This paper was published by the American Psychological Association as:
Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*, 2159-2173. doi:10.1037/dev0000240

Sources of Developmental Change in the Efficiency of Information Search

Azzurra Ruggeri
University of California, Berkeley,
and Max Planck Institute for
Human Development, Berlin, Germany

Tania Lombrozo, Thomas L. Griffiths, Fei Xu
University of California, Berkeley

Abstract:

Children are active learners: they learn not only from the information people offer and the evidence they happen to observe, but by actively seeking information. However, children's information search strategies are typically less efficient than those of adults. In two studies, we isolate potential sources of developmental change in how children (7- and 10-year-olds) and adults search for information. To do so, we develop a hierarchical version of the 20-questions game, in which participants either ask questions (Study 1) or test individual objects (Study 2) to discover which category of objects within a nested structure (e.g., animals, birds, or owls) has a novel property. We also develop a computational model of the task, which allows us to evaluate performance in quantitative terms. As expected, we find developmental improvement in the efficiency of information search. In addition, we show that participants' performance exceeds random search, but falls short of optimal performance. We find mixed support for the idea that children's inefficiency stems from difficulty thinking beyond the level of individual objects or hypotheses. Instead, we reveal a previously undocumented source of developmental change: Children are significantly more likely than adults to continue their search for information beyond the point at which a single hypothesis remains, and thus to ask questions and select objects associated with zero information gain. This suggests that one crucial source of developmental change in information search efficiency lies in children's "stopping rules."

Keywords:
information search, active learning, 20-questions game, cognitive development

Children actively engage with their environment from an early age. As soon as they can sit or walk, infants spontaneously grab and manipulate objects and approach or avoid people. As language develops, young children ask about the meaning of words, request the labels of objects, and inquire about the many new and puzzling phenomena they encounter. Piaget (1954) theorized that active engagement with the world is a crucial component of learning. As children explore, they come across information that does not fit within their existing conceptual structures. This cognitive *disequilibrium* generates a state of uncertainty that is uncomfortable, and that motivates children to devote their mental energy and efforts to adapting or developing new conceptual structures that better accommodate the new information. Piaget advocated the view that children are *active learners*, and described behaviors of his own children that looked very much like testing hypotheses with self-generated data.

Although Piaget did not develop a full-fledged theory of how active learning works, more recent and systematic research in cognitive development has supported the idea that children are indeed active learners. For instance, there is evidence that infants, toddlers, and preschoolers control their attention based on how much information can be gained from the environment (e.g., Kidd, Piantadosi, & Aslin, 2012), and that even young children have some rudimentary ability to generate evidence to figure out the causal structure of the world (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012; Schulz, Gopnik, & Glymour, 2007; Sim & Xu, 2014).

Beyond direct exploration through observations and interventions, children can actively gather evidence from people by asking questions (Callanan & Oakes, 1992; Campos & Stenberg, 1981; Chouinard, 2007; Meltzoff, 1988a, 1988b, 1990). When engaged in conversations with adults, preschool-aged children ask on average over 100 questions per hour (Chouinard, 2007). According to Tomasello (1999a, 1999b), the ability to seek out information from knowledgeable others gives us a particular evolutionary advantage, allowing us to learn efficiently. Indeed, research to date has shown that young children ask questions that support the acquisition of relevant and reliable information. For example, preschool-aged children ask domain-appropriate questions (Greif,

*This research was supported by the Seventh European Community Framework Programme Marie Curie International Outgoing Fellowship awarded to A. Ruggeri. T. Lombrozo was supported by a McDonnell Scholar Award on Understanding Cognition. F. Xu was supported by a China Distinguished Foreign Expert Award. We thank Celina Vicuna, Minh-Thy Nguyen, Meg Bishop, Samika Kumar, Simone Riley, and Justine Hoch for assistance in recruiting, data collection, and coding, as well as Zi Lin Sim and Anita Todd for their helpful comments. We would also like to thank all the students and teachers from the Madera Elementary School, in El Cerrito, and the Laurence Hall of Science, in Berkeley, for providing research space.**Correspondence concerning this article should be addressed to Azzurra Ruggeri, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee, 94, 14195 Berlin, Germany*
Email: ruggeri@mpib-berlin.mpg.de

Kemler Nelson, Keil, & Gutierrez, 2006), and they preferentially consult reliable informants (Birch, Vauthier, & Bloom, 2008; Corriveau & Harris, 2009; Gweon, Pelton, Konopka, & Schulz, 2014; Koenig & Harris, 2005; Mills, Legare, Bills, & Mejias, 2010; Mills, Legare, Grant, & Landrum, 2011; Sabbagh & Baldwin, 2001; Sobel & Corriveau, 2010). Even 2- and 3-year-olds have reasonable expectations about which responses count as satisfying answers to their questions: They tend to agree and ask follow-up questions when adults provide explanatory answers, but reask their original question, or provide their own explanations, otherwise (Frazier, Gelman, & Wellman, 2009; see also Lombrozo, 2016). Finally, we know that 4-year-olds have the capacity to acquire relevant information through strategic questions, and that by age 5 they reliably *use* this information to solve problems (Chouinard, 2007; Legare, Mills, Souza, Plummer, & Yasskin, 2013; Ruggeri & Feufel, 2015; Ruggeri & Katsikopoulos, 2013).

These findings provide compelling evidence that children's questions play an important role in learning: They are generated with expectations about the content and reliability of possible answers, and those answers are evaluated and eventually used to guide behavior. At the same time, there is evidence that children's questions tend to be *inefficient*. To investigate the efficiency of children's questions, studies have used variants of the "20-questions game," in which participants try to identify an unknown target object by asking as few yes-or-no questions as possible, either generating the questions from scratch (e.g., Chouinard, 2007; Legare et al., 2013; Mosher & Hornsby, 1966; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015) or selecting them from a list of given alternatives (Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014; Ruggeri & Feufel, 2015). This research has found that the ability to ask efficient questions undergoes a large developmental change from age 4 to adulthood. Younger children tend to ask "hypothesis-scanning" questions, which offer tentative solutions by targeting individual hypotheses or objects (e.g., "Is it the dog?") and typically support a less efficient path to the correct solution (Ruggeri & Lombrozo, 2015). In contrast, older children and adults more readily ask "constraint-seeking" questions, which can more efficiently partition the hypothesis space by targeting superordinate categories or features that are shared by multiple hypotheses (e.g., "Is it an animal?" or "Does it have a tail?").

It thus appears that asking questions is a powerful mechanism for learning, but one that children do not master until late childhood. This raises important questions about the *sources* of younger children's inefficiency, and about what *changes* over the course of development. In the present work, we take a new approach to studying these questions, testing 7-year-olds, 10-year-olds and adults in a novel information search paradigm. Specifically, we employ two "hierarchical" versions of the 20-questions task in order to isolate distinct hypotheses about potential developmental sources of inefficiency. In these hierarchical tasks, objects conform to a hierarchical structure (e.g., there are two owls, in a set of four birds, in a set of eight animals), and the solution to the task can occur at any hierarchical level (e.g., owls, birds, or animals).

Our hierarchical task allows us to explore two hypotheses. The first hypothesis is that children's inefficiency stems from difficulty going beyond the level of individual *objects* (e.g., "Is it this dog?") to ask questions that target higher-order properties of objects, such as shared features (e.g., "Does it have a tail?") or category membership (e.g., "Is it an animal?"). The second hypothesis is that children experience difficulty going beyond individual *hypotheses* (whether hypotheses correspond to individual objects or to groups of objects) to generate questions that target multiple hypotheses. Having solutions that can correspond to *groups* of objects (e.g., dogs or animals) allows us to pull apart the object level from the hypothesis level.

We also analyze our results within a formal computational framework, which allows us to measure performance in absolute terms. By differentiating between *nonoptimal* questions (i.e., those that are informative, but not optimally) and *unnecessary* questions (i.e., those that are not informative at all), this quantitative analysis allows us to investigate a third hypothesis about children's inefficiency in asking questions: that they fail to recognize when enough information has been collected to solve the task, and therefore continue asking questions beyond the point at which they are informative. In the sections that follow, we explain these alternative hypotheses and how our tasks and formal framework address them.

Young Children's Inefficiency: Single Objects or Single Hypotheses?

Why do younger children tend to ask hypothesis-scanning questions in a 20-questions task, despite their typical inefficiency? A dominant explanation is that young children have a hard time going beyond the object level, failing to spontaneously represent, and therefore target, more abstract categories or features. Consistent with this idea, Ruggeri and Feufel (2015) found that scaffolding higher-level representations facilitated children's ability to ask constraint-seeking questions. In their study, 7- and 10-year-old children, as well as adults, were presented with 20 cards on a computer screen, each of which contained a word label (e.g., "dog" or "sheep"). Participants were randomly assigned to one of two experimental conditions based on the specificity of the label: a basic-level condition (e.g., "dog") or a subordinate-level condition (e.g., "Dalmatian"). The authors found that participants were more likely to ask constraint-seeking questions in the former condition than in the latter, suggesting that more abstract labels facilitated a shift away from object-based reasoning when generating questions. They also found that the ability to generate more abstract features (e.g., "a dog is a mammal" or "a dog has four legs") is one factor that affects performance and that developed within their age range (see also Herwig, 1982). With even younger children, Legare and colleagues (2013) found that the ability to identify and to flexibly categorize objects based on alternative features (e.g., color and pattern) predicted how well 4- to 6-year-olds could generate effective questions.

The traditional 20-questions task used in prior research has a crucial limitation: the solution to the task is always a single object. As a result, individual hypotheses (i.e., candidate solutions) necessarily correspond to individual *objects* (e.g., the dog). A tendency to ask hypothesis-scanning questions could reflect a tendency to represent the problem at the *object level,* a tendency to do so in terms of *individual hypotheses*, or both. In other words, the traditional 20-questions task cannot differentiate between two possibilities: (a) that younger age groups (i.e., 7-year-olds as compared to 10-year-olds, and 10-year-olds as compared to adults) have more difficulty going beyond object-level representations to

generate questions that target higher-order features (such as “mammal” for animals, or “brown-haired” for people), and (b) that younger age groups have a harder time going beyond individual hypotheses when representing the hypothesis space, and therefore fail to ask questions that can efficiently partition the hypothesis space by targeting multiple hypotheses at once.

