"Why is Toma late to school again?"

Preschoolers identify the most informative questions

Azzurra Ruggeri

Zi L. Sim

Fei Xu

University of California, Berkeley

(In press, Developmental Psychology)

Abstract

The current study investigates whether preschoolers are able to successfully identify the most effective among given questions, adapting their reliance on different types of questions (*constraint-seeking* vs. *hypothesis-scanning*) based on the quantitative measure of *expected information gain*. Children were presented with storybooks describing the reasons why a fictional character, Toma, was late to school over several days. In three experiments with five-year-old children, we manipulated the frequency and likelihoods of the reasons presented. Children were asked to identify which of two given questions would be more effective in finding out why Toma was late to school again. In a fourth experiment, we investigated whether preschoolers are *adaptive* learners, that is, whether they can identify the most effective question iteratively, and we extended our investigation to younger preschoolers (3- and 4-year-olds). We find that children assessed the effectiveness of different types of questions based on the hypothesis space currently under consideration, and this adaptation may be guided by expected information gain. Overall, our results suggest that over the preschool years, children begin to develop the computational foundations that support successful question-asking strategies.

Keywords: expected information gain, question-asking, cognitive development.

"Why is Toma late to school again?" Preschoolers identify the most informative questions

Asking questions is a powerful learning tool. Children ask questions about a variety of topics numerous times a day. In a sample analyzed by Chouinard (2007), 2- to 5-year-olds asked an average of 107 questions per hour while engaged in conversation with adults. Their inquiring behavior is purposeful, intended to fill a knowledge gap or resolve some inconsistency, to seek explanations, and, more generally, to test and extend their developing understanding of the world (Carey, 1985; Chouinard, 2007; Frazier, Gelman, & Wellman, 2009; Gopnik & Wellman, 1994; Harris, 2012; Piaget, 1954; Wellman, 2011).

Research to date has shown that young children ask domain-appropriate questions (Callanan & Oakes, 1992; Greif, Kemler Nelson, Keil, & Guterrez, 2006; Hickling & Wellman, 2001), have reasonable expectations about which responses count as answers to their questions (Frazier et al., 2009), and can use the answers they receive to solve problems (Chouinard, 2007; Legare, Mills, Souza, Plummer, & Yasskin, 2013). We also know that children direct their questions toward more reliable informants (Birch, Vauthier, & Bloom, 2008; Corriveau & Harris, 2009; Koenig & Harris, 2005; Mills, Legare, Bills, & Mejias, 2010; Mills, Legare, Grant, & Landrum, 2011) and they privilege more informative cues (Nelson, Divjak, Martignon, Gudmundsdottir, & Meder, 2013).

Previous studies have examined the development of children's ability to ask questions by using variations of the Twenty Questions game, in which children have to identify a target object or category of objects within a given set (e.g., "What kind of objects can be found on Planet Apres?) by asking as few yes-no questions as possible, e.g., “Are animals found on Planet Apres?” (see Mosher & Hornsby, 1966; Nelson et al., 2014; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015; Ruggeri, Lombrozo, Griffiths & Xu, 2016).

In most of these studies, researchers measure children’s question-asking ability by analyzing their usage of *constraint-seeking* vs. *hypothesis-scanning* questions. Constraint-seeking questions target a category of objects or a feature shared by multiple objects, such as “Can animals be found on Planet Apres?” They stand in contrast to hypothesis-scanning questions, which target a single object within the given set, such as “Can this dog be found on Planet Apres?” Constraint-seeking questions are usually considered to be more effective than hypothesis-scanning questions because they are able to rule out multiple hypotheses (objects, categories of objects or reasons) at each step of the search process (Mosher & Hornsby, 1966). Legare and colleagues (2013) showed that preschoolers as young as 4 are able to generate a majority of *effective* constraint-seeking questions, as opposed to redundant or ineffective questions (i.e., questions that do not discriminate among different hypotheses; see Legare et al., 2013). Their study design does not allow for a direct comparison between children’s usage of constraint-seeking and hypothesis-scanning questions, because in their procedure children were only allowed to ask one hypothesis-scanning question (i.e., “Is it the card with the small spotted red bird?”). However, previous research provided empirical evidence that preschoolers’ question generation is strongly characterized by a hypothesis-scanning approach. Indeed, Herwig (1982) found that *all* of the questions generated by preschoolers in a 20-questions task are hypothesis-scanning questions. By age 7, children still predominantly use hypothesis-scanning questions (Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015). However, children transition to using more constraint-seeking questions over the course of development, until constraint-seeking becomes the dominant strategy in adulthood (Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015).

Generating constraint-seeking questions from scratch depends on children's verbal knowledge, categorization skills, and previous experience. For example, one needs to identify features that can be used to group hypotheses into different categories, categorize objects correctly according to those features, and label those categories. Indeed, the developmental change in the effectiveness of children's questions has been explained by an increasing ability to generate object-general features that can be used to cluster similar objects into categories (e.g., quadrupeds vs. non-quadrupeds, see Ruggeri & Feufel, 2015). This leaves open the possibility that if children are not required to *generate* these high-level object features themselves, the ability to *select* the most informative within a set of given questions may be observed earlier than the ability to ask effective questions from scratch. Indeed, previous work shows that 5- to 7-year-old children are more efficient when selecting among given questions than when generating questions. When presented with a forced choice between a constraint-seeking question and a hypothesis-scanning question, 46% of the questions selected by five-year-olds and about 60% of those selected by first and second graders were constraint-seeking questions, as compared to 0% (five-year-olds) and less that 20% (first and second graders) of their *generated* questions (Herwig, 1982; see also Ruggeri & Feufel, 2015).

Although constraint-seeking questions are traditionally considered to be more effective than hypothesis-scanning questions, they are not always the most effective. Indeed, the informativeness of each question type varies depending on the problem being considered, e.g., the number of hypotheses available and their likelihoods (Ruggeri & Lombrozo, 2015; see also Todd, Gigerenzer, & the ABC Research Group, 2012). For example, with only three equally likely candidate hypotheses, hypothesis-scanning questions are just as informative as constraint-seeking questions. Moreover, when the alternative hypotheses considered are not all equally likely, a hypothesis-scanning question that targets a single high-probability hypothesis (e.g., one that has a 70% probability of being correct) can be more informative than a constraint-seeking question that targets several hypotheses with a small summed probability (e.g., 30%). Furthermore, not all constraint-seeking questions are equally effective: For example, a constraint-seeking question that partitions the hypothesis space evenly is on average more informative than a constraint-seeking question that partitions the same space unevenly. Given these considerations, studies with adults have often used more formal quantitative measures such as *expected information gain* to capture the effectiveness of different information search strategies (Chin, Payne, Fu, Morrow, & Stine-Morrow, 2015; Nelson, McKenzie, Cottrell, & Sejnowski, 2010; Oaksford & Chater, 1994; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003).

