on test trials, a sample of five green balls is less probable than a sample of five red balls; therefore they should look longer at the five green ball sample. For the 8 infants who saw the mostly green box on test trials, the opposite prediction holds.²

Alternatively, if infants simply look longer at samples of balls with Velcro, they should look longer at the samples with Velcro regardless of the probability of obtaining those balls given the proportions in the population. For half of the infants, the balls with Velcro constituted the majority of balls in the population and for the other half of the infants, the balls with Velcro constituted the minority of balls in the population. Therefore, if infants simply look longer at samples with Velcro, this should result in infants looking about equally at the more probable versus the less probable samples.

*Results*

² The probability of obtaining a sample of 5 green balls from the mostly green box = 60/72 * 59/71 * 58/70 * 57/69 * 56/68 = 0.39034 when sampling without replacement. The probability of obtaining a sample of 5 red balls from the mostly green box = 12/72 * 11/71 * 10/70 * 9/69 * 8/68 = 0.00006 when sampling without replacement. The more probable outcome is 6896 times more likely than the less probable outcome.
Preliminary analyses found no effects of gender, content of the box (mostly green vs. mostly red), color of balls with Velcro, or order of the familiarization trials. Subsequent analyses collapsed over these variables. A second observer who was unaware of the order of the trials coded all test trials. Inter-observer reliability averaged 96%.

An ANOVA examined the effects of test trial order (majority-color sample first or minority-color sample first), trial pair (1, 2, and 3), and outcome (majority-color sample vs. minority-color sample). There was a main effect of outcome, $F(1,15) = 7.35, p = .016$, effect size ($\eta^2_p$) = .33. Infants looked reliably longer at the minority-color sample outcomes ($M = 9.81s, SD = .71$) than the majority-color sample outcomes ($M = 6.21s, SD = 1.25$) (see Fig. 4). Twelve of sixteen infants looked longer at the minority-color sample outcomes, Wilcoxon signed-ranks test: $z = -2.17, p = .03$. There were no other main effects or interactions.

Next we tested the alternative hypothesis that infants simply tend to look longer at samples of balls covered in Velcro. Similarly to Experiment 1, we split infants into two groups. Infants in the Consistent group ($n = 8$) saw minority-color samples that also happened to consist of balls covered in Velcro (i.e., infants saw a mostly red box with the green balls covered in Velcro). In this group, infants would look longer at a sample of five green balls if they were reasoning about probability or if they were simply inclined to look longer at balls with Velcro. Infants in the Inconsistent group ($n = 8$) saw minority-color samples that consisted of balls not covered in Velcro (i.e., a mostly red box with red balls covered in Velcro). In this group, infants would look longer at a sample of five green balls if reasoning about probability and a sample of five red balls if they were simply looking longer at balls with Velcro. An ANOVA revealed no interaction between consistency and outcome, $F(1,14) = 0.16, p = .7$, effect size ($\eta^2_p$) = .011. Therefore, infants did not differ in their looking times at the majority-color sample outcomes and
minority-color sample outcomes depending on whether or not the balls in the samples were covered in Velcro. This finding makes the interpretation that infants simply looked longer at samples containing balls with Velcro in Experiment 1 extremely unlikely.

Discussion

When no balls were demonstrated to have the property of being stuck, infants looked longer at, for example, a sample of five green balls being removed from a mostly red box than a sample of five red balls being removed from a mostly red box. That is, infants used a probabilistic cue to make inferences about populations given the composition of a sample. This result replicated earlier findings suggesting that 8- and 11-month-old infants can make inferences from samples to populations. The results of this experiment also ruled out the interpretation that infants simply looked longer at the unexpected outcome in Experiment 1 because the physical constraint was correlated with whether or not the balls in the samples were covered in Velcro.