To differentiate between these possibilities, we developed a novel, hierarchical version of the 20-questions task in which the solution was a category of objects (e.g., “all animals” or “all birds”) rather than an individual object (e.g., “this blue owl”), with the consequence that hypotheses do not correspond to individual objects. More specifically, our hierarchical 20-questions task differs from traditional versions in two ways: (a) participants receive an array of objects that can be classified into a symmetrical nested structure organized at three category levels (Table 1), each including the same number of objects: superordinate (e.g., 8 animals and 8 plants), basic (e.g., among the animals there are 4 fish and 4 birds), or subordinate (e.g., among the birds, there are 2 different owls and 2 different parrots); and (b) participants are told that *some* objects share a novel property (e.g., existing on an alien planet, or activating a machine) and are asked to find out which objects do so. Participants’ task is not to identify *one object*, but to discover at what category level objects share a novel property (e.g., is it the animals, the birds, or the owls that are found on planet Apres?), where the solution can occur at any category level. For instance, the solution could be that “birds” are found on the planet, in which case participants would be told “yes” if they asked whether the owls are on planet Apres, and “some” if they asked whether animals are on planet Apres.

The studies we report employed a second innovation relevant to differentiating between the theories that younger children are object-bound versus hypothesis-bound. We developed two versions of the hierarchical 20-questions task: one in which participants could ask questions (question-asking) and another in which they could only select individual objects to test (object-selection). Whereas questions can target individual objects (e.g., “is the blue owl on planet Apres?”) or individual hypotheses at different levels (e.g., “are all the animals on planet Apres?,” or “are all the birds on planet Apres?”), the object-selection task requires that all participants work at the object level, as they effectively have no choice but to “ask” about individual objects (e.g., the blue owl).

If a key source of developmental change in the efficiency of information search concerns the ability to generate questions that go beyond individual objects by targeting higher-order features, then we would expect a developmental improvement in performance in our question-asking version of the hierarchical 20-questions task. Specifically, we would expect older age groups to be better able to generate questions at higher levels of the hierarchy (e.g., “are all the animals found on the planet?”), implementing a more efficient search than that available to object-bound younger children (e.g., “is the blue owl on the planet?”). However, we would expect this developmental difference to disappear in the object-selection version of the task, which forces both adults and children to proceed with an object-based search.

If children’s inefficiency instead (or additionally) stems from difficulty reasoning beyond individual hypotheses, we would predict a different profile for performance. Specifically, we would expect older age groups to more readily recognize the relationships between hypotheses and therefore to generate queries “top-down,” capitalizing on the hierarchical structure of the hypothesis space to begin testing superordinate hypotheses first, then basic, and then subordinate. As we show with simulations involving a formal model of the task (detailed below), solutions at higher levels of the hierarchy (e.g., “animals”) can be reached more efficiently than those at lower levels (e.g., “owls”), not only for question-asking, but also for object-selection. If only adults can exploit the hierarchical structure of the hypothesis space, we would expect them to reach solutions more quickly for solutions at superordinate versus subordinate levels for both question-asking and for object-selection. However, we would expect children to be less capable of doing so in either case.

Table 1
Materials and Scenarios Used in Studies 1 and 2

Category level	Scenario							
	What kinds of objects are present on the far-away planet Apres?							
Superordinate	Animals				Plants			
Basic	Fish		Birds		Trees		Flowers	
Subordinate	Goldfish	Clownfish	Parrots	Owls	Apple trees	Pine trees	Tulips	Daisies
	What kinds of objects can turn on this machine?							
Superordinate	Clothes				Furniture			
Basic	Shirts		Shoes		Tables		Chairs	
Subordinate	Long sleeves	Short sleeves	Flip-flops	Boots	Dining tables	Desks	Rocking chairs	High chairs
	What kinds of objects can enter this magic box?							
Superordinate	Vehicles				Fruit			
Basic	Cars		Planes		Apples		Berries	
Subordinate	Vans	Sports cars	Airplanes	Helicopters	Yellow apples	Green apples	Raspberries	Blueberries

Originally published in: Developmental Psychology, 52(12), 2016, p. 2161

Young Children's Inefficiency: Nonoptimal Questions or Unnecessary Questions?

We suggested above two different reasons why children could fail to generate the most effective questions. An additional possibility is that children do not (only) ask nonoptimal questions (i.e., those that are only moderately informative), but unnecessary questions (i.e., those that provide no information at all). Specifically, younger children, older children, and adults could differ in their "stopping rule," or in their ability to effectively apply it. A stopping rule establishes when enough information has been collected to terminate information search and make a decision, and it is thought to be a crucial building block for decision making (Gigerenzer, Todd, & the ABC Research Group, 1999). If younger age groups have an overly conservative stopping rule, they may continue to ask questions beyond the point at which the answers yield information relevant to solving a given problem.

Previous research in decision-making involving other information search paradigms—such as information boards from which children can reveal information prior to making a decision—has found that younger children tend to adopt more exhaustive and inefficient approaches than older children and adults (Davidson, 1991a, 1991b, 1996; Gregan-Paxton & John, 1995, 1997; Howse, Best, & Stone, 2003; Mata, von Helversen, & Rieskamp, 2011; Ruggeri & Katsikipoulos, 2013). These results suggest that younger children's stopping rules could be more conservative, or simply less accurate, than those of older children and adults, potentially leading them to ask unnecessary questions in a 20-questions paradigm. In fact, Legare and colleagues (2013) found that about 20% of the questions generated by preschoolers were "confirmatory" in the sense that they requested information that had already been provided. However, prior work has not analyzed search performance in quantitative terms, which makes it difficult to differentiate nonoptimal queries from unnecessary queries. For our tasks we developed a model of each participant's hypothesis space at each time point that allows us to evaluate whether children ask questions or select objects beyond the point at which a single hypothesis remains.

Before explaining the details of each study, we present a formal analysis of performance on these tasks, which is necessary to explore and disentangle potential sources of developmental change in the efficiency of information search. This framework is built on a Bayesian model of learning, with optimal learning defined in terms of information gain. We explain each of these components in turn.

Bayesian Model

Performance on the hierarchical 20-questions task can be modeled within a Bayesian framework originally developed for concept learning and generalization (Tenenbaum & Griffiths, 2001). The learner's hypothesis space is the set of hypotheses concerning which set of objects has a target property (e.g., activating a machine). In our case, the hypothesis space consists of 14 alternative hypotheses, corresponding to all the object categories at any hierarchical level: the two superordinate levels, the four basic levels, and the eight subordinate levels.

We did not consider single-object hypotheses (e.g., only the blue owl), as participants were explicitly told that the property applied to more than one object. Moreover, we did not consider disjunctive hypotheses, that is, the combination of objects across categories, such as "a boot or a desk can turn on the machine." Such hypotheses were never provided by participants as possible solutions. Finally, because participants were told that all categories, at any level, were equally likely to be correct, we assumed that participants initially expected all hypotheses to be equally likely, regardless of their level in the hierarchy.

We assume that participants update their beliefs after each observation by evaluating the hypotheses still under consideration according to Bayes' rule, computing their *posterior probability* $p(h|X)$ given all observations X, which is proportional to the product of their *prior probabilities* $p(h)$ and *likelihoods* $p(X|h)$:

$$p(h|X) = \frac{p(X|h)p(h)}{\sum_{h'} p(X|h')p(h')}$$

The prior $p(h)$ represents participants' expectations about how likely the candidate hypotheses are. Matching the instructions given to participants, we assume a uniform prior. The likelihood $p(X|h)$ represents how likely it is that X will be observed if h is true. In line with the structure of the task, we assume that for each observation $x \in X$, $p(x|h)$ is 1 if the observation is compatible with h and 0 otherwise, and that observations are independently conditioned on h, so $p(X|h)$ is just the product of $p(x|h)$ for each observation x. The posterior $p(h|X)$ is thus a function of the observations X and prior beliefs about the probability of each candidate hypothesis considered.

Expected Information Gain

Several possible measures can be used to compute how informative each search option is, whether it's a question to ask or an object to test. These include probability gain, impact, and expected savings (Nelson, 2005). Following previous research on the 20-questions task (Nelson et al., 2014; Ruggeri & Feufel, 2015), we measured the informativeness of participants' queries in terms of their expected information gain (e.g., Oaksford & Chater, 1994). At each step of the search process, an optimal learner evaluates all possible actions in terms of their *information gain* (*IG*), computed by subtracting the *predicted posterior entropy* from the *prior entropy*:

$$IG = H_{\text{prior}} - H_{\text{posterior}}.$$

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

$$H_{\text{prior}} = -\sum_{h} p(h)\log_2 p(h)$$

The predicted posterior entropy $H_{\text{posterior}}$ refers to the predicted uncertainty after the action is chosen and the corresponding feedback is observed. 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:

$$H_{\text{posterior}} = \sum_{x_i} p(x_i|X)H(x_i)$$

where x_i is a possible observation, $p(x_i|X)$ is the probability of that observation resulting from taking the candidate action given all the

information from previous observations, and $H(x_i)$ is the entropy of the posterior distribution over hypotheses after observing x_i. More formally,

$$p(x_i|X) = \sum_h p(x_i|h)p(h|X)$$

$$H(x_i) = -\sum_h p(h|X, x_i)\log_2 p(h|X, x_i)$$

With these formal tools, we can compare children's and adults' performance to that of an optimal learner and better identify sources of inefficiency in their information search.

