Journal of Experimental Child Psychology 110 (2011) 499–519

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

It's in the sample: The effects of sample size and sample diversity on the breadth of inductive generalization

Chris A. Lawson [a,*], Anna V. Fisher [b]

[a] Department of Psychology, Saint Joseph's University, Philadelphia, PA 19131, USA
[b] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 13 April 2010
Revised 1 July 2011
Available online 3 August 2011

Keywords:
Generalization
Reasoning
Induction
Inference
Sample size
Categorization

ABSTRACT

Developmental studies have provided mixed evidence with regard to the question of whether children consider sample size and sample diversity in their inductive generalizations. Results from four experiments with 105 undergraduates, 105 school-age children (M = 7.2 years), and 105 preschoolers (M = 4.9 years) showed that preschoolers made a higher rate of projections from large samples than from small samples when samples were diverse (Experiments 1 and 3) but not when samples were homogeneous (Experiment 4) and not when the task required a choice between two samples (Experiment 2). Furthermore, when a property occurred in large and diverse samples, preschoolers exhibited a broad pattern of projection, generalizing the property to items from categories not represented in the evidence. In contrast, adults followed a normative pattern of induction and never attributed properties to items from categories not represented in the evidence. School-age children showed a mixed pattern of results.

Introduction

From an early age, children are able to generate inductive inferences about the properties of novel objects. Given evidence about a single *bird* with "omentum inside", 4-year-olds readily predict that other *birds* also have "omentum inside" (Gelman, 1988; see Hayes, 2007, for a review). Children generalize properties from a known exemplar (e.g., *bird*) to novel cases (e.g., other *birds*, *animals*) if the exemplars share taxonomic relations (Gelman, 2003) or are perceptually similar (Sloutsky & Fisher, 2004). It has also been shown that not all properties are equally generalizable (Carey, 1985; Gelman,

* Corresponding author. Fax: +1 610 660 1803.
E-mail address: clawson@sju.edu (C.A. Lawson).

0022-0965/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2011.07.001

1988; Keil, Smith, Simons, & Levin, 1998). For example, children are willing to generalize the property "has important internal parts" but not the property "is dusty" (Keil et al., 1998). This phenomenon entails two related questions. First, what is the basis on which children limit the generalizability of some properties (e.g., being dusty)? Second, what is the basis on which children decide that some properties can be generalized broadly (e.g., from a single *bird* to other *birds* and other *animals*)?

The first question has been recently addressed in studies demonstrating that the regularity with which a property occurs in a sample of evidence influences children's willingness to generalize this property (Lawson & Fisher, 2008). For example, Lawson and Fisher (2008) showed that 5-year-olds are willing to generalize properties that occur with high regularity (e.g., 70–100% of the time) but not properties that occur with low regularity (e.g., 30% of the time). These findings show that children can use statistical regularity of a property to infer which properties should *not* be generalized.

However, it is still unclear what information children use to infer that some properties can be generalized broadly. The current studies examined the possibility that children learn about the generalizability of properties by attending to the distribution of evidence in the input. There is ample evidence showing that young learners can successfully use distributional evidence in the available input in category and language learning tasks (Colunga & Smith, 2002; Rakison & Hahn, 2004; Rogers & McClelland, 2004; Samuelson, Schutte, & Horst, 2009; Sloutsky & Fisher, 2008). The current studies were designed to examine whether information about the distribution of a property within a sample would limit or broaden the scope of children's inductive projections.

Among other things, distributions of properties differ in (a) the diversity of exemplars for which a particular property can occur and (b) the number of instances for which a property has occurred. For instance, the property *has fins* occurs in a less diverse set of items than the property *has eyes*, and the number of animals that have fins is far smaller than the number of animals that have eyes. We tested the possibility that if a property has occurred with high regularity for a relatively large number of diverse items, children may infer that the property is highly common and, thus, can be generalized to a broad range of exemplars. In other words, the breadth of property projections may be influenced, among other factors, by sample size and sample diversity. In the remainder of this Introduction, we review the findings on sample size and sample diversity effects in adults and children and then describe a series of studies designed to explore the role of these statistical properties on children's and adults' inductive generalizations.

Sample size and sample diversity effects in adults

In a now classic set of studies, Osherson, Smith, Wilkie, Lopez, and Shafir (1990) asked adult participants to decide which of two samples of evidence provided the best support for a given conclusion. For instance, adults were asked to judge which sample of evidence provided better support for the conclusion "*gorillas* secrete uric acid crystals". The majority of participants chose the larger sample (e.g., a *fox*, a *pig*, and a *wolf*) over a smaller sample (e.g., a *pig* and a *wolf*) and chose the more diverse sample (e.g., *lions* and *giraffes*) over the less diverse sample (e.g., *lions* and *tigers*) (see also Heit, 1998; Lopez, 1995). Osherson and colleagues (1990) posited two converging processes to explain sample size and sample diversity effects in adults. First, solving these inductive problems involves the ability to generate the appropriate conclusion category that spans the evidence and conclusion exemplars (e.g., *mammal* in the above examples). The second process involves the ability to decide that one of the samples provides better coverage of the conclusion category. The implication from these results is that adults recognize that larger and more diverse samples provide better coverage of a conclusion category than smaller and less diverse samples.