The current study has three main goals. First, we focus on preschoolers' *judgments* of the effectiveness of given questions, disentangling the process involved in selecting the most informative questions from the processes involved in generating effective questions from scratch. We test the hypothesis that children's ability to select more informative questions may emerge earlier in development than their ability to generate these questions. Second, we consider how the qualitative distinction between constraint-seeking and hypothesis-scanning questions maps onto the more formal distinction between more and less informative questions using *expected information gain*. Expected information gain measures how much a question reduces the uncertainty in the hypothesis space considered (see section below). Although it is unlikely that learners actually compute expected information gain as it is done in computational models, this formal measure gives us a *computational level mechanism* for comparing the effectiveness of different questions. In the developmental literature, to our knowledge, no study has investigated whether and how a formal measure such as expected information gain may capture preschoolers' question-asking or question-selection behavior (for 7- to 10-year-old children see Nelson et al., 2014; Ruggeri et al., 2016). Third, we ask if preschoolers are *adaptive* learners—whether they are able to implement effective information-search strategies iteratively based on feedback (see Ruggeri & Lombrozo, 2015).

Having such a nuanced understanding of a question's informativeness, which goes beyond a simple consideration of its type, builds upon a more basic capacity to understand and reason with frequencies and probabilities. Recent research suggests that infants are already capable of rudimentary probabilistic reasoning (Denison, Reed, & Xu, 2013; Denison & Xu, 2010a; 2010b; 2014; Teglas, Girotto, Gonzalez, & Bonatti, 2007; Teglas et al., 2011; Xu & Denison, 2009; Xu & Garcia, 2008). Moreover, a growing body of research suggests that infants and preschoolers are already able to *use* probabilistic information to form judgments, to make predictions and generalizations, and to guide their information search (Denison & Xu, 2014; Gweon, Tenenbaum, and Schulz, 2010; Kushnir & Gopnik, 2005). Children are able to integrate prior probabilities with feedback and subsequent evidence (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Girotto & Gonzalez, 2008; Gonzalez & Girotto, 2011) and make inferences that are consistent with the general principles of Bayesian inference (e.g., Eaves & Shafto, 2012; Ruggeri et al., 2016; Schulz, Bonawitz, & Griffiths, 2007).

The present experiments investigate whether preschoolers are sensitive to the statistical structure of a given causal scenario, adapting their reliance on different question types (constraint-seeking vs. hypothesis-scanning) depending on their informativeness as measured by expected information gain. To do so, we use a causal version of the 20-questions game (Ruggeri & Lombrozo, 2015), in which participants are asked to identify, among a given set of hypotheses, the reason why something happened (i.e., “Why was the monster Toma late for school?”). Whereas most 20-questions game used in the literature consider an hypothesis space with uniform prior (i.e., all hypotheses are equally likely to be correct), this version allows us to easily manipulate the likelihood of the available hypotheses and therefore provide different prior distributions over the given hypothesis space (see Nelson, Meder, & Jones, 2016, for in-depth discussion of 20-question games with unequal priors). For example, we present children with the reasons why Toma was late for school over several days, and manipulate the frequency of the given reasons, so that some occurred more often than others (e.g., “On three days Toma was late because he woke up late”).

Formal Framework: Expected Information Gain

Although several possible measures can be used to compute how informative different questions are (e.g., probability gain, impact, expected savings, path length; see Nelson, 2005), we followed previous research that has used the 20-questions task (Eimas, 1970; Nelson et al., 2014; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015; Ruggeri, Lombrozo, Griffiths, & Xu, in press) and measured the informativeness of questions in terms of their expected stepwise information gain. Expected stepwise information gain (see Chin, Payne, Fu, Morrow, & Stine-Morrow, 2015; Nelson, McKenzie, Cottrell, & Sejnowski, 2010; Oaksford & Chater, 1994; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003) measures the reduction of entropy (Shannon entropy; Shannon, 1948) –that is, the uncertainty as to which hypothesis is correct–upon asking a certain question (see Lindley, 1956). Within this framework, the best questions are the ones that maximize the reduction of entropy, allowing the learner to move from a state of uncertainty (e.g., "Why was the boy late to school?") closer to a state of certainty (e.g., "The boy was late to school because he woke up late.") with the fewest number of questions. It is important to note that, in our studies, alternative measures to compute the informativeness of a question (such as probability gain or path length) would have led to identical predictions.

Formally, expected information gain of each question can be computed by subtracting the *expected posterior entropy* from the *prior entropy*:

$$IG = H_{prior} - H_{posterior} \qquad \text{Eq. (A.1)}$$

The entropy H embodies the uncertainty about which of the candidate hypotheses is true. Its computation is based on the probabilities (p) associated with each of the candidate hypotheses (h). The prior entropy H_{prior} defines the status of uncertainty preceding every action:

$$H_{prior} = -\sum_h p(h) \log_2 p(h) \qquad \text{Eq. (A.2)}$$

The predictive posterior entropy $H_{\text{posterior}}$ refers to the predicted uncertainty after the question is asked and the answer is received. The predicted posterior entropy is measured as the sum of the entropies corresponding to each possible future scenario weighted according to the probability of that scenario. Because in our task there are two possible answers to each question (yes/no), $H_{\text{posterior}}$ is computed as the sum of:

$$H_{\text{posterior}} = p(x_{\text{yes}}|X)H(x_{\text{yes}}) + p(x_{\text{no}}|X)H(x_{\text{no}}) \qquad \text{Eq. (A.3)}$$

To our knowledge, expected information gain has never been used as a formal measure to capture preschoolers' learning behavior. An example of how expected information gain was calculated in our studies can be found in Appendix A.

Overview of the studies

In four experiments, preschoolers are given a simple causal inference task about why a monster, Toma, was late to school. In the first three experiments (Experiments 1A-1C), we test the hypothesis that 5-year-olds are able to select the most effective question across a variety of scenarios. In particular, we hypothesize that children rely on different types of questions (constraint-seeking vs. hypothesis-scanning) based on their expected information gain in a scenario, rather than based on the probability of positive feedback (Experiment 1B) or the salience associated with the single most frequent hypothesis (Experiment 1C). In Experiment 2, we replicate and extend our investigation developmentally to include younger preschoolers (3- and 4-year-olds). Additionally, we examine the possibility that preschoolers are *adaptive* learners, revising their judgments of effectiveness of different question types iteratively by taking into account how the hypothesis space changes due to feedback.

A sample size of 25 – 30 participants was targeted in our experiments based on prior research (see Ruggeri & Lombrozo, 2015; Ruggeri & Feufel, 2015).

Experiment 1A

Method

Participants. Participants were 60 5-year-olds (36 female, M_{age} = 62.4 months; *SD* = 7.9 months) recruited from local children's museums and schools. Five additional children were excluded from the analyses for failing to respond to the test question (N = 2), or due to parental interference (N = 3). Participants were randomly assigned to one of two conditions: Uniform or Skewed (see below).