Overview of Studies

In two studies, we explore the three potential sources of inefficiency in children's information search identified above. These sources are not mutually exclusive—children could face challenges going beyond the object level, going beyond individual hypotheses, *and* implementing an effective stopping rule. However, these sources can be isolated and identified with appropriate tasks and analyses. Our studies use two hierarchical versions of the 20-questions task, one involving question-asking (Study 1) and the other object-selection (Study 2). In both studies we tested three age groups: 7-year-olds, 10-year-olds, and adults. These age ranges were motivated by prior research suggesting a strong developmental shift in children's information search strategies between the ages of 7 and 10 (Mosher & Hornsby, 1966; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015).

Study 1: Question Asking

In Study 1, participants completed the hierarchical 20-questions task by asking questions. We additionally modeled optimal and random performance using our formal analysis. The task and analyses allowed us to test three hypotheses. First, if children's inefficiency is related to difficulty going beyond the object level to identify higher-order features to target in generating questions, then we expect to replicate prior work in finding a large developmental change in the efficiency of information search, with adults requiring fewer questions to reach the solution than 7-year-olds, and 10-year-olds falling in between. Second, if children's inefficiency as compared with adults is related to the challenge of going beyond individual hypotheses to ask questions that target multiple hypotheses at once, then we would additionally expect children, but not adults, to fail to identify the solution with fewer questions when it is at a higher level. Finally, if children's inefficiency stems from an overly conservative stopping rule, then we would expect to see children continuing to ask questions beyond the point at which only a single hypothesis remains.

Method

Participants. Participants in Study 1 were 24 children in second grade (10 female, M_{age} = 90.5 months; *SD* = 5.56), and 23 children in fifth grade (8 female, M_{age} = 119.4 months; *SD* = 12.7), recruited from a primary school and a local children's museum, as well as 23 university students (15 female, M_{age} = 21.1 years; *SD* = 2.6 years). Although most participants were from White, middle-class backgrounds, a range of ethnicities reflecting the diversity of the population were represented.
Materials. We used three scenarios for the hierarchical 20-questions task (see Table 1), each involving 16 distinct objects. In the "planet Apres" scenario, participants had to find out what kind of objects can be present on a far-away planet; in the "machine" scenario, the participants had to find out what kind of objects can turn on a machine; and in the "magic box" scenario, participants had to find out what kind of objects can enter a magic box. Participants completed all three scenarios in a random order, with the three hierarchical levels of the objects constituting the solution (subordinate, basic, and superordinate) randomly assigned to the three scenarios. The category of objects constituting the solution was chosen randomly (e.g., Fish, Birds, Trees or Flowers could be the solution for the basic level condition in the planet Apres scenario).
Procedure. The experimental session started with a short familiarization phase, aimed at making the children comfortable playing with the tablet on which the objects were displayed and following the experimenter's instructions. In three trials, participants were presented with an array of six objects and had to identify and select some of the objects by touching them on the screen in response to the experimenter's instructions. For example, they were asked to touch all the circles or all the stars. After the familiarization, participants began the first of three test trials, in which they had to identify the target category by asking yes-or-no questions. After each question (e.g., "Can birds be found on planet Apres?"), participants received feedback from the experimenter with a response of "yes," "no," or "some." The response "some" was provided when only some members (e.g., parrots) of the category targeted (e.g., birds) shared the novel property (e.g., being found on planet Apres). Participants were prompted to put a red ("no" feedback) or green ("yes" feedback) frame around the object(s) to which their question referred by touching the object(s) on the tablet. For example, if the experimenter said "yes" to whether birds can be found on planet Apres, the participant was asked to put a green frame around each bird. The experimenter made sure that participants framed all and only the objects each question applied to. This procedure ensured that the participants understood the feedback and it reduced memory demands, as participants did not have to remember their questions or the feedback received as the task proceeded. Cross-category questions—that is, those that targeted groups of objects belonging to different categories (e.g., "Does it have something brown?" in the planet Apres scenario, which targets all trees and the brown owl)—were not answered: Children were told that the computer did not know the answer to that question, and they were prompted to ask another question.[1] After receiving feedback, participants could choose whether to ask another question or guess the solution. Participants could ask questions and guess the solution as often as they wanted, but they were told to find the solution with as few questions as possible.

---

[1] Overall, participants asked 29 cross-category questions, that is, questions that targeted groups of objects belonging to different categories, such as "Does it have something brown?" in the planet Apres scenario (*N*7-year-olds = 18; *N*10-year-olds = 3; *N*adults = 8). Such questions were not answered ("I don't know the answer to this question, can you try another question?") and were excluded from further analyses.

At the end of the three trials constituting the experimental session, participants performed a sorting task to determine whether they understood the hierarchical structure of a scenario and were able to verbally label categories at each level. They were given 16 cards, representing the objects with which they were presented in the second experimental trial, and had to sort them into two piles (superordinate level) such that the two groups differed from each other but the objects within each group were similar. The second step required sorting the cards within each of these two subgroups into two new piles (basic level), and the third step required sorting the cards within each of these four subgroups into two new piles (subordinate level). The sorting task eventually produced eight groups of two objects each. At each step, participants were asked to name each group they had sorted the objects into. When participants, at any stage of the sorting task, did not organize the objects according to the expected hierarchical categorization (e.g., if they sorted them by color), the experimenter prompted the participant to sort the objects differently (i.e., "is there another way to sort the objects into two groups?"), without suggesting any specific way to do so.

Simulations

To model the behavior of an optimal learner, we simulated 10,000 step-optimal search trajectories for the question-asking paradigm. A step-optimal search trajectory is one that selects, at each step, the question with the highest expected information gain. For example, the most efficient first question targets a superordinate category (e.g., "Are animals on planet Apres?"; *IG* (information gain) = 1.29). Questions that target a basic (e.g., birds) or subordinate (e.g., parrots) category would achieve lower information gain (basic: *IG* = 1.16; subordinate: *IG* = 0.75). If the feedback on this first question is "yes," the search is over, the solution being "all animals are on planet Apres." If the feedback is "no," the most informative follow-up question targets a basic-level plant category (i.e., either trees or flowers, *IG* = 1.56). Targeting the superordinate category plants (*IG* = 0.60) or a subordinate-level plant category (e.g., daisies, *IG* = 0.98) would achieve a lower information gain (Figure 1, top panel). Targeting any animal category would not provide any information (*IG* = 0). If the feedback is "some," the most informative follow-up question targets a basic-level animal category (i.e., either fish or birds, *IG* = 1.46). Targeting a subordinate-level animal category (e.g., parrots) would achieve a lower information gain (*IG* = .92). Targeting any plant category would not provide any information (*IG* = 0).

Results

Simulation results for optimal learners. For the question-asking paradigm, the optimal question-asking strategy reached the solution, on average, in 2.85 questions. An analysis of variance (ANOVA) with the number of questions needed to reach the solution as a dependent variable and solution level (3 levels: subordinate, basic, superordinate) as a between-subjects variable found that this number varied with the hierarchical level of the solution (Figure 2), $F(2, 299) = 60.03$, $p < .001$, $\eta^2 = .288$. A Bonferroni-corrected multiple-comparisons analysis revealed that fewer questions were needed when the solution was at the superordinate level (*M*superordinate = 2.08 *SE* = 0.14) than at the basic level (*M*basic = 2.80, *SE* = 0.12, $p < .001$), and that fewer questions were needed when the solution was at the basic level than at the subordinate level (*M*subordinate = 3.35, *SE* = 0.07, $p < .001$; see Figure 2).

Empirical results for human learners. Overall, participants required an average of 4.24 questions to reach the solution. Even adults (*M*adults = 3.36, *SE* = 0.35) required significantly more questions than the simulated optimal question asker (2.85), $t(321) = 3.143$, $p = .002$.

To analyze the overall efficiency of participants' question asking, we performed a mixed ANOVA with the number of questions asked prior to providing the correct solution as a dependent variable, age group (3: 7-year-olds, 10-year-olds, adults) as a between-subjects variable, and solution level (3: superordinate, basic, subordinate) and trial number (3) as within-subject variables. This analysis revealed a main effect of age group, $F(2, 67) = 5.29$, $p = .007$, $\eta^2 = .136$ (see Figure 2). A Bonferroni-corrected multiple-comparisons analysis confirmed that 7-year-olds (*M*7-year-olds = 4.92, *SE* = 0.34) asked more questions than adults (*M*adults = 3.36, *SE* = 0.35, $p = .006$). We did not find differences between the number of questions asked by 7- and 10-year-olds (*M*10-year-olds = 4.38, *SE* = 0.35, $p = .807$), nor by 10-year-olds and adults ($p = .126$).

We also found a main effect of solution level, $F(2, 67) = 20.02$, $p < .001$, $\eta^2 = .320$. Mirroring the findings for optimal question-asking, a Bonferroni-corrected multiple-comparisons analysis found that participants asked fewer questions in the superordinate condition (*M*superordinate = 3.37, *SE* = 0.26) than in the basic condition (*M*basic = 4.08, *SE* = 0.24, $p = .038$), and in the basic condition than in the subordinate condition (*M*subordinate = 5.21, *SE* = 0.31, $p < .001$; see Figure 2). We did not find an interaction between age group and solution level ($p = .120$), nor effects of scenario or trial number.

Participants' initial questions. Although we failed to find a significant interaction between age group and the hierarchical level of the solution, an analysis of participants' initial questions does suggest that adults were better able to capitalize on the hierarchical structure to ask informative questions (Figure 3).

Across the three trials, adults asked a larger number of first questions at the most informative superordinate level (*IG* = 1.29; *M*adults = 1.87, *SE* = 0.16) than did older children (*M*10-year-olds = 1.13, *SE* = 0.20), $t(162) = -3.33$, $p = .001$, who, in turn, asked a larger number of such questions than younger children (*M*7-year-olds = 0.41, *SE* = 0.16), $t(145) = -3.02$, $p = .003$. Symmetrically, adults asked fewer first questions at the least informative subordinate level (*IG* = 0.75; *M*adults = 0.13, *SE* = 0.06) than older children (*M*10-year-olds = 0.29, *SE* = 0.10), $t(162) = -3.14$, $p = .002$, who, in turn, asked fewer such questions than younger children (*M*7-year-olds = 0.69, *SE* = 0.17), $t(145) = -2.14$, $p = .034$. The number of initial questions at the basic level (*IG* = 1.16) did not vary significantly across age groups (*p*s > .05): *M*7-year-olds = 0.79, *SE* = 0.18, *M*10-year-olds = 0.92, *SE* = 0.17, *M*adults = 0.55, *SE* = 0.12.