The findings of Experiments 1 and 2 suggest that, when infants are presented with physical knowledge that conflicts with probabilistic knowledge, they can correctly choose between the two. Infants appear to override probabilistic information and instead use information regarding a physical constraint to guide their expectations about samples and populations. Experiment 3 asks whether infants can apply a physical constraint in a more challenging situation. In Experiment 3, we show infants a large population containing three sets of balls with a ratio of 5 stuck green balls to 4 red balls to 1 yellow ball. We ask whether infants can keep in mind that the green balls are stuck and then estimate which of two samples is more likely—four red balls and one yellow ball or four yellow balls and one red ball. If infants’ looking behavior suggests that they find the sample of four yellow balls and one red ball less likely than the

---

3 A second analysis was done to examine whether infants looked longer at samples covered in Velcro. We re-coded the data such that balls with Velcro were considered unexpected and balls without Velcro were considered expected. There was no main effect of outcome, $F(1,15) = 0.11, p > .05$, effect size ($\eta^2$) = .007.
sample of four red balls and one yellow ball, this will provide further evidence that infants can incorporate domain-specific knowledge while using this probabilistic inference mechanism.

4. Experiment 3

Method

Participants

Participants were forty 11-month-old infants (20 girls and 20 boys; mean age = 11;3 months;days, range = 10;16 to 11;15). Twenty infants (10 girls and 10 boys) were randomly assigned to each of two conditions: the Physical Constraint condition (mean age = 11;01), and the No Physical Constraint condition (mean age = 11;02). An additional 6 and 3 infants in each condition were tested but not included in the final sample due to experimenter error (1), parental interference (3), or not completing the study due to fussiness (5).

Materials

Ping-pong Balls. A total of 292 (146 green, 73 red, and 73 yellow) Ping-pong balls were used. All of the green balls had six white Velcro strips glued to them.

Boxes and Containers. Boxes and containers were the same as those in Experiments 1 and 2 except that there were three colors of Ping-pong balls. In the Physical Constraint condition, the green balls were glued to the boxes and containers and in the No Physical Constraint condition they were not. The small white box used in the Free Play Period and the Plexiglas container used in Demonstration Phase 1 contained two green, two red, and two yellow balls. The small box used in Demonstration Phase 2 contained six green, six red, and six yellow balls.

The large box used in the Familiarization Trials and Test Trials had a total of 200 balls with a ratio of 5 green (with Velcro): 4 red: 1 yellow balls on one side and 5 green (with Velcro): 4 yellow: 1 red balls on the other side.
Apparatus

The Apparatus was the same as Experiment 1.

Procedure

The procedure was the same as in Experiment 1 with any differences resulting from the fact that there were three colors of balls.

Physical Constraint Condition

Free Play Period. At the beginning of the experiment, the experimenter showed infants the small white box with two green, two red and two yellow balls for about 1 minute. She allowed the infants to play with the balls and encouraged them to pick the balls up so that the infants could see that the balls were discrete objects and that the green balls were stuck to the box.

Demonstration Phase 1. The experimenter brought out the long Plexiglas container with two green, two red, and two yellow balls in a random order. She picked up (in counterbalanced order) the red balls one at a time while saying, “Look at this one! See this?” She then did the same with the yellow balls. Then, she grasped the green balls one at a time and said, “Look at this one. It’s stuck!” This phase lasted for approximately 30 seconds, after which experimenter cleared the container from the stage.

Familiarization trials. Four familiarization trials were conducted in exactly the same manner as in Experiment 1. These trials lasted approximately 4 minutes.

Demonstration Phase 2. This phase was identical to Experiment 1 except that the experimenter picked up 1 red ball and 1 yellow ball and grasped one green ball. This phase lasted approximately 45 seconds.
Test trials. Six test trials were conducted in exactly the same manner as in Experiment 1. Infants either saw the box with a ratio of 5 green: 4 red: 1 yellow balls or a ratio of 5 green: 4 yellow: 1 red balls for all six test trials. On alternating trials, the experimenter pulled out a sample of four red balls and one yellow ball from the box or a sample of four yellow balls and one red ball from the box. This phase lasted approximately 7 minutes.

Design

In each condition, 10 infants saw the box with a ratio of 5 green: 4 red: 1 yellow balls during test trials and 10 infants saw the box with a ratio of 5 green: 4 yellow: 1 red balls. The order of the familiarization trials (5 green: 4 red: 1 yellow vs. 5 green: 4 yellow: 1 red) and the order of the samples on the test trials (four red and one yellow or four yellow and one red first) were counterbalanced across infants. The order during the Demonstration Phases (green, red or yellow first) was randomized across infants.