Other researchers have described the preference for diverse evidence as a normative inductive strategy (Heit, 1998; Heit, Hahn, & Feeney, 2005; Lo, Sides, Rozelle, & Osherson, 2002). For example, Heit and colleagues (2005) provided several examples from the history and philosophy of science whereby it has been argued that diverse samples support greater precision in conclusions. Sample size effects have also been described as normative from the perspective that larger samples are preferable from the statistical law of large numbers (Sedlmeier & Gigerenzer, 1997). As a set of normative principles, sample size and sample diversity should increase inductive accuracy. In the context of category induction, a normative prediction is that more evidence about a set of exemplars

from the same category should elicit generalizations to only those exemplars from the category relative to smaller samples of evidence (Xu & Tenenbaum, 2007). In other words, a larger sample should provide better evidence than a smaller sample about the kind of instances that have the property.

Developmental evidence for sample size and sample diversity effects

Several studies have provided evidence that young children are insensitive to sample size and sample diversity until after 8 years of age. For example, Lopez, Gelman, Gutheil, and Smith (1992) presented 5- and 8-year-olds with evidence about a large sample (5 exemplars) of mammals for which they were told that a *raccoon*, *leopard*, *skunk*, *tiger*, and *giraffe* each had "property X" and about a small sample (3 exemplars) of mammals for which a *skunk*, *tiger*, and *giraffe* each had "property Y". Participants were then asked to predict whether a different exemplar from the mammal category (e.g., a bear) would have property X. Children were equally likely to project property X and property Y to the novel exemplar regardless of sample size (Gutheil & Gelman, 1997, provided similar results).

Research also indicates that children fail to consider sample diversity in their inductive generalizations prior to 7 years of age (Li, Cao, Li, Li, & Deák, 2009; Lopez et al., 1992; Rhodes, Gelman, & Brickman, 2008). Lopez and colleagues (1992) showed that 7-year-olds generalized a property to a novel *rabbit* whether they were given evidence about a diverse sample (e.g., a *lion* and *giraffe* had property Y) or a homogeneous sample of evidence (e.g., a *lion* and *tiger* had property X).

Several explanations for the lack of sample size and sample diversity effects in young children appeal to limitations in the ability to compare samples of evidence. For instance, Lopez and colleagues (1992) argued that children are unable to compute the category that is covered by the exemplars in both samples of evidence and the target exemplar. Others have argued that children's ability to compare samples is limited by a focus on within-category similarities (Gutheil & Gelman, 1997) or expectations about the typicality of members within each sample (Rhodes et al., 2008). Thus, it is possible that children's generalizations are not influenced by sample size and diversity information because they lack the cognitive skills necessary to represent the categories covered by the samples of evidence or because they are unable to effectively compare different samples of evidence.

However, some evidence indicates that young children are sensitive to sample size and sample diversity in the course of generalization. For example, Xu and Tenenbaum (2007) observed that 3-year-olds generalized novel object labels narrowly—only to the categories represented in the evidence sample (e.g., Dalmations)—when the evidence sample consisted of 3 exemplars (e.g., 3 Dalmatians); however, when the evidence sample consisted of only 1 exemplar, children and adults generalized novel labels more broadly—to the categories that were represented in the evidence sample as well as to the categories outside of the evidence sample (e.g., from 1 Dalmatian to other Dalmatians as well as to other dogs). In addition, Jacobs and Narloch (2001) found that children made higher estimates about the number of category members that would have a property after learning that 30 exemplars have the property than after learning that 1 or 3 exemplars have the property.

With regard to sample diversity, Rhodes, Gelman, and Brickman (2010) found that 5-years-olds generalized evidence about a diverse sample of dogs exclusively to other subordinates (e.g., Dalmatians) when the evidence was said to come from an expert but generalized more broadly (e.g., to Dalmatians and turtles) when the evidence came from a novice. Heit and Hahn (2001) showed that 5-year-olds predicted that an individual who demonstrated diverse preferences (e.g., played with a soccer ball, a baseball, and a football) was more likely to make a diverse range of choices (e.g., about which toys to play with) than was a person with a homogeneous set of preferences (e.g., always played with a soccer ball).

Overall, prior research yielded mixed results with regard to children's reliance on sample size and sample diversity information in the course of inductive generalization. This mixed pattern of results may stem from context sensitivity in children's induction. For instance, children's inferences were not influenced by sample size when they needed to choose between two samples (Gutheil & Gelman, 1997; Li et al., 2009; Lopez et al., 1992; Rhodes et al., 2008); at the same time, children were sensitive to sample size information when the task did not require making a comparison between two samples (Jacobs & Narloch, 2001; Xu & Tenenbaum, 2007). Therefore, it is possible that children may have

difficulty in evaluating and comparing two samples of evidence rather than being generally insensitive to sample size and sample diversity information.

Another possible reason for a mixed pattern of findings in prior research is the presence of pragmatic cues in a small subset of the studies (Rhodes et al., 2010; Xu & Tenenbaum, 2007). For example, children showed sensitivity to sample diversity when told that an expert supplied the sample of evidence (