Participants' performance versus optimal history-matched and random history-matched models. The analyses presented above demonstrate that children and adults fell short of optimal information search, which represents a potential upper bound on their performance. However, they do not provide a reasonable lower bound for comparison, nor do they reveal how the efficiency of information search may have varied over the course of inquiry. To assess performance in a more fine-grained way, we compared

the expected information gain of participants' questions at each time point with two additional models: the optimal history-matched model, which provides an upper bound on expected performance, and the random history-matched model, which provides a lower bound.
The optimal history-matched model selects at each step the question that has the highest information gain, considering the current hypothesis space as defined by a given participant's previous selections. The random history-matched model selects a question at random. This random selection is repeated 10 times at

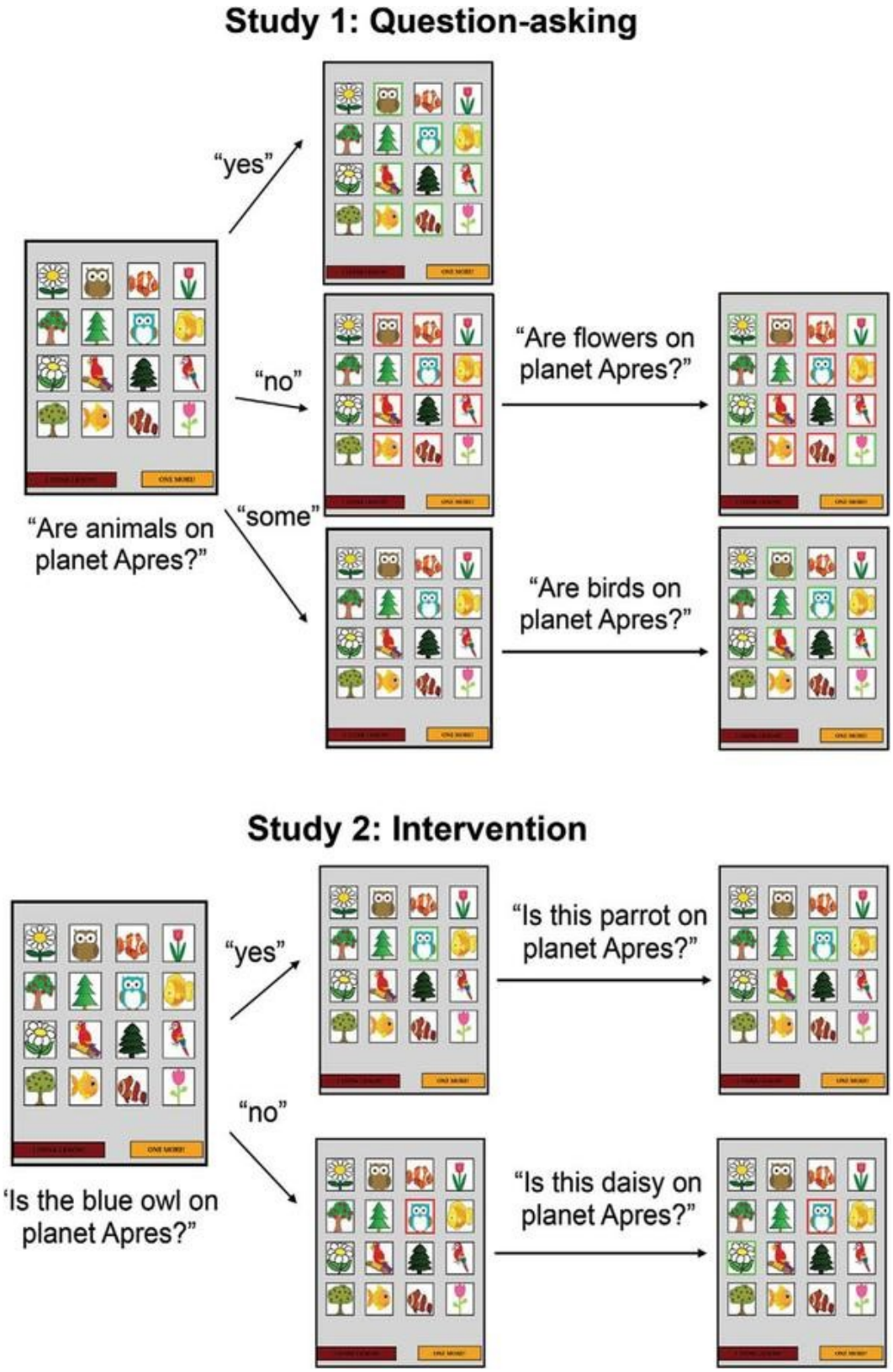
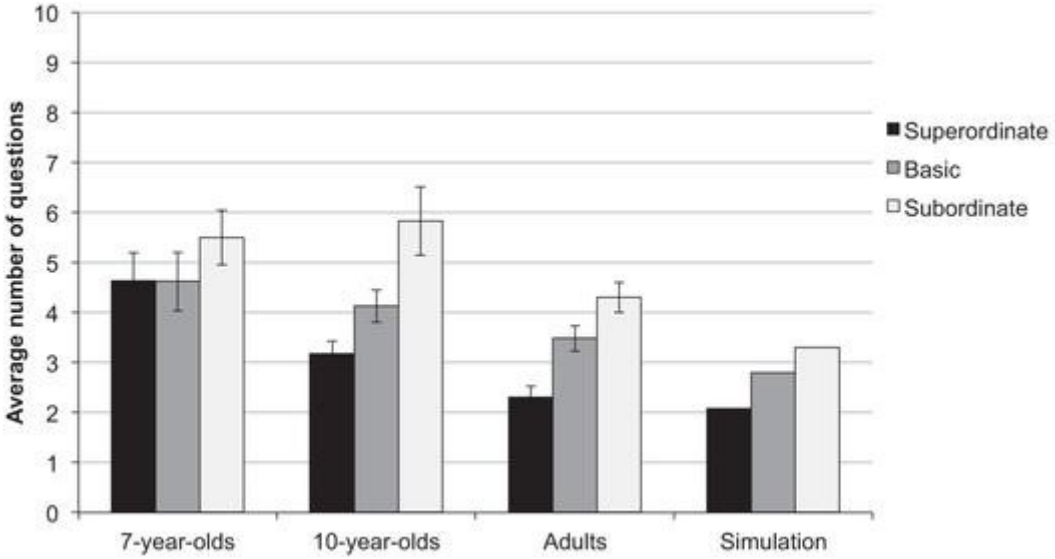


Figure 1. First steps of the step-optimal search path for the question-asking (Study 1, top panel) and object-selection (Study 2, bottom panel) paradigms in the planet Apres scenario. See the online article for the color version of this figure.

Originally published in: Developmental Psychology, 52(12), 2016, p. 2165

Figure 2. Study 1: question-asking. Average number of questions asked before offering the solution, displayed by category level of the solution (i.e., solution condition). Error bars represent 1 *SEM* in each direction.



each step, with replacement, to simulate a representative random average, and we considered the average information gain of the 10 randomly selected questions.

A repeated-measures ANOVA with age group (3: 7-year-olds, 10-year-olds, adults) as a between-subjects factor and model (3: participants, optimal history-matched, random history-matched) as a within-subject factor found that participants' average information gain (*M*participants = 0.83, *SE* = 0.02) was higher than the information gain resulting from a random selection (*M*random = 0.50, *SE* = 0.01), but lower than that resulting from the optimal history-matched model (*M*optimal = 1.06, *SE* = 0.02), $F(2, 134) = 697.97$, $p < .001$, $\eta^2 = .91$. This analysis confirms that participants fell short of optimality, but additionally demonstrates that their performance was significantly better than chance.

This approach to quantifying participants' performance also allowed us to analyze the efficiency of information search at each time point, and to identify unnecessary questions. Figure 4 displays the average information gain as predicted by the optimal history-matched and random history-matched models and the actual information gain of participants' questions as the task unfolded. The figure suggests two sources of inefficiency. For the first several questions—until the hypothesis space was narrowed down to a single hypothesis—the average information gain associated with participants' questions fell below that of the optimal history-matched model (Table 2)—that is, they asked nonoptimal questions. But additionally, some participants continued to ask questions beyond the point at which only a single hypothesis remained—that is, they asked unnecessary questions. Such questions would necessarily have zero information gain and therefore contributed to overall inefficiency. In fact, 83% of the 7-year-olds ($n = 20$), 70% of the 10-year-olds ($n = 16$), and 52% of the adults ($n = 12$) asked, in at least one trial, more questions than strictly necessary to identify a single hypothesis. A repeated-measures ANOVA revealed an effect of age group on the number of questions asked beyond the point at which a single hypothesis remained, $F(2, 67) = 4.50$, $p = .015$, $\eta^2 = .118$. A Bonferroni-corrected multiple-comparisons analysis confirmed that the 7- and 10-year-olds asked, on average, more additional questions (*M*7-year-olds = 1.28, *SE* = 0.22; *M*10-year-olds = 0.81, *SE* = 0.23) than adults (*M*adults = 0.32, *SE* = 0.23, $p = .011$). We did not find any