No Physical Constraint Condition

The No Physical Constraint condition was exactly the same as the Physical Constraint condition except that all of the balls were movable.

Design

The design was the same as the Physical Constraint Condition.

Predictions

Physical Constraint Condition We predicted that infants would detect the physical constraint (green balls are immovable) and therefore they should look longer at a sample that is less probable than one that is more probable given the population of Ping-pong balls in the box. For example, for the 10 infants who saw the population with a ratio of 5 green: 4 red: 1 yellow balls on test trials, a sample of four yellow and one red balls is less probable than a sample of
four red and one yellow balls.⁴ For the 10 infants who saw the population with a ratio of 5 green: 4 yellow: 1 red balls, the opposite is true.

_No Physical Constraint Condition_ Because the green balls were movable, the probability of obtaining either sample (four red and one yellow or four yellow and one red) from a box containing 50% green balls is extremely low. However, it is technically more likely to obtain a sample of four red and one yellow balls than a sample of four yellow and one red balls from a box with a ratio of 5 green: 4 red: 1 yellow balls (and vice versa for the population with a ratio of 5 green: 4 yellow: 1 red balls).⁵ Due to the fact that these probabilities are both so low (< 2%), we predicted that infants will find both outcomes unexpected and will look about equally at both outcomes.

_Coding_

For the 10 infants who saw the population with a ratio of 5 green: 4 red: 1 yellow during test trials, we coded the sample of four red and one yellow balls as the majority-color sample and

---

⁴ Infants could carry out either of the following computations in order to succeed at this task:
1) The probability of obtaining a sample of a particular order of 4 yellow and 1 red balls from a box containing 50 green, 40 yellow and 10 red balls = 40/50 * 39/49 * 38/48 * 37/47 * 10/46 = 0.08627. The probability of obtaining a sample of a particular order of 4 red and 1 yellow balls from this box = 10/50 * 9/49 * 8/48 * 7/47 * 40/46 = 0.00079.
2) The probability of obtaining a sample of 4 yellow and 1 red balls, irrespective of order = 0.08627 * 5 = 0.43134. The probability of obtaining a sample of 4 red and 1 yellow balls from this box, irrespective of order = 0.00079 * 5 = 0.00396.

The second set of computations seems a more plausible strategy than the first, although it is possible to succeed at this task via some alternative strategy.
The more probable outcome is 109 times more likely than the less probable outcome for both sets of computations.

⁵ 1) The probability of obtaining a sample of a particular order of 4 yellow and 1 red balls from a box with 50 green, 40 yellow and 10 red balls = 40/100 * 39/99 * 38/98 * 37/97 * 10/96 = 0.00243. The probability of obtaining a sample of a particular order of 4 red and 1 yellow balls from this box = 10/100 * 9/99 * 8/98 * 7/97 * 40/96 = 0.00002.
2) The probability of obtaining a sample of 4 yellow and 1 red balls, irrespective of order = 0.0024 * 5 = 0.01214. The probability of obtaining a sample of 4 red and 1 yellow balls from this box, irrespective of order = 0.00002 * 5 = 0.00011.

The more probable outcome is 109 times more likely than the less probable outcome for both sets of computations.

However, all four probabilities have less than 2% chance of occurring.
the sample of four yellow and one red balls as the minority-color sample. For the 10 infants who saw the population with a ratio of 5 green: 4 yellow: 1 red during test trials, trials were coded in the opposite way. This coding was used for both conditions (Physical Constraint Condition and No Physical Constraint Condition).

Results

Preliminary analyses found no effects of gender, content of the box on test trials (5 green: 4 red: 1 yellow vs. 5 green: 4 yellow: 1 red), order of samples on test trials (4 red and 1 yellow first or 4 yellow and 1 red first) or order of the familiarization trials. Subsequent analyses collapsed over these variables. A second observer who was unaware of the order of the trials coded a random sample of 10 infants from each condition. Inter-observer reliability averaged 97% and 95% for the two conditions, respectively.

An ANOVA found no differences between average looking times on the familiarization trials across conditions, $F(3,111) = 0.047, p > .5$; $M_{\text{physical constraint}} = 8.58s$, $SD_{\text{physical constraint}} = 5.01s$. $M_{\text{no physical constraint}} = 10.33s$; $SD_{\text{no physical constraint}} = 5.40s$.