Design and procedure. Participants were presented with a storybook, displayed on a computer screen. The story introduced Toma, a monster from Planet Apres, who is often late to school, and illustrated the reasons why Toma was late to school over several days. Each day was represented on a different page of the storybook (e.g., "On Day 6, Toma was late because he was watching TV"), and a clipart was used to illustrate the reason why Toma was late on that day (e.g., a television; see Figure 1).

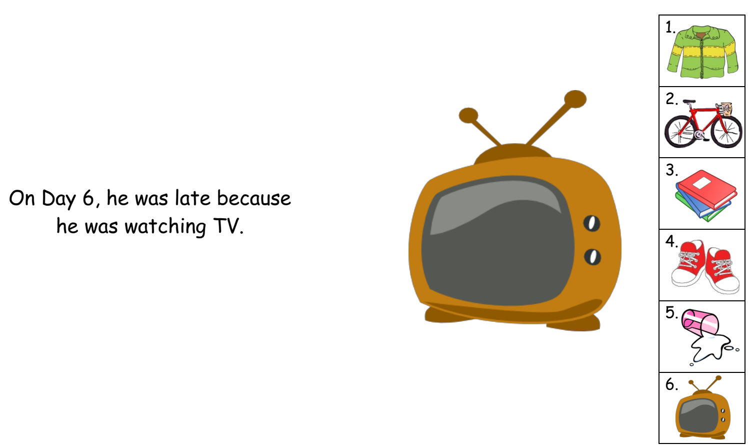

Figure 1. Example page of a storybook from the first series of experiments presenting the reasons why Toma was late to school over several days. Each day was represented on a different page of the storybook, and a clipart (e.g., a television) was used to illustrate the reason why Toma was late on that day.

Children were randomly assigned to one of two experimental conditions: Uniform or Skewed. In the story presented to participants in the *Uniform* condition, Toma had been late to school on six days, each day for a different reason. Therefore a total of six different hypotheses were included in this condition (see Figure 2 and Table B1 of the Appendix).

In the story presented to participants in the *Skewed* condition, Toma had been late to school on eight days. On five out of eight days, Toma was late to school because he woke up late, and on the other three days he was late for three different reasons. Therefore a total of four different hypotheses were included in this condition (see Figure 2 and Table B1 of the Appendix). To ensure that the information gain of the hypothesis-scanning question presented at test (see below) was higher than that of the constraint-seeking question in the Skewed condition, it was necessary to present more instances in this condition (8 days) than in the Uniform condition (6 day), but overall a smaller number of distinct hypotheses.

After being presented with all reasons why Toma had been late for school on the previous days, children were told that Toma was late to school again today, and that his monster friends, Dax and Wug, wanted to find out why. Toma proposes a game: "I won't tell you; you have to find out. You can ask me questions to find out. The first who finds out wins!" The children were then presented with the questions that Dax and Wug asked to find out why Toma was late to school again (see Figure 2). One of the monsters (Dax or Wug, counterbalanced across participants) asked a constraint-seeking question targeting multiple hypotheses for why Toma was late to school (e.g., Dax said, "Toma, were you late because you could not find something?", which targets the following three hypotheses: He could not find his jacket, or he could not find his books, or he could not find his shoes). The other monster asked a hypothesis-scanning question targeting a single hypothesis (e.g., Wug said, "Toma, were you late because your bike was broken?"). The hypotheses targeted by each question were also illustrated in two thought bubbles containing the corresponding cliparts previously used to represent the various reasons for Toma's tardiness over the last several days. At the bottom of the same page, children also saw a graphical summary of the reasons why Toma had been late in the past days, one clipart for each day that Toma had been late, so that the reasons that occurred on more days were represented multiple times (Figure 2).

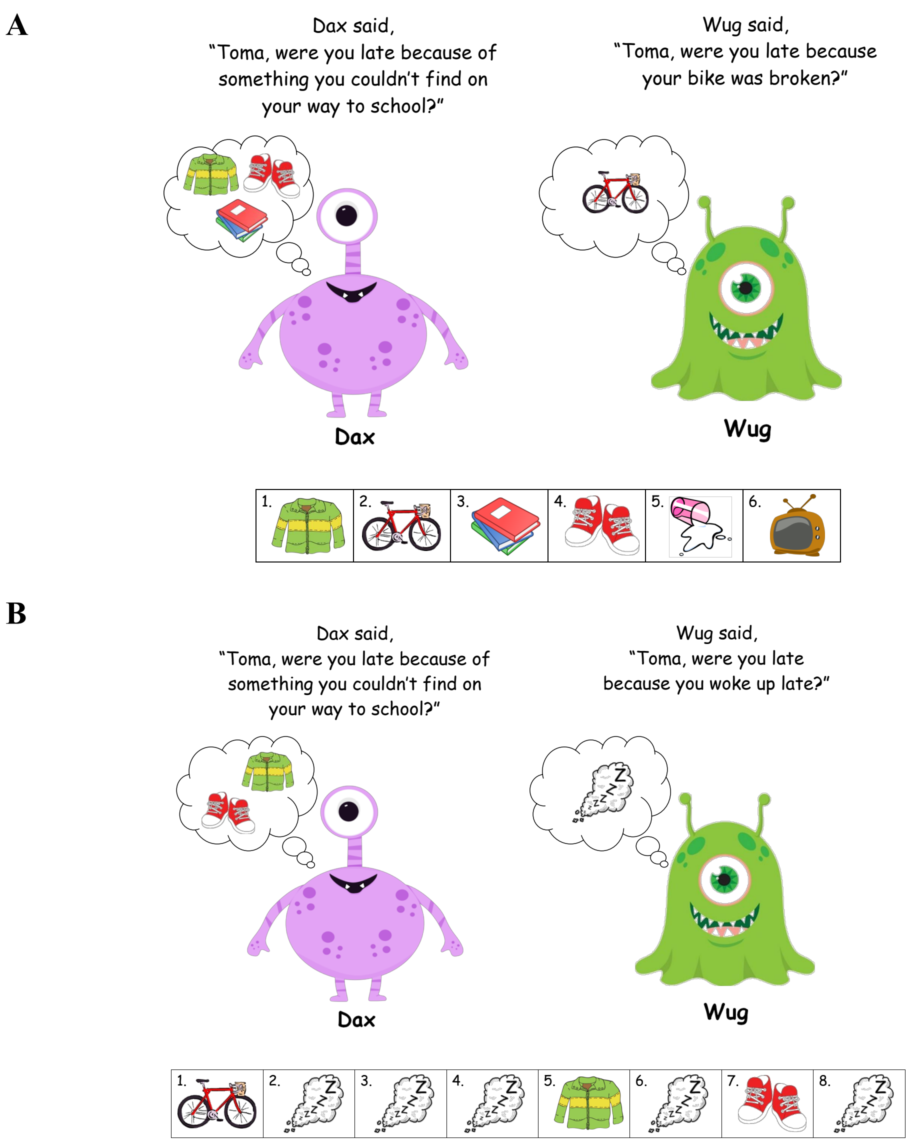

Figure 2. Displays presented at test in the Uniform (Figure 2A) and Skewed (Figure 2B) conditions of Experiment 1A. Children were asked to select which monster would find out first why Toma was late to school. One of the monsters asked a constraint-seeking question targeting multiple hypotheses, whereas the other monster asked a hypothesis-scanning question targeting a single hypothesis. At the bottom of the page, children were reminded of the reasons why Toma was late on previous days, using the corresponding cliparts.