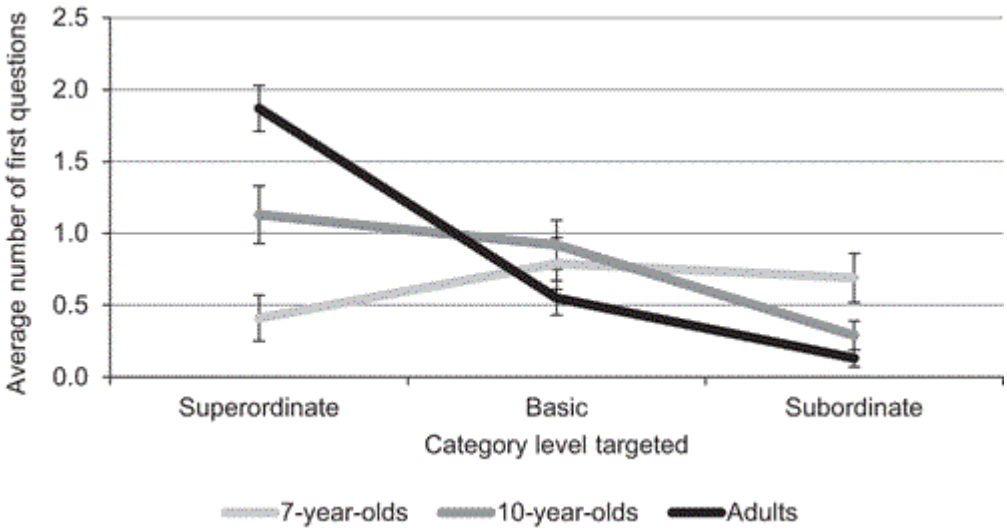
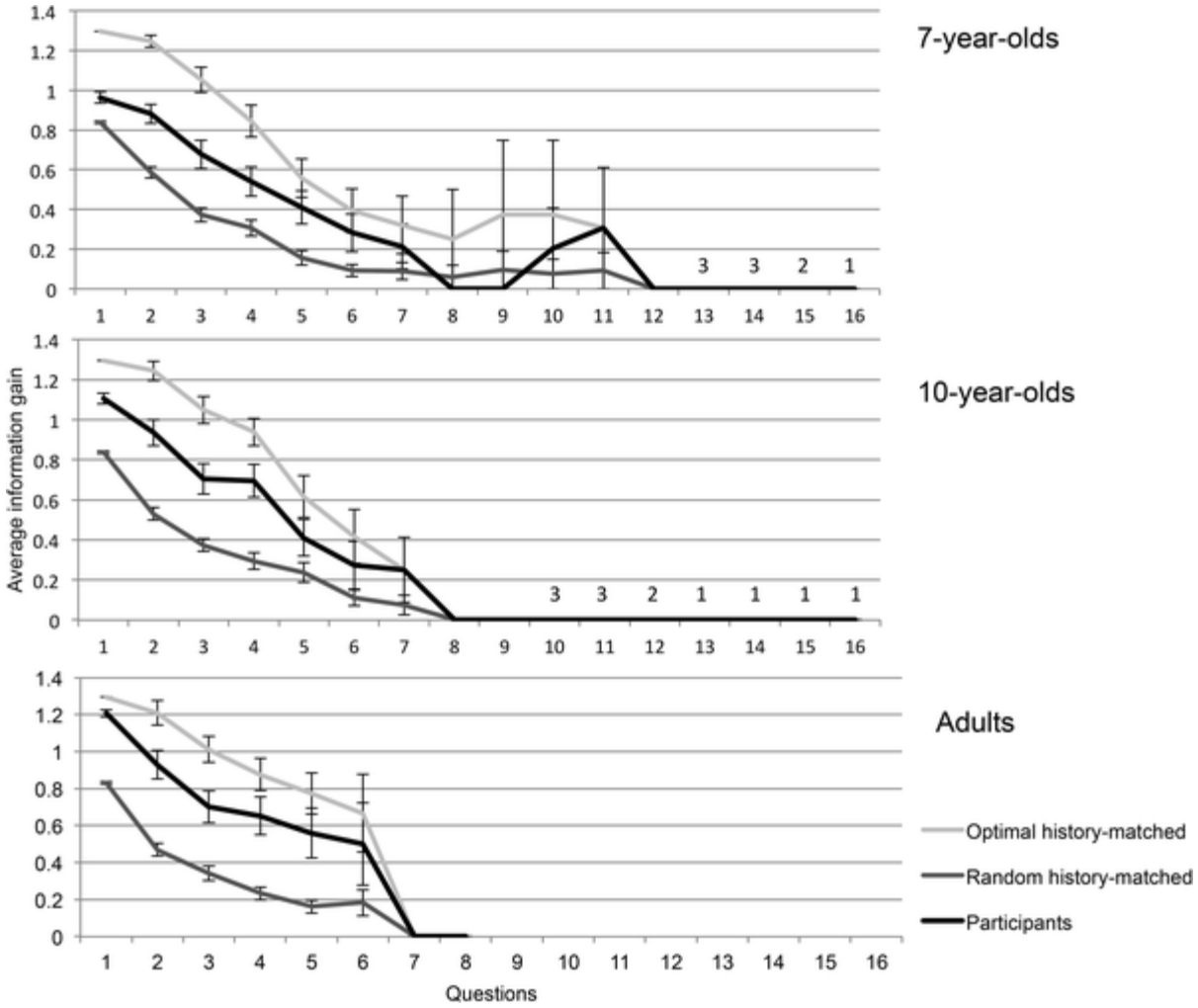


Figure 3. Study 1: question-asking. Average number of first questions (across the three trials) asked at the superordinate, basic, and subordinate levels by participants in Study 1 (question asking), displayed by age group. Error bars represent 1 *SEM* in each direction.

Figure 4. Study 1: question asking. Average information gain predicted by the optimal history-matched and random history-matched model and actual information gain of participants' questions across questions asked, displayed by order of question and age group. The numbers over the *x*-axis represent the number of participants that asked the corresponding number of questions. Error bars represent 1 *SEM* in each direction.



differences in the number of additional questions asked by 7- and 10-year-olds (p = .448), nor by 10-year-olds and adults (p = .396).

One possibility is that children asked unnecessary questions in the hope of receiving positive feedback, which could itself be rewarding. Counter to this hypothesis, the majority of unnecessary questions received *negative* feedback. However, positive feedback was more likely for the younger children. A univariate ANOVA revealed an effect of age group on the percentage of unnecessary questions that received positive feedback, $F(2, 55) = 9.18$, $p < .001$, $\eta^2 = .272$. A Bonferroni-corrected multiple-comparisons analysis confirmed that, on average, a higher percentage of the unnecessary questions asked by 7-year-olds received positive feedback (*M*7-year-olds = 30.00%, *SE* = 4.43%) as compared to 10-year-olds (*M*10-year-olds = 2.36%, *SE* = 7.54%) and adults (*M*adults = 1.18%, *SE* = 5.78%).

Sorting task. Most participants successfully completed the sorting task without help (7-year-olds: 84%; 10-year-olds: 88%; adults: 97%). With the exception of three younger children (12%), the remaining participants were all able to complete the sorting task with help (i.e., after being prompted at least once by the experimenter to reorganize the objects in a different way). We found no correlation between performance in the sorting task and performance in the question-asking task, and our results are not

Table 2
Percentage of Questions Asked (Study 1) or Objects Selected (Study 2) With Information Gain Zero Across All Three Trials, Displayed by Study, Age Group, and Question/Object Number

		Block															
Study	Age group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Study 1: question-asking	7-year-olds	0	0	9	26	43	71	73	100	100	50	50	100	100	100	100	100
	10-year-olds	0	0	0	10	40	56	71	100	100	100	100	100	100	100	100	100
	Adults	0	0	22	26	46	25	100	100								
Study 2: object-selection	7-year-olds	0	0	0	13	26	13	60	65	44	86	82	100	82	100	100	100
	10-year-olds	0	4	8	0	14	37	60	64	67	63	100	67	100	100	100	100
	Adults	0	0	0	17	11	19	38	50	50							

affected by excluding from all analyses those participants who could not complete the sorting task.

Discussion

Study 1 found an anticipated developmental change in the efficiency of information search: 7-year-olds required significantly more questions to solve the task than adults did, with 10-year-olds falling in between. However, our findings and analyses go beyond prior work in several ways. First, we found that children were able to solve a hierarchical version of the 20-questions task, which itself requires that children entertain hypotheses beyond the level of individual objects. Moreover, we found that children's performance fell short of optimality, but was quantifiably better than chance. Second, we found mixed support for the idea that one source of children's inefficiency is an inability to operate beyond the level of individual hypotheses to consider questions that more efficiently partition the hypothesis space by subsuming multiple hypotheses. On the one hand, we did find that adults were significantly more likely than children to begin by asking questions at the superordinate level, and that children were correspondingly more likely to ask initial questions at the subordinate level. This supports the idea that adults are better at appreciating the relationships between hypotheses and exploiting the hierarchical structure of our task, and could reflect both differences in search strategies and potentially increasing familiarity with 20-questions outside the lab, where queries often begin at high levels (e.g., "animal, vegetable, or mineral?"). On the other hand, we failed to find an interaction between age group and solution level, suggesting that adults' tendency to ask more superordinate questions did not translate into a greater advantage in reaching higher-level solutions.

Finally, we found that children not only asked questions that were less informative than those of adults, but also that children more often asked questions that were unnecessary (in the sense that they yielded no information at all). Specifically, children failed to reliably offer a solution at the point a single hypothesis remained, instead asking additional questions. One interpretation is that children have a more conservative stopping rule than adults, or that they are less accurate in assessing when the conditions for stopping apply. In the General Discussion section we return to these results in light of the findings from Study 2.