An Omnibus ANOVA examined the effects of test trial order (majority-color sample outcome first or minority-color sample outcome first), trial pair (1, 2, and 3), outcome (majority-color sample vs. minority-color sample) and condition (Physical Constraint vs. No Physical Constraint). There was a main effect of trial pair, $F(2,76) = 5.32$, $p = .007$, average looking decreased over time. There was also an interaction between outcome and condition, $F(1,38) = 6.32$, $p = .016$, effect size ($\eta^2$) = .14 (see Fig. 5). There were no other main effects or interactions.

Given the interaction between Outcome and Condition and our specific a priori looking time predictions, ANOVAs were performed to analyze each condition separately.
Physical Constraint Condition

An ANOVA examined the effects of test trial order (majority-color sample outcome first or minority-color sample outcome first), trial pair (1, 2, and 3), and outcome (majority-color sample outcome vs. minority-color sample outcome). There was a main effect of outcome, $F(1,19) = 14.66$, $p = .001$, effect size ($\eta^2_p$) = .44. Infants looked reliably longer at the minority-color sample outcomes ($M = 7.45s$, $SD = 3.70s$) than the majority-color sample outcomes ($M = 5.53s$, $SD = 3.85s$). Sixteen of twenty infants looked longer at the unexpected outcomes, Wilcoxon signed-ranks test: $z = -2.91$, $p = .004$. There were no other main effects or interactions.

No Physical Constraint Condition

An ANOVA examined the effects of test trial order (majority-color sample outcome first or minority-color sample outcome first), trial pair (1, 2, and 3), and outcome (majority-color sample vs. minority-color sample). There was a main effect of trial pair, $F(2,38) = 4.37$, $p = .020$; average looking decreased over time. There was no main effect of outcome, $F (1,19) = .21$, $p = .65$, effect size ($\eta^2_p$) = .01. Infants looked about equally at the minority-color sample outcomes ($M = 8.07s$, $SD = 4.03$) and the majority-color sample outcomes ($M = 7.72$, $SD = 4.00$). Nine of twenty infants looked longer at the minority-color sample outcomes, Wilcoxon signed-ranks test: $z = -.30$, $p = .77$. There were no other main effects or interactions.

To test the hypothesis that both outcomes in the No Physical Constraint Condition were unexpected, we conducted four post-hoc t-tests comparing the mean looking times obtained from the Physical Constraint and No Physical Constraint Conditions: First, we compared the majority-color sample outcome in the Physical Constraint condition ($M = 5.53s$) with the majority-color sample outcome in the No Physical Constraint condition ($M = 7.72s$), $t(38) = -2.10$, $p = .02$. Second, we compared the majority-color sample outcome in the Physical Constraint condition
(M = 5.53s) with the minority-color sample outcome in the No Physical Constraint condition, (M = 8.07s), t(38) = -2.12, p = .02. Third, we compared the minority-color sample outcome in the Physical Constraint condition (M = 7.45s) with the majority-color sample outcome in the No Physical Constraint condition (M = 7.72s), t(38) = -0.50, p = .62. Finally, we compared the minority-color sample outcome in the Physical Constraint condition (M = 7.45s) with the minority-color sample outcome in the No Physical Constraint condition (M = 8.07s), t(38) = -0.24, p = .81. These results indicate that the mean looking times for the minority-color sample outcome in the Physical Constraint condition and the majority-color sample and minority-color sample outcomes in the No Physical Constraint condition were not reliably different from one another. The majority-color sample outcome for the Physical Constraint condition was different from each of the other three means.

**Discussion**

In the Physical Constraint condition, infants were given evidence that green balls were physically constrained inside the box. They looked longer at a sample of, for example, four yellow balls and one red ball than a sample of four red balls and one yellow ball from a box with a ratio of 5 stuck green balls to 4 red balls to 1 white ball. This suggests that infants kept in mind that green balls were physically constrained and could thus be disregarded from probability calculations. Excluding the green balls allowed infants to correctly make inferences regarding the likelihood of a mostly yellow or mostly red sample from a box with either a greater proportion of red or yellow balls.