The informativeness of the constraint-seeking and hypothesis-scanning questions depended on the conditions children were assigned to.

Uniform condition. In the Uniform condition, the constraint-seeking question ("Toma, were you late because you could not find something?") targeted three hypotheses that occurred on three of the six days: Toma was late because he could not find his jacket, or he could not find his books, or he could not find his shoes. The information gain for the constraint-seeking question was exactly IG = 1 (see Appendix A and Figure 2).

The hypothesis-scanning question ("Toma, were you late because your bike was broken?") targeted a single hypothesis that occurred on one of the six days. The information gain for the hypothesis-scanning question was IG = .66 (see Appendix A and Figure 2). Therefore, in the Uniform condition, the constraint-seeking question was more informative.

Skewed condition. In the Skewed condition, the constraint-seeking question ("Toma, were you late because of something you could not find?") targeted two hypotheses, each occurring on a different day: Toma was late because he could not find his jacket or he could not find his shoes. The information gain for the constraint-seeking question was IG = .81 (see Appendix).

The hypothesis-scanning question ("Toma, were you late because you woke up late?") targeted the single most frequent hypothesis that occurred on five of the eight days. The information gain for the hypothesis-scanning question was IG = .94 (see Appendix A). Therefore, in the Skewed condition, the hypothesis-scanning question was more informative.

Children were then asked to indicate which of the two friends, Dax or Wug, would win the game, that is, to find out first why Toma was late to school again today. We accepted both verbal responses (e.g., the monster's name or color) and points towards either monster.

Results and discussion

Preliminary analyses found no effects of gender or the order in which the two monster friends (i.e., Dax and Wug) were presented. Subsequent analyses were collapsed over these variables.

In the Uniform condition, 70% of the children selected the monster asking the constraint-seeking question as the winner (i.e., the one to find out first why Toma was late to school again), exact binomial *p* (two-tailed) = .042. In the Skewed condition, 73% of the children selected the monster asking the hypothesis-scanning question as the winner, exact binomial *p* (two-tailed) = .016. A chi-square test confirmed the difference between these two distributions, $\chi^2(2, N = 60) = 11.28$, $p < .001$. In both conditions, the majority of children chose the question that had a higher expected information gain, regardless of question type.

How did children compare the effectiveness of the two monsters' questions? One intriguing possibility is that children based their judgments on the information gain associated with each question. However, an alternative possibility is that children might have simply selected the question targeting the most frequent reason for Toma being late in the previous days (e.g., waking up late), therefore likely to be the one with the highest probability of receiving positive feedback (i.e., a "yes" response). The design of Experiment 1A does not allow us to distinguish between these two interpretations: Both the constraint-seeking question in the Uniform condition and the hypothesis-scanning question in the Skewed condition have higher information gain, but they also have a higher probability of receiving positive feedback. We test this alternative explanation in Experiment 1B.

Experiment 1B

Method

Participants. Participants were 54 5-year-olds (29 female, M_{age} = 64.7 months; *SD* = 9.6 months) recruited at local museums and schools. Twelve additional children were excluded from the analyses for failing to respond to the test question (N = 5), experimenter error (N = 2), or parental interference (N = 5). None of these children participated in Experiment 1A.

Design and procedure. We tested children in a modified Skewed condition. The hypothesis space based on past frequencies was designed to pit a question with higher information gain against a question with the highest probability of receiving positive feedback.

Each child was randomly assigned to one of two storybooks. The storybooks had the same cover story as in Experiment 1A and shared a same statistical structure, but they featured two different sets of specific reasons in order to reduce potential effects related to children's idiosyncratic preferences. In both stories, Toma had been late to school on eight days. On five out of eight days, Toma was late to school for the same reason (e.g., he could not find his shoes), and on the other three days he was late for three different reasons. Therefore, a total of four different hypotheses were included (see Table B1 of the Appendix).

In both stories, the constraint-seeking question (e.g., "Toma, were you late because of something you could not find?") targeted two hypotheses that occurred over six days: The most frequent hypothesis, which occurred on five of the days (i.e., he could not find his shoes), plus one other hypothesis, which occurred on just one of the days (e.g., he could not find his jacket). The information gain for this constraint-seeking question was IG = .81. In contrast, the hypothesis-scanning question ("Toma, were you late because you could not find your shoes?") targeted the most frequent hypothesis, which occurred on five days. The information gain for the hypothesis-scanning question was IG = .95. Therefore, the hypothesis-scanning question was more informative. However, the constraint-seeking question had a higher probability of resulting in positive feedback (p = .75, since it targeted 6 out of 8 days) as compared to the hypothesis-scanning question (p = .625, since it targeted 5 out of 8 days).

Results and Discussion

Preliminary analyses found no effects of gender or the order in which the two monster friends (i.e., Dax and Wug) were presented. Subsequent analyses were collapsed over these variables.

When predicting who would find out first why Toma was late to school, 70% of the children selected the monster who asked the more informative hypothesis-scanning question, exact binomial p (two-tailed) = .004, even though this question had a lower probability of resulting in positive feedback. A chi-square test showed no difference between the distributions obtained for the two different storybooks, $\chi^2(2, N = 54) = 0.01$, p = .95.

The results of Experiment 1B rule out the alternative interpretation that children in Experiment 1A judged the questions' effectiveness according to the probability of receiving positive feedback. With the use of two storybooks featuring different stimuli, it is also unlikely that our results were driven by children's idiosyncratic preferences. However, in both Experiment 1A and 1B, children might have used past frequencies as a salient cue for identifying the most effective question, thus selecting the question that targeted the single most frequent hypothesis (e.g., waking up late). Experiment 1C tests this alternative interpretation.

Experiment 1C

Method

Participants. Participants were 54 5-year-olds (24 female, M_{age} = 65.6 months; *SD* = 8.5 months) recruited at local museums and schools. Six additional children were excluded from the analyses because they failed to answer the test question (N = 3) or experimenter error (N = 3). None of these children participated in Experiments 1A or 1B.

Design and procedure. We tested children in a modified Skewed condition. The hypothesis space was designed to pit a question with higher information gain against a question targeting the most frequent hypothesis.

Once again, each child was randomly assigned to one of two storybooks, sharing the same statistical structure but featuring two different sets of specific reasons in order to reduce potential effects related to children's idiosyncratic preferences. In both stories, Toma had been late to school on ten days. On three out of ten days, Toma was late to school for the same reason (e.g., he woke up late), and on the other seven days, he was late for seven different reasons. Therefore a total of eight different hypotheses were included (see Table B1 of the Appendix).