Study 2: Object Selection

In Study 2, participants completed the hierarchical 20-questions task by selecting individual objects to "test." After an object was selected for testing, participants observed whether or not it possessed the relevant property. We again modeled optimal and random performance using our formal analysis. This task and the corresponding analyses allowed us to further evaluate our three hypotheses about potential sources of inefficiency in children's information search. First, if children's inefficiency is related to difficulty going beyond the object level to identify higher-order features to target in generating questions, then we would expect to erase (or at least attenuate) the developmental differences in efficiency found in Study 1 (question asking) when it comes to testing individual objects, as this mode of inquiry effectively forces all participants to test their hypotheses via individual objects. Second, as in Study 1, if children's inefficiency is related to the challenge of going beyond individual hypotheses to make queries that target multiple hypotheses at once, then we would expect children to fail to identify the solution with fewer questions when it is at a higher level (superordinate vs. basic, and basic vs. subordinate). Finally, if children's inefficiency stems from a suboptimal stopping rule, then we would expect to see children continuing to test objects beyond the point at which a single hypothesis remains, as we observed with question asking in Study 1.

Method

Participants. Participants in Study 2 were 22 children in second grade (7 female, M_{age} = 90.0 months; *SD* = 6.2 months) and 23 children in fifth grade (11 female, M_{age} = 119.6 months; *SD* = 11.7 months), recruited from a primary school and a local children's museum, as well as 22 university students (16 female, M_{age} = 23.8 years; *SD* = 5.7 years). Although most participants were from White, middle-class backgrounds, a range of ethnicities reflecting the diversity of the population were represented. None of the participants in Study 2 had participated in Study 1.

Task, materials, and procedure. The task, materials, and procedure used in Study 2 were identical to Study 1 (including the familiarization phase and the sorting task), with one crucial procedural difference: instead of asking yes-or-no questions as in Study 1, participants could only test individual objects to find out which objects shared a novel property (e.g., being found on planet Apres). After each selection, participants received feedback from the experimenter with a response of "yes" or "no" and were prompted to put a red ("no" feedback) or green ("yes" feedback) frame around the object selected by touching the object on the tablet. Any errors were corrected by the experimenter.

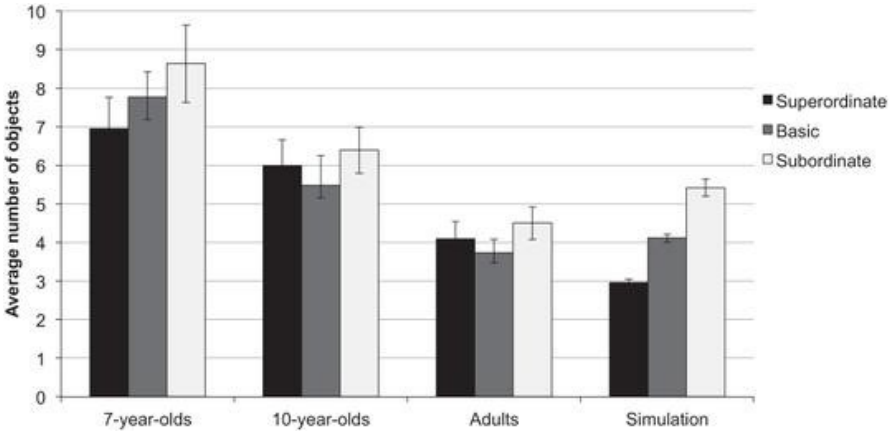
Simulations. As in Study 1, we simulated a total of 10,000 step-optimal search trajectories for the object-selection paradigm. A step-optimal search trajectory is one that selects, at each step, the object with the highest expected information gain (Figure 5). For example, suppose that a participant starts by selecting the blue owl (*IG* = 0.75). If the feedback is "yes," then the solution could be "only owls are on planet Apres," "all birds are on planet Apres," or "all animals are on planet Apres."[2]

In this case, the most informative follow-up selection would be one of the remaining animals (*IG* = .98). Selecting the brown owl or one of the plants would not provide any information (*IG* = 0). If the feedback is "no," then the solution could not be "owls," "birds," or "animals." In this case, the most informative follow-up selection would be one of the plants (*IG* = 0.85): Selecting the other owl would provide no information (*IG* = 0), and selecting one of the other animals would achieve lower information gain (fish: *IG* = 0.69; parrot: *IG* = 0.44; see Figure 1, bottom panel).

---

[2]Because we told participants that the solution included more than one of the given objects, they knew the solution could not be "only the blue owl is on planet Apres."

Figure 5. Study 2: object selection. Average number of objects selected before offering the solution, displayed by solution condition. Error bars represent 1 *SEM* in each direction.



Results

Simulation results for optimal learners. The optimal object-selection strategy reached the solution, on average, with 4.17 selections. This was significantly more than the 2.85 questions needed, on average, in the question-asking paradigm, $t(598) = 4.295$, $p < .001$. Again, this number varied with the hierarchical level of the solution (see Figure 5), $F(2, 299) = 137.66$, $p < .001$, $\eta^2 = .481$. As with question asking, a Bonferroni-corrected multiple-comparisons analysis revealed that fewer questions were needed when the solution was at the superordinate level (*M*superordinate = 2.96, *SE* = 0.09) than at the basic level (*M*basic = 4.12, *SE* = 0.09, $p < .001$), and that fewer questions were needed when the solution was at the basic level than at the subordinate level (*M*subordinate = 5.42, *SE* = 0.22, $p < .001$; see Figure 5).

Empirical results for human learners. Overall, participants required an average of 5.95 selections to reach the solution. Adults (*M*adults = 4.11, *SE* = 0.44) required no more selections than the optimal model's 4.17, $p = .773$.

To analyze the overall efficiency of participants' object-selection choices, we performed a mixed ANOVA with the number of objects selected prior to providing the correct solution as the dependent variable, age group (3: 7-year-olds, 10-year-olds, adults) as a between-subjects variable, and solution level (3: superordinate, basic, subordinate) and trial number (3) as within-subject variables. This analysis revealed a main effect of age group, $F(2, 64) = 21.16$, $p < .001$, $\eta^2 = .398$ (see Figure 5), with fewer objects tested with increasing age. A Bonferroni-corrected multiple-comparisons analysis confirmed that 7-year-olds selected more objects (*M*7-year-olds = 7.79, *SE* = 0.44) than 10-year-olds (*M*10-year-olds = 5.96, *SE* = 0.43, $p = .013$), and 10-year-olds more than adults (*M*adults = 4.11, *SE* = 0.44, $p = .012$). Notably, we did not find significant effects of solution level ($p = .175$), scenario, or trial; nor were there any interactions.

Participants' performance versus optimal history-matched and random history-matched models. As in Study 1, we compared the expected information gain of participants' object selections with two additional models: the optimal history-matched and random history-matched models. Our analysis showed that participants' average information gain (*M*participants = 0.60, *SE* = 0.02) was higher than the information gain resulting from a random selection (*M*random = 0.45, *SE* = 0.01) but lower than that resulting from the optimal history-matched model (*M*optimal = 0.76, *SE* = 0.01), $F(2, 128) = 429.49$, $p < .001$, $\eta^2 = .87$. The analysis also revealed a main effect of age group, $F(2, 64) = 12.74$, $p < .001$, $\eta^2 = .29$, but no interactions.

As with question asking, we found evidence for two sources of inefficiency (Figure 6). For the first several object selections—until the hypothesis space was narrowed down to a single hypothesis—the average information gain associated with participants' choices fell below that of the optimal history-matched model (see also Table 2). But additionally, some participants continued to select objects beyond the point at which a single hypothesis remained. Overall, 86% of the 7-year-olds ($n = 19$), 87% of the 10-year-olds ($n = 20$), but only 48% of the adults ($n = 10$) selected, in at least one trial, more objects than they needed to offer the solution. A repeated-measures ANOVA showed an age group effect on the number of objects selected beyond the point at which a single hypothesis remained, $F(2, 64) = 10.61$, $p < .001$, $\eta^2 = .249$. A Bonferroni-corrected multiple-comparisons analysis confirmed that the 7- and 10-year-olds tested on average more additional objects (*M*7-year-olds = 2.23, *SE* = 0.33; *M*10-year-olds = 1.17, *SE* = 0.32) than adults (*M*adults = 0.09, *SE* = 0.33, $p < .001$).

Again, we considered whether younger children were selecting additional objects merely to obtain reinforcing positive feedback. We found that a majority of questions received negative feedback, with no effect of age (*M*7-year-olds = 44.26%, *SE* = 8.58%; *M*10-year-olds = 49.62%, *SE* = 8.80%; *M*adults = 29.18%, *SE* = 15.66%; $p = .528$).

Sorting task. All participants were able to complete the sorting task. However, some participants (36% of the 7-year-olds and 30% of the 10-year-olds) had to be prompted by the experimenter at least once during the task to reorganize the objects in a different way in order to produce the hierarchical categorization expected. As in Study 1, there was no relationship between performance on the sorting task and on the object-selection task.

Discussion

Study 2 reinforced several findings from Study 1. First, we found expected developmental changes in the efficiency of information search, with an increase in efficiency (marked by a decrease in the number of objects tested) across development. We also found that children's performance fell short of the optimal learner's, but exceeded that of a learner who made random selections.

With regard to our key hypotheses, we failed to find support for the idea that forcing all learners to generate queries at the object level would erase or attenuate developmental differences. In fact, the effect of age was *greater* in Study 2 ($\eta^2 = .398$) than in Study 1 ($\eta^2 = .136$). This challenges the idea that children's inefficiency in question asking stems largely from difficulty going beyond the level of individual objects in making queries, since adults still performed much better when similarly constrained. However, we also failed to find support for the idea that adults surpass children by better exploiting the hierarchical structure of the hypothesis space. The simulations confirmed that higher-level solutions (e.g., superordinate vs. basic) should have been found more efficiently, but *neither* children nor adults exhibited this pattern.

Finally, we did find support for our third hypothesis: that children differ from adults in their stopping rule. As in Study 1, children were considerably more likely than adults to continue to

select objects beyond the point at which a single hypothesis remained, and thus beyond the point at which they should have been able to provide the solution.