---

6 According to the Bonferroni technique, these p values are marginally significant (Bonferroni requires an alpha level of .0125 for these tests).
In contrast, infants in the No Physical Constraint condition were not given evidence that green balls were stuck inside the box. They looked about equally at both samples. This suggests that infants did not find either of the outcomes less probable in this condition.

Furthermore, the analyses indicating that there were no reliable differences in looking times at both outcomes in the No Physical Constraint condition and the minority-color sample outcome in the Physical Constraint condition suggest that infants may have found all three of these outcomes improbable. It is also possible, however, that infants in the No Physical Constraint condition were simply unable to reason about the probability of three sets of balls without the benefit of excluding one of the sets. This may have caused them to look equally at the two samples in this condition. Teasing apart whether infants found both of these outcomes unexpected or whether they were simply unable to reason about three sets of balls is an empirical question regarding the computational limits of this mechanism. However, the fact that infants’ looking times on these trials were relatively long and were similar in length to the looking time on the minority-color sample trials in the Physical Constraint condition makes the alternative interpretation less plausible. Regardless, findings from the No Physical Constraint condition control for the possibility that infants in the Physical Constraint condition did not detect the physical constraint and were instead computing which of two very unlikely probabilities was least likely. Taken together, the results of these two conditions suggest that infants in the Physical Constraint condition were sensitive to, and were able to apply a physical constraint while deploying the statistical inference mechanism.

These results provide the first evidence that infants can succeed in a situation in which they must take into account both domain-specific physical knowledge and statistical information in the same task. The results from this experiment go beyond previous studies in that infants
were asked to apply a physical constraint and then make a more subtle discrimination between probable and improbable events as opposed to possible and impossible events.

5. General Discussion

Results from three experiments suggest that 11-month-old infants can integrate a physical constraint of immobility into a statistical inference mechanism. In Experiment 1, we found that infants looked longer when an experimenter removed from a box a sample of five balls that should have been physically constrained, regardless of whether the probability of obtaining constrained balls was high or low based on the proportions in the box. In Experiment 2, we replicated earlier findings suggesting that infants can make inferences from samples to populations. In Experiment 3, infants were able to exclude one set of balls due to the physical constraint and then compute probabilities over two remaining sets of balls. They looked longer at a sample of four yellow and one red balls than a sample of four red and one yellow balls being drawn from a box with a ratio of 5 green to 4 red to 1 yellow balls, if the green balls were physically constrained. In contrast, when infants were not shown that green balls with Velcro were constrained, they looked about equally at both of these samples. These results demonstrate that infants can override probabilistic knowledge in favor of domain-specific knowledge (Experiments 1 and 2), and furthermore, we provide the first evidence that infants can integrate domain-specific knowledge with probabilistic information under the appropriate conditions (Experiment 3).

Along with the experiments conducted by Teglas et al. (2007), the present experiments suggest that 11- to 12.5-month-old infants can take into account two principles of physical reasoning – immobility and solidity – while deploying a statistical inference mechanism. At 11 months, infants are also able to reason about psychological constraints under similar conditions.
(Xu & Denison, 2009). Previous studies have only asked infants to choose between using their physical or psychological knowledge versus using the probabilistic information provided by the experimental context. In Experiment 3, we provide the first evidence for a bona fide case of integrating domain-specific knowledge in statistical inference: the infants had to first apply their physical knowledge about immobility to exclude one set of balls, then they had to compute probabilities on the remaining two sets of balls.

The results of the current experiments also shed light on the nature of the statistical inference mechanism. The findings of Xu and Garcia (2008) were consistent with at least two interpretations: infants may have estimated the proportions of red and white balls in the box and made some predictions about the samples based on probability. Alternatively, infants may have expected the samples to roughly match the proportions of red and white balls inside the box – a version of the representativeness heuristic (e.g., Tversky & Kahneman, 1974). The Physical Constraint Condition in Experiment 3 provides suggestive evidence that the matching hypothesis may not be correct: in this case, neither sample – four yellow and one red balls or four red and one yellow balls – resembled the composition of the box – half green balls, and some red and some yellow balls. Future studies will investigate this question more directly. For example, one might create a situation where the population consists of 80% white balls and 20% red balls, and the majority of the white balls, but not all, are stuck inside the box such that the proportions of white and red balls are reversed once the cohesion violation is taken into account. This manipulation will ensure that perceptual matching between the sample and the population would result in a gross miscalculation of probabilities.