In both stories, the constraint-seeking question (e.g., “Toma, were you late because you could not find something”) targeted four of the different hypotheses that occurred over four days: Toma was late because he could not find his shoes, or he could not find his jacket, or he could not find his books, or he could not find his lunchbox. The information gain for the constraint-seeking question was IG = .95.

The hypothesis-scanning question (“Toma, were you late because you woke up late?”) targeted the single most frequent hypothesis, which occurred on three days. The hypothesis-scanning question had a lower information gain of IG = .88. Therefore, the constraint-seeking question was more informative, even though the hypothesis-scanning question targeted the single most frequent hypothesis.

Results and Discussion

Preliminary analyses found no effects of gender or the order in which the two monster friends (i.e., Dax and Wug) were presented. Subsequent analyses were collapsed over these variables.

Overall 72% of the children selected the monster asking the constraint-seeking question, exact binomial p (two-tailed) < .001. A chi-square test showed no significant difference between the distributions obtained for the two different storybooks, $X^2(2, N = 54) = 0.66$, $p = .41$.

The results of Experiment 1C rule out the interpretation that children in the Skewed conditions of Experiments 1A and 1B selected the hypothesis-scanning question simply because it targeted the single most frequent hypothesis. With the use of two sets of storybooks, it is also unlikely that the results of Experiment 1C were driven by children's idiosyncratic preferences.

Discussion of Experiments 1A, 1B and 1C

Experiments 1A-1C examined whether 5-year-old children were able to make predictions based on the informativeness of the presented questions. Across three experiments, we found that preschoolers were sensitive to the statistical structure of the hypothesis space presented and judged the quality of the given questions in a way that was consistent with information gain: They selected the monster asking the question with higher information gain, regardless of whether it was a constraint-seeking or hypothesis-scanning question.

This claim is supported by our results showing that children in our task appeared not to rely on simpler strategies. First, although constraint-seeking questions are usually considered superior to hypothesis-scanning questions, children reliably judged a hypothesis-scanning question as more effective when the distribution of hypotheses resulted in the latter having a higher information gain (Experiment 1A). Second, children did not simply judge questions according to the probability of receiving positive feedback, although this strategy would require a considerably simpler computation than that of information gain (Experiment 1B). Finally, children did not rely on a heuristic based on frequency—they did not judge the question targeting the single most frequent hypothesis as more effective (Experiment 1C).

In all three experiments, children were presented with only the *first* question that the monster friends asked. Based on that information, they were asked to predict which monster would find out first why Toma was late to school. In other words, we asked children to choose the best *first* question to ask, and established that 5-year-old children can make accurate one-shot judgments of the effectiveness of given questions.

In real life, however, depending on the feedback received to the first question, a learner may have to ask several additional questions to reach the solution. The most informative follow-up question might be of a different type from what was the most informative first question. For example, it could be that, although the most informative first question was a hypothesis-scanning question, the most informative follow-up question is a constraint-seeking question. In this sense, question asking is a form of adaptive learning that requires the learner to reassess and adjust the inquiry strategy along the way, depending on how the hypothesis space changes after having received feedback.

In Experiment 2 we investigate whether preschoolers are *adaptive* learners, that is, whether they can identify the most effective question iteratively, depending on how the hypothesis space changes due to feedback. To do that, we present children with cover stories similar to those used in Experiment 1A-1C, and ask them to select, between two given questions (one constraint-seeking and one hypothesis-scanning, differing in informativeness), the one they think Toma's friend, Wug, should ask to find out why Toma was late for school again. We then provide children with feedback to the selected question (yes or no), present them with a new hypothesis space (revised according to the feedback received), and ask them to select again, between two new questions (one constraint-seeking and one hypothesis-scanning, differing in informativeness), the one they think Wug should ask to find out why Toma was late for school again.

Additionally, to test whether there are any developmental changes in preschoolers' information-search strategies, we extend our investigation to include younger preschoolers (3- and 4-year-olds).

Experiment 2

Method

Participants. Participants were 100 three- to five-year-olds (45 female, M_{age} = 60.16 months; *SD* = 12.76 months) recruited at local museums and schools. None of the children participated in Experiments 1A-1C.

Design and procedure. In Experiment 2 children were presented with a shortened version of the storybook used in Experiments 1A-1C, in which the reasons for Toma being late to school were presented all within one page (*original hypothesis space;* see Table B2 of the Appendix). Children were asked to count with the experimenter the number of times Toma had been late for each of the reasons presented (e.g., "On this day, Toma was late because he woke up late. On how many days was Toma late because he woke up late? Let's count together! One, two… ten days. For ten days he was late because he woke up late.").

Children were told that Toma was late to school again today, and that his monster friend, Wug, wanted to find out why. Toma proposes a game: "I won't tell you, you have to find out. You can ask me questions to find out. The sooner you find out, the bigger the prize! Wug, what is your first question?" As in the previous experiments, we first presented children with graphical summaries to help them remember the reasons why Toma was late on previous days. The children were then given two different questions–a constraint-seeking and a hypothesis-scanning question— and asked to indicate the question they thought Wug should ask (see Figure 3).

One of the two questions presented had a higher information gain than the other. For example, in one of the conditions the constraint-seeking question in the original hypothesis space (“Were you late because you had to do something?”, targeting 7 out of the 14 given hypotheses) had higher information gain (IG = 1.00) than the hypothesis-scanning question (“Were you late because you spilled milk on your clothes?”, targeting two out of the 14 given hypotheses; IG = .59).

Because we were interested in whether those children who selected the question with higher information gain would be able to do the same iteratively, independent of the types of questions considered, the game continued *only* if children selected the question with the higher information gain. Children who selected the question with the higher information gain (i.e., in our example, the constraint-seeking question) were presented with Toma’s answer to the selected question, which was always “no” (e.g., “No, I was not late because I had to do something. Wug, what is your next question?”).

Children were then shown, on the bottom of a new page, an updated representation of the hypothesis space (*revised hypothesis space*; “These are now the reasons why Toma could be late for school, right?”), which excluded the hypotheses ruled out by Toma’s “no” feedback to the first question selected (see Figure 3). On the same page, children were presented with two new follow-up questions that Wug could ask to find out why Toma was late to school—one constraint-seeking (e.g., “Were you late because you had to go somewhere?”) and one hypothesis-scanning question (e.g., "Were you late because you woke up late?"). Children were again asked to indicate which question Wug should ask.