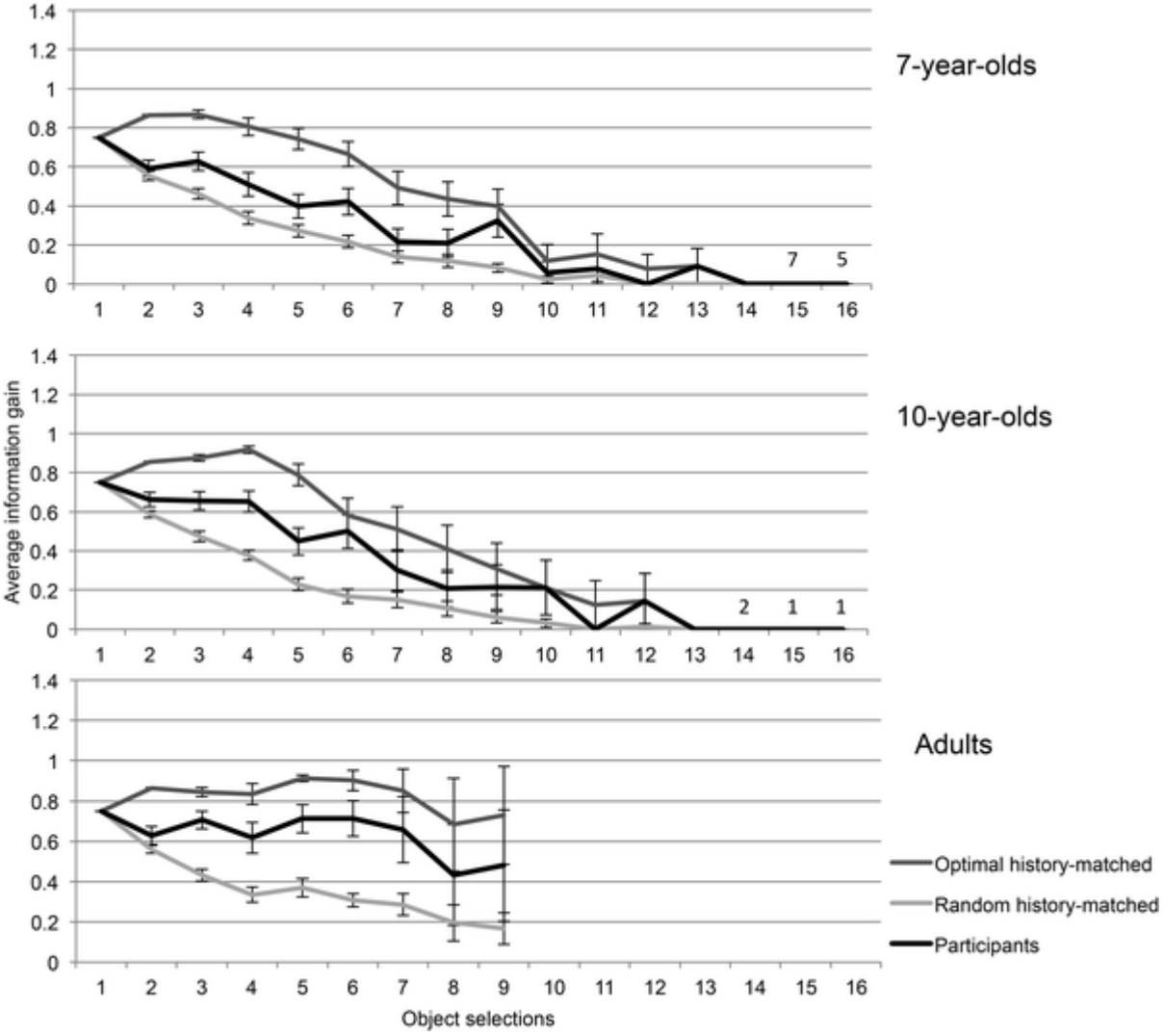


Figure 6. Study 2: object-selection. Average information gain predicted by the optimal history-matched and random history-matched model and actual information gain of participants' object-selection across queries, displayed by order of objects selected and age group. The numbers over the participants' line (*x*-axis) represent the number of participants that selected that many objects. Error bars represent 1 *SEM* in each direction.

General Discussion

Across two studies we investigated potential sources of developmental change in the efficiency of information search. In particular, we developed a novel task involving learning over a hierarchically structured domain, investigated information search through both question asking and object selection, and provided corresponding formal analyses.

Overall, our findings revealed the expected developmental shift in the efficiency of information search, with a progression from age 7 to age 10 to adulthood for both question asking and object selection. The finding that children's questions and object selections were less efficient than those of adults replicates prior research (Davidson, 1991a, 1991b; Mosher & Hornsby, 1966; Ruggeri & Feufel, 2015). But what drives this developmental change in efficiency? We return to the three hypotheses put forth in the introduction.

First, can children's inefficiency relative to adults be explained by adults' greater ability to go beyond the object level when seeking information? If adults' superior performance stems from the ability to go beyond individual objects to generate higher-level questions (Ruggeri & Feufel, 2015), then one would expect a larger developmental change in the efficiency of information search in Study 1 (question asking) than in Study 2 (object selection), which required that all participants make queries at the object level. Instead we found the reverse, with the effect size associated with age group in the object-selection paradigm (η^2 = .398) larger than that in the question-asking paradigm (η^2 = .136). With these relatively familiar categories and explicit instructions to identify which "*kinds* of objects" had a target feature, children appeared to be as capable of going beyond the object level as adults.

A second hypothesis is that adults outperform children because they are better able to go beyond individual hypotheses to consider the relationships between hypotheses, thereby exploiting the hierarchical structure of the task to ask questions or select objects that better partition the hypothesis space. Consistent with this idea, adults were significantly more likely than children to begin by asking questions at the most informative superordinate level (Ruggeri & Feufel, 2015). However, our optimal learning models also revealed that capitalizing on the structure of the hypothesis space

would result in fewer steps to reach solutions at the superordinate level, followed by the basic level, followed by the subordinate level—both for question asking and for object selection. We found this effect of solution level for question asking, but not for object selection. In both cases, however, there was no interaction between solution level and age group. This suggests that *both* children and adults were able to capitalize on hierarchical structure to ask efficient questions, but that neither children nor adults were able to do so when it came to object selection.

These findings challenge the second hypothesis, but some alternative explanations are worth considering. For instance, it could be that participants did not have uniform expectations as to the category-level (superordinate, basic, or subordinate) of the solution. In particular, they might have assumed that a solution at a superordinate-level was less likely given the task of selecting *individual* objects versus asking questions. Consistent with this idea, adults actually outperformed the optimal learner in Study 2 when the solution was at the subordinate level, suggesting that they may have optimized their search for this outcome, paying a cost in other conditions. Future research should investigate whether and how the types of actions available to the learner (e.g., asking questions vs. selecting objects) influence the expectations a learner brings to a task.

Finally, we found compelling evidence for a third candidate source of children's inefficiency: a tendency to ask unnecessary questions and to make unnecessary object selections. In both paradigms, we found that children were significantly more likely than adults to continue their search beyond the point at which a single hypothesis remained, and thus to ask questions and select objects associated with zero information gain. One interpretation of these results is that children's stopping rule differed from that of adults. In particular, children might have sought confirming evidence even when it was not strictly informative. This interpretation is supported by comments made by the children as they selected additional objects (e.g., "I think I know, but let me ask just one more question, to be sure"). Although these results are surprising in light of previous research showing that children of this age tend to be *over*confident (e.g., Finn & Metcalfe, 2014; Salles, Semelman, Sigman, & Calero, 2016), they are consistent with research on children's decision making, which finds that younger children tend to be more exhaustive in their search than older children (Davidson, 1991a, 1991b). Moreover, Legare and colleagues (2013) found that the preschoolers (4- to 6-year-olds) who asked more confirmatory questions (i.e., redundant based on the information already gathered) performed better overall in a 20-questions game, suggesting that the confirmatory strategy, although inefficient, may be useful. Looking for confirming evidence could indeed make sense when there is uncertainty about the hypothesis space, the feedback one has received, or the constancy of what is being learned. As novice learners in a noisy world, children might do well to err on the side of obtaining extra feedback.

Other interpretations of children's "unnecessary" questions and object selections are possible. First, children might have entertained more hypotheses than those considered in our model's hypothesis space. For example, they might have considered disjunctive hypotheses (e.g., a desk *or* a high chair can have the target property). This possibility deserves further study, but it is worth noting that the children never spontaneously offered such hypotheses as solutions during the search task. Second, children may have sought the rewarding experience of confirming an identified hypothesis (Nickerson, 1998), independently of whether the feedback was positive ("yes") or negative ("no"). Finally, children might have formed a preference for a particular hypothesis early on in the task, and consequently required more evidence before abandoning or revising it. Future work could ask participants to report what they think the most likely hypothesis is at each step of the search process, which could not only verify our assumptions about the hypothesis space entertained at the outset, but also shed light on how hypotheses are revised in response to feedback (Bonawitz, Denison, Gopnik, & Griffiths, 2014). Additionally, asking participants to express their *confidence* in the hypotheses entertained at each stage of the search process could provide insights into why children are more exhaustive in their search, and might account for some of the individual differences within age groups.

While we did find some variation in performance across our two information search paradigms—question-asking versus object-selection—the similarities are more striking than the differences. In particular, both studies support the conclusion that children differ from adults in their pursuit of questions and queries beyond the point at which the solution should be known. This similarity is especially notable in light of the differences between these search paradigms. The question-asking paradigm focuses on children's ability to acquire information from informants (Einav & Robinson, 2011; Mills et al., 2010; Mills, Legare, Grant, & Landrum, 2011). In contrast, the object-selection paradigm focuses on children's ability to explore and generate their own data (Cook, Goodman, & Schulz, 2011; Legare, 2012; Schulz & Bonawitz, 2007; Sim & Xu, 2014).