Several questions remain open from the current experiments. First, what is the developmental trajectory in terms of infants’ abilities to integrate domain-specific knowledge in
statistical inference? The probabilistic inference mechanism that we explored is present at 8 months of age (Xu & Garcia, 2008) and infants’ reason correctly about core object principles very early in development. We chose to test 11-month-old infants in the reported experiments as a starting point because the current tasks appear much more difficult than the experiments completed by 8-month-old infants in Xu and Garcia (2008) and the experiments completed by young infants in most of the physical reasoning experiments (Spelke, 1994). In addition, the current experiments are meant to be compatible with findings suggesting that 11-month-old infants can integrate a psychological constraint into this mechanism (Xu & Denison, 2009) and the findings suggesting that 11- to 12.5-month-old infants can reason about a solidity constraint in another statistical inference mechanism (Teglas et al, 2007). An interesting future direction of this work will investigate whether younger infants are capable of reasoning about both physical and psychological constraints in the context of this statistical inference task. Given young infants’ robust knowledge about physical reasoning principles (e.g., Spelke et al. 1992), it seems likely that younger infants will be able to apply physical knowledge in their statistical computations.

Second, what is the computational limit of this statistical inference mechanism? In the No Physical Constraint Condition of Experiment 3, infants looked about equally at the two samples even though one was more probable than the other. Two interpretations are possible: infants may have found both samples improbable since half of the balls in the box were green and neither sample contained a single green ball. Alternatively, infants may have found it taxing to reason about three sets of balls. Future studies may address this issue directly by asking infants to estimate the probabilities of the samples given a population of 50% green balls, 40% red balls, and 10% yellow balls, e.g., two green, two red, and one yellow balls vs. four red and one yellow
balls. Some visual attention studies suggest that adults can keep track of a maximum of three sets of objects in large arrays (e.g., Alvarez & Cavanagh, 2004; Halberda, Sires, & Feigenson, 2006). For infants, the set size limit is an open empirical question.

In sum, our findings suggest that when substantive domain-specific knowledge is in conflict with probabilistic information, infants are able to override probabilistic information and reason according to the domain-specific constraint. Furthermore, infants are able to use the domain-specific constraints to decide how to compute probabilities appropriately. The findings of these and related experiments demonstrate that the study of statistical learning mechanisms and domain-specific knowledge should go hand in hand.
References


Acknowledgment

Thanks to the infants and their parents for participating in our studies. We would also like to thank Krista Byers-Heinlein, Kathryn Dewar, Mijke Rhemtulla and Henny Yeung for helpful comments on the design of the experiments and Sarah Heller, Maria Ho and Vanessa Waechtler for their assistance in running the experiments. Thanks also to Susan Birch and Janet Werker for helpful comments on an earlier version of this manuscript. This research was supported by a grant from the Natural Science and Engineering Research Council (NSERC), Canada, to F.Xu.
Figure Captions

*Figure 1.* Schematic representation of the procedure of Experiment 1.

*Figure 2.* Average looking time at the mobile objects sample and immobile objects sample outcomes in Experiment 1.

*Figure 3.* Average looking time at the majority-color sample and minority-color sample outcomes in Experiment 1 for the Consistent Group and the Inconsistent Group.

*Figure 4.* Average looking time at the majority-color sample and minority-color sample outcomes in Experiment 2.

*Figure 5.* Average looking time at the majority-color sample and the minority-color sample outcomes in the Consistent condition and the “majority-color sample” and the “minority-color sample” in the Inconsistent condition in Experiment 3.
Figure 1.

Mobile Objects Sample Outcome

1. 

2. 

3. 

4. 

5. 

Immobile Objects Sample Outcome

1. 

2. 

3. 

4. 

5.
Figure 2.
Figure 3.
Figure 4.
Figure 5.

Outcome Based on Constraint and Box Proportions

Constraint Condition  No Constraint Condition

Time (seconds)