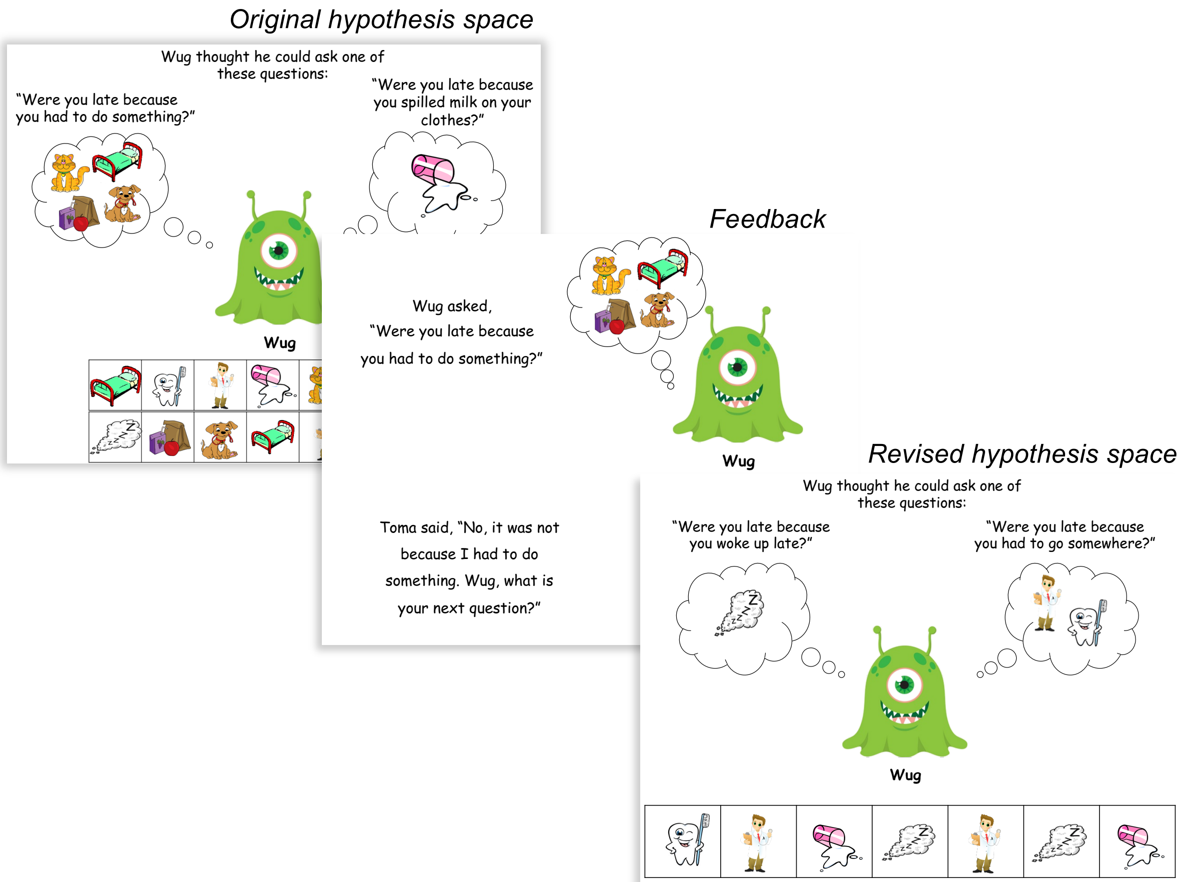

Figure 3. An example of the displays presented in Experiment 2. Children were asked to select the question that Wug should ask to find out why Toma was late to school today. Children who selected the question with the higher information gain in the original hypothesis space were given a "no" feedback. They were then shown the revised hypothesis space, and were asked to indicate which of two questions Wug should ask now. The choice was always between a constraint-seeking question, targeting multiple hypotheses, and a hypothesis-scanning question, targeting a single hypothesis. The questions varied in informativeness, as measured by expected information gain.

Children were randomly assigned to four possible conditions, across which we manipulated the question type (constraint-seeking or hypothesis-scanning) that was more informative in the original and revised hypothesis spaces. In Conditions 1 and 2, the type of question that was more informative changed between the original and the revised hypothesis space (*dynamic conditions*): In Condition 1, the hypothesis-scanning question was more informative in the original hypothesis space, whereas the constraint-seeking question was more informative in the revised hypothesis space. In Condition 2, the constraint-seeking question was more informative in the original hypothesis space, whereas the hypothesis-scanning question was more informative in the revised hypothesis space. In Conditions 3 and 4, the same type of question was more informative both in the original and revised hypothesis spaces (*static conditions;* Condition 3: hypothesis-scanning question; Condition 4: constraint-seeking question). Table 1 shows the information gain associated with the constraint-seeking and hypothesis-scanning questions presented in the original and revised hypothesis space for Conditions 1 to 4.

Table 1

Information Gain (IG) Associated with the Constraint-seeking (CS) and Hypothesis-scanning (HS) Questions Presented in the Original and Revised Hypothesis Space for Experiment 2.

		Original hypothesis space		Revised hypothesis space	
Condition Type	Condition	CS question	HS question	CS question	HS question
Dynamic	1	.75	**.86**	**1.00**	.81
	2	**.99**	.94	.81	**.95**
Static	3	.94	**.99**	.92	**1.00**
	4	**1.00**	.59	**.99**	.86

Note. Bolded numbers indicate the question with a higher IG in each hypothesis space.

Results

Overall sample. For the original hypothesis space, collapsed across the four conditions, 58% of all participants (58 out of 100) selected the question with higher information gain, exact binomial p (one-tailed) = .067 (the use of one-tailed test is justified because we had a clear hypothesis about the direction of the difference between groups based on results of Experiments 1A-1C). To examine accuracy rates, we performed a logistic regression analysis with age (in months) and condition type (Dynamic vs. Static) as predictors. The Wald criterion demonstrated that only age (p = .046) made a significant contribution to predicting accuracy, whereas condition type was not a significant predictor (p = .813). The exp(B) value indicated that older preschoolers had an increased likelihood of selecting the question with higher information gain (by 1.03 times).

When we consider only the children who selected the question with higher information gain in the original hypothesis space, 69% of all participants (40 out of 58) selected the question with higher information gain in the revised hypothesis space, exact binomial p (one-tailed) = .003. A logistic regression analysis, with age (in months) and condition type (Dynamic vs. Static) as predictors, revealed that neither age (p = .701) or condition type (p = .092) were significant predictors.

Overall, 40% of all children (40 out of 100) selected the question with higher information gain in both the original and the revised hypothesis space. This is significantly different from chance (25%), exact binomial p (one-tailed) < .001.

Age group median split analyses. To further investigate the age differences revealed by the logistic regression analysis, we split children into two age groups at the median age: younger preschoolers (50 participants, 25 per condition type, M_{age} = 49.5 months; *SD* = 6.56 months) and older preschoolers (50 participants, 24 and 26 per condition type, M_{age} = 70.82 months; *SD* = 7.35 months).

For the original hypothesis space, collapsed across the two condition types, 66% (33 out of 50) of the older preschoolers selected the question with higher information gain, exact binomial *p* (one-tailed) = .016, as compared to only 50% (25 out of 50) of the younger preschoolers (*p* = 1.00). A chi-square test revealed a marginal difference between the two age groups ($\chi^2(1, N = 100) = 2.63$, *p* = .078). A chi-square test showed no difference between children's performance in the static conditions vs. the dynamic conditions, and this was the case for both the younger (*p* = .500) and the older preschoolers (*p* = .347).