Despite the similarities observed here, future work should systematically investigate how information search strategies differ as a function of the method for obtaining information, and how people engage in the metareasoning task of selecting the best search strategy. For instance, we found that both children and adults learned more efficiently by asking questions than by testing individual objects; however, in many real-world contexts, asking questions might not be possible or desirable. First, knowledgeable informants might not be available. Second, to ask efficient questions one needs to categorize and name groups of objects (e.g., birds) or target them by generating shared features (e.g., animals that fly). This is particularly challenging in unfamiliar domains, especially for children (Ruggeri & Feufel, 2015). Third, asking questions might be less effective when the hierarchical structure does not match the true causal structure of what is being learned. In those cases, a bottom-up object-selection strategy might be easier to implement and less likely to mislead. In particular, asking questions based on a learner's prior category knowledge could hinder opportunities for serendipitous learning, whereas intervening in the world by trying out one object at a time could provide evidence that overrides existing category representations and gives rise to new categories.

Prior work has identified several developmental changes within our age range that could help explain changes in the efficiency of information search, including cognitive flexibility (Legare et al., 2013), the ability to identify and generate abstract features (Ruggeri & Feufel, 2015), the ability to statistically analyze the hypothesis space and recognize the most effective features to target (Nelson et al., 2014; Ruggeri, Sim, & Xu, 2016), and vocabulary

(Ruggeri, Walker, Lombrozo, & Gopnik, 2016). It is notable, however, that these changes can only partially account for the shifts we document in children's performance. In particular, they are unlikely to explain children's tendency to make uninformative queries. More research is needed to investigate whether and how developmental changes in working memory capacity (Katz, Bereby-Meyer, Assor, & Danziger, 2010), planning abilities (e.g., the ability to plan one step ahead; see Völter & Call, 2014), and metacognitive control underlie the efficiency of information search across the life span and across individuals. Additionally, children and adults might have different ways to operationalize a query's informativeness (e.g., optimizing vs. satisficing), and this operationalization could be influenced by the different costs (e.g., time and effort) incurred by children and adults when implementing a given strategy, as well as by their motivation.

In sum, our studies replicate prior work in documenting a developmental shift in the efficiency of information search, and additionally compare two information search paradigms and offer a quantitative analysis of performance. We show that even 7-year-olds perform reliably better than a random learner, and that even adults tend to fall short of optimality. We also evaluate three hypotheses about the sources of children's inefficiency, and succeed in identifying an important but previously undocumented developmental change: over development, children become less likely to make queries beyond the point at which the solution should be known. Our findings thus raise new and important questions about children's decisions and information search strategies.

Investigating the "active child" is one of the enduring themes in the study of developmental psychology. Asking questions and exploration are powerful tools through which children can actively learn about the changing world around them, testing and revising their theories. They are therefore invaluable sources of insight into *how* children learn. Previous research has not investigated the development of information search strategies computationally, nor compared different information search paradigms. Doing so paves the way for a more formal investigation of children's active learning toolbox, including the various information search strategies available to them, how efficiently they are implemented at different developmental stages, and the metastrategies employed in selecting among them. Our tasks and analyses thus provide initial steps toward a more formal approach to understanding hypothesis testing and learning from middle childhood to adulthood and suggest promising directions for further investigation of information search throughout the life span.

References

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-yearolds spontaneously use others' past performance to guide their learning. *Cognition, 107,* 1018–1034. http://dx.doi.org/10.1016/j.cognition.2007.12.008

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology, 74,* 35–65. http://dx.doi.org/10.1016/j.cogpsych.2014.06.003

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64,* 215–234. http://dx.doi.org/10.1016/j.cogpsych.2011.12.002

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

Campos, J. J., & Stenberg, C. R. (1981). Perception, appraisal, and emotion: The onset of social referencing. In M. E. Lamb & L. R. Sherrod (Eds.), *Infants' social cognition: Empirical and social considerations* (pp. 273–314). Hillsdale, NJ: Erlbaum.

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development, 72,* vii–ix.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition, 120,* 341–349. http://dx.doi.org/10.1016/j.cognition.2011.03.003

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science, 12,* 426–437. http://dx.doi.org/10.1111/j.1467-7687.2008.00792.x

Davidson, D. (1991a). Children's decision making examined with an information-board procedure. *Cognitive Development, 6,* 77–90. http:// dx.doi.org/10.1016/0885-2014(91)90007-Z

Davidson, D. (1991b). Developmental differences in children's search of predecisional information. *Journal of Experimental Child Psychology, 52,* 239–255. http://dx.doi.org/10.1016/0022-0965(91)90061-V

Davidson, D. (1996). The effects of decision characteristics on children's selective search of predecisional information. *Acta Psychologica, 92,* 263–281. http://dx.doi.org/10.1016/0001-6918(95)00014-3

Einav, S., & Robinson, E. J. (2011). When being right is not enough: Four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological Science, 22,* 1250–1253. http://dx.doi .org/10.1177/0956797611416998

Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32,* 1–9. http://dx.doi .org/10.1016/j.learninstruc.2014.01.001

Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult-child conversation. *Child Development, 80,* 1592–611. http://dx.doi.org/10.1111/j.14678624.2009.01356.x

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press. Gregan-Paxton, J., & John, D. R. (1995). Are young children adaptive decision makers? A study of age differences in information search behavior.

Journal of Consumer Research, 21, 567–580. http://dx.doi .org/10.1086/209419

Gregan-Paxton, J., & John, D. R. (1997). The emergence of adaptive decision making in children. *Journal of Consumer Research, 24,* 43–56. http://dx.doi.org/10.1086/209492

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

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are underinformative. *Cognition, 132,* 335–341. http://dx.doi.org/10.1016/j.cognition.2014.04.013

Herwig, J. E. (1982). Effects of age, stimuli, and category recognition factors in children's inquiry behavior. *Journal of Experimental Child Psychology, 33,* 196–206. http://dx.doi.org/10.1016/0022-0965(82) 90015-7

Howse, R. B., Best, D. L., & Stone, E. R. (2003). Children's decision making: The effects of training, reinforcement, and memory aids. *Cognitive Development, 18,* 247–268. http://dx.doi.org/10.1016/S08852014(03)00023-6

Katz, I., Bereby-Meyer, Y., Assor, A., & Danziger, S. (2010). Children's adaptive pre-decisional search behavior: Effects of memory and number

of alternatives. *Journal of Economic Psychology, 31,* 17–24. http://dx.doi.org/10.1016/j.joep.2009.09.003

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE, 7,* e36399. http://dx.doi.org/10 .1371/journal.pone.0036399

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261–1277. http://dx.doi.org/10.1111/j.1467-8624.2005.00849.x

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development, 83,* 173–185. http://dx.doi.org/10.1111/j.1467-8624.2011.01691.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. http://dx.doi.org/10.1016/j.jecp.2012.07.002

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences, 20,* 748–759. http://dx.doi.org/10.1016/j.tics.2016.08.001

Mata, R., von Helversen, B., & Rieskamp, J. (2011). When easy comes hard: The development of adaptive strategy selection. *Child Development, 82,* 687–700. http://dx.doi.org/10.1111/j.1467-8624.2010.01535.x

Meltzoff, A. N. (1988a). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24,* 470–476. http://dx.doi.org/10.1037/0012-1649.24.4.470

Meltzoff, A. N. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59,* 217–225. http://dx.doi.org/10.2307/1130404

Meltzoff, A. N. (1990). Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 139– 167). Chicago, Il: University of Chicago Press.

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,* 533–560. http://dx.doi.org/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,* 539–560. http://dx.doi.org/10.1016/j.jecp.2011.06.003

Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. R. Olver, T. M. Greenfield, J. R. Hornsby, H. J. Kenney, M. Maccoby, . . . (Eds.), *Studies in cognitive growth* (pp. 86–102). New York, NY: Wiley.

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112,* 979–999. http://dx.doi.org/10.1037/0033-295X.112.4.979

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,* 74–80. http://dx.doi.org/10.1016/j.cognition.2013.09.007

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220. http://dx.doi.org/10.1037/1089-2680.2.2.175

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631. http://dx.doi.org/10.1037/0033-295X.101.4.608

Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Ballantine Books. http://dx.doi.org/10.1037/11168-000

Ruggeri, A., & Feufel, M. A. (2015). How basic-level objects facilitate question-asking in a categorization task. *Frontiers in Psychology, 6(July),* 918. http://dx.doi.org/10.3389/fpsyg.2015.00918

Ruggeri, A., & Katsikopoulos, K. V. (2013). Make your own kinds of cues: When children make more accurate inferences than adults. *Journal of Experimental Child Psychology, 115,* 517–535. http://dx.doi.org/10.1016/j.jecp.2012.11.007

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

Ruggeri, A., Sim, Z. L., & Xu, F. (2016). *Five-year-old children identify the most informative questions*. [Manuscript submitted for publication].

Ruggeri, A., Walker, C., Lombrozo, T. & Gopnik, A. (2016). *Strategic use of questions to solve a causal inference task in early childhood*. [Manuscript in preparation].

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72,* 1054–1070. http://dx.doi.org/10.1111/1467-8624.00334

Salles, A., Ais, J., Semelman, M., Sigman, M., & Calero, C. I. (2016). The metacognitive abilities of children and adults. *Cognitive Development, 40,* 101–110. http://dx.doi.org/10.1016/j.cogdev.2016.08.009

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43,* 1045–1050. http://dx.doi.org/10.1037/0012-1649.43.4.1045

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10,* 322–332. http://dx.doi.org/10.1111/j.1467-7687.2007.00587.x

Sim, Z., & Xu, F. (2014). Acquiring inductive constraints from selfgenerated evidence. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81,* 669–679. http://dx.doi.org/10.1111/j.1467-8624.2009.01422.x

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. http://dx.doi.org/10.1017/S0140525X01000061

Tomasello, M. (1999a). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (1999b). The human adaptation for culture. *Annual Review of Anthropology, 28,* 509–529. http://dx.doi.org/10.1146/annurev.anthro.28.1.509

Völter, C. J., & Call, J. (2014). Younger apes and human children plan their moves in a maze task. *Cognition, 130,* 186–203. http://dx.doi.org/10 .1016/j.cognition.2013.10.007

Originally published in: Developmental Psychology, 52(12), 2016, p. 2173