When we consider only the children who selected the question with higher information gain in the original hypothesis space, 70% of the older preschoolers (23 out of 33) selected the question with higher information gain in the revised hypothesis space, *p* (one-tailed) = .018. Although the younger preschoolers did not select the most informative first question at a level different from chance (50%) in the original hypothesis space, 68 % (17 out of 25) selected the question with higher information gain in the original hypothesis space. This proportion is marginally different from chance (50%), exact binomial *p* (one-tailed) = .054. For the revised hypothesis space, a chi-square test revealed no difference between the two age groups (*p* = .577). There were also no differences between children's performance across the two condition types in the revised hypothesis-space, for either younger (*p* = .387) or the older preschoolers (*p* = .105).

Discussion

Performance in the original hypothesis space replicated the results found in Experiments 1A-1C, showing that older preschoolers' (5-year-olds) judgments were robust across different distributions and types of hypotheses. However, we found a strong developmental effect on preschoolers' accuracy in selecting the question with higher information gain in the original hypothesis space: Only half of the younger preschoolers were able to successfully select the most informative question, as compared to 66% of the older preschoolers. The younger preschoolers might be less sensitive to the statistical structure of the environment, and lack the computational abilities needed to select informative questions. Indeed, Sobel et al. (2009) showed that 5-year-olds have probabilistic reasoning capacities that 3- and 4-year-olds do not have. For example, whereas 3- and 4-year-olds were able to generalize causal properties of objects to new members of the same set given deterministic, but not probabilistic data, 5-year-olds reliably generalized in both situations. Future research may investigate more thoroughly, from a developmental perspective, the relationship between children's ability to understand and reason with frequencies and probabilities and their ability to select informative questions.

We also found that the majority of the children who succeeded in the original hypothesis space (i.e., those who selected the question with higher information gain) also succeeded in the revised hypothesis space. This result suggests that those preschoolers who succeeded in the original hypothesis space are *ecological* learners: They can judge the effectiveness of the questions presented iteratively, rather than being limited to one-shot judgments of the most informative first question. Furthermore, they selected questions based on their informativeness within each scenario, instead of choosing according to the type of question (i.e., constraint-seeking vs. hypothesis-scanning), thereby demonstrating an early ability to revise their judgments of the effectiveness of different question types depending on the *current* hypothesis space.

General Discussion

Across four experiments, we find that over the preschool years, 3- to 5-year-old children begin to develop the computational foundations for asking informative questions. The results of Experiments 1A – 1C indicate that 5-year-olds are able to select the most informative *first* question between two presented alternatives, regardless of the question type (constraint-seeking vs. hypothesis-scanning). Experiment 2 shows that older preschoolers are *adaptive* learners: they are able to select the most informative question iteratively, based on the current hypothesis space. In contrast, younger preschoolers have not fully developed the ability to select the most informative question based on information gain. However, our results also show that those younger preschoolers who have developed this ability are adaptive learners, like the older preschoolers: They, too, are able to reassess the effectiveness of the questions iteratively, depending on the current hypothesis space.

Our results also suggest that children's judgments and behaviors are well captured by the formal measure of expected information gain: Preschoolers judge the effectiveness of questions according to how well they are expected to reduce the learner's uncertainty about the true solution in the scenario considered. Although it is unlikely that learners compute information gain as in our model, as we had acknowledged in the introduction, it is striking to observe how well this formal measure predicts children's judgments. Thus, we provide evidence for a computational level mechanism for selecting informative questions during the preschool years.

We note that we did not find that children's performance reflected the varying levels of difference in information gain between the two given questions. In particular, although in some of the presented problems the difference in information gain between the two given questions was rather small, we still found that children selected the more effective question, as measured by information gain. We speculate that our sample sizes might not have provided enough power to detect such differences. It would be worthwhile in future research to investigate whether the magnitude of the difference in information gain between two given questions mediates performance or whether there is a difference threshold beyond which participants are able to identify the most informative question and–in that case–whether such threshold change with age.

Other measures and/or information search strategies, which may be more psychologically plausible than information gain, may be able to account for the data we observed. As noted in Nelson (2005), it is not trivial to choose a formal measure that best explains people's choice of actions in active learning scenarios. For example, our model assumes children will consider all the presented reasons independently, weighting them evenly. However, this is not necessarily the case. For example, if Toma has often been late because he overslept, children might think of him as a chronic "over-sleeper". As a result of this characterization, children may consider it far more likely for Toma to be late today because he overslept again, as compared to the likelihood actually borne out by observed data. More research is necessary to test these alternative models in order to provide evidence that information gain best captures children's judgments in the domain of question-asking behavior, as well as to identify possible heuristics children may implement to approximate information gain calculations (for example the split-half heuristic, see Navarro & Perfors, 2011; Nelson et al. 2014; Nelson, Meder, & Jones, 2016; or the maximum-entropy question heuristic, see Markant, Settles, Gureckis; 2015).

In sum, by eliminating the need for children to generate questions from scratch, we demonstrate that 5-year-old children and, to some extent, even younger preschoolers (3- and 4-year-olds) are sensitive to the relative informativeness of different questions. Our results show that the computational machinery to support effective question-asking may already be present by three years of age. Future research will investigate whether young children are able to generate their own questions based on their effectiveness, and how learners implement heuristics to approximate information gain computations.

References

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018–1034.

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, *7*, 213–233. doi:10.1016/0885-2014(92)90012-G

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Chin, J., Payne, B. R., Fu, W., Morrow, D. G., & Stine-Morrow, E. A. L. (2015). Information foraging across the life span: Search and switch in unknown patches. *Topics in Cognitive Science*, n/a–n/a. doi:http://dx.doi.org/10.1111/tops.12147

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, *72*, vii–ix. doi:10.1111/j.1540-5834.2007.00412.x

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology, 64*(3), 161–185.

Corriveau, K. H., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science, 12*, 426–437.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The Sampling Hypothesis. *Cognition, 126(2)*, 285–300.

Denison, S., Reed, C., & Xu, F. (2013). Probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology, 49(2),* 243–249.

Denison, S., & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science, 34*, 885–908.

Denison, S., & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13,* 798–803.

Denison, S., & Xu, F. (2014) The origins of probabilistic inference in human infants. *Cognition*, *130,* 335–347.

Eimas, P. D. (1970). Information processing in problem solving as a function of developmental level and stimulus saliency. *Developmental Psychology*, *2*(2), 224–229. doi:10.1037/h0028746

Frazier, B. N., Gelman, S.A, & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult-child conversation. *Child development*, *80*, 1592–611. doi:10.1111/j.1467-8624.2009.01356.x

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106(1)*, 325–344.

Gonzalez, M., & Girotto, V. (2011). Combinatorics and probability: Six- to ten-year-olds reliably predict whether a relation will occur. *Cognition, 120*, 372–379.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). New York: Cambridge University Press.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science*, *17*, 455–459. doi:10.1111/j.1467-9280.2006.01727.x

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

Halberda, J. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–8. doi:10.1038/nature07246. PMID 18776888.

Harris, P. L. (2012). Trusting what you're told: How children learn from others. Cambridge, MA: Harvard University Press. doi:10.4159/harvard.9780674065192

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation, *37*, 668–683. doi:10.1037/70012-1649.37.5.668

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*(9), 678–683. http://doi.org/10.1111/j.1467-9280.2005.01595.x

Legare, C. H., Mills, C. M., Souza, A. L., Plummer, L. E., & Yasskin, R. (2013). The use of questions as problem-solving strategies during early childhood. *Journal of Experimental Child Psychology*, *114*, 63–76. doi:10.1016/j.jecp.2012.07.002

Libertus, M.E. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*. *14*(6), 1292–1300. doi:10.1111/j.1467-7687.2011.01080.x.

Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-Directed Learning Favors Local, Rather Than Global, Uncertainty. *Cognitive Science*, *40*(1), 100–120. doi:10.1111/cogs.12220

Mills, C. M., Legare, C. H., Bills, M., & Mejias, C. (2010). Preschoolers Use Questions as a Tool to Acquire Knowledge From Different Sources. *Journal of Cognition and Development*, *11*(4), 533–560. doi:10.1080/15248372.2010.516419

Mills, C. M., Legare, C. H., Grant, M. G., & Landrum, A. R. (2011). Determining who to question, what to ask, and how much information to ask for: the development of inquiry in young children. *Journal of Experimental Child Psychology*, *110*(4), 539–60. doi:10.1016/j.jecp.2011.06.003

Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. *Studies in Cognitive Growth*. New York, NY: Wiley.

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological Review*, *118*(1), 120–134. doi:10.1037/a0021110

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition*, *130*(1), 74–80. doi:10.1016/j.cognition.2013.09.007

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science, 21*(7), 960-969. doi:http://dx.doi.org/10.1177/0956797610372637

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608-631.

Piaget, J. (1954). *The construction of reality in the child.* New York: Ballantine books.

Ruggeri, A., Feufel, M. A. (2015). How basic-level objects facilitate asking efficient questions in a categorization task. *Frontiers in Psychology, 6*(918), 1-13. doi:10.3389/fpsyg.2015.00918

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143,* 203–216. doi:10.1016/j.cognition.2015.07.004

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*(12), 2159–2173. http://doi.org/10.1037/dev0000240

Ruggeri, A., Walker, C., Lombrozo, T. & Gopnik, A. Strategic use of questions to solve a causal inference task in early childhood. *Manuscript in preparation.*

Sobel, D. M., Sommerville, J. a., Travers, L. V., Blumenthal, E. J., & Stoddard, E. (2009). The Role of Probability and Intentionality in Preschoolers' Causal Generalizations. *Journal of Cognition and Development, 10*(4), 262–284. http://doi.org/10.1080/15248370903389416.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*(3), 453-489. doi:http://dx.doi.org/10.1016/S0364-0213(03)00010-7

Teglas, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. Proceedings of the

National Academy of Sciences of the United States of America, 104, 19156–19159.

Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. Science, 332, 1054–1058.

Todd, P. M., Gigerenzer, G., & the ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.

Wellman, H. M. (2011). Reinvigorating explanations for the study of early cognitive development. Child Development Perspectives, 5, 33–38.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition, 112*, 97–104.

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

Appendix A

Experiment 1A – Uniform condition

Hypothesis-scanning question. For a hypothesis-scanning question, the probability of getting a 'yes' answer is 1/6, whereas the probability of getting a 'no' answer is 5/6:

$$H_{\text{posterior}} = \frac{1}{6} H(x_{\text{yes}}) + \frac{5}{6} H(x_{\text{no}})$$

using Eq. (A.2):

$$H(x_{\text{yes}}) = 0$$

$$H(x_{\text{no}}) = 2.32$$

Therefore:

$$H_{\text{posterior}} = \frac{1}{6}\,(0) + \frac{5}{6}\,(2.32) = 1.93$$

To obtain the information gain for the hypothesis-scanning question, we use Eq. (A.1):

$$IG = 2.59 - 1.93 = 0.66$$

Constraint-seeking question. The constraint-seeking question in the Uniform condition of Experiment 1 targets three of the six hypotheses, therefore the probability of getting a 'yes' or a 'no' answer is 3/6:

$$H_{\text{posterior}} = \frac{3}{6} H(x_{\text{yes}}) + \frac{3}{6} H(x_{\text{no}})$$

using Eq. (A.2):

$$H(x_{\text{yes}}) = 1.59$$

$$H(x_{\text{no}}) = 1.59$$

Therefore:

$$H_{\text{posterior}} = \frac{3}{6}\,(1.59) + \frac{3}{6}\,(1.59) = 1.59$$

To obtain the information gain for the hypothesis-scanning question, we use Eq. (A.1):

$$IG = 2.59 - 1.59 = 1$$

Experiment 1A – Skewed condition

In the Skewed condition, there are four hypotheses presented. One hypothesis (h_{freq}) occurs on five out of eight instances, while the other three hypotheses (h_{infreq}) each occur once. Using Eq. (A.2):

$$H_{\text{prior}} = -\left[\frac{5}{8}\left(h_{freq}\right)\log_2\frac{5}{8}\left(h_{freq}\right) + 3\left(\frac{1}{8}\left(h_{infreq}\right)\log_2\frac{1}{8}\left(h_{infreq}\right)\right)\right] = 1.55$$

Hypothesis-scanning question. For the hypothesis-scanning question, the probability of getting a 'yes' answer is 5/8, whereas the probability of getting a 'no' answer is 3/8. Using Eq. (A.4):

$$H_{\text{posterior}} = \frac{5}{8}H\left(x_{\text{yes}}\right) + \frac{1}{8}H\left(x_{\text{no}}\right)$$

Using Eq. (A.2):

$$H\left(x_{\text{yes}}\right) = 0$$

$$H\left(x_{\text{no}}\right) = 1.59$$

Therefore:

$$H_{\text{posterior}} = \frac{5}{8}\,(0) + \frac{3}{8}\,(1.59) = 0.59$$

To obtain the information gain for the hypothesis-scanning question, we use Eq. (A.1):

$$IG = 1.55 - 0.59 = 0.94$$

Constraint-seeking question. The constraint-seeking question targets two of the three infrequent hypotheses, and the probability of getting a 'yes' answer is 1/4, whereas the probability of getting a 'no' answer is 3/4:

$$H_{\text{posterior}} = \frac{1}{4} H(x_{\text{yes}}) + \frac{3}{4} H(x_{\text{no}})$$

using Eq. (A.2):

$$H(x_{\text{yes}}) = 1$$

$$H(x_{\text{no}}) = 0.65$$

Therefore:

$$H_{\text{posterior}} = \frac{1}{4}\ (1) + \frac{3}{4}\ (0.65) = 0.74$$

To obtain the information gain for the hypothesis-scanning question, we use Eq. (A.1):

$$IG = 1.55 - 0.74 = 0.81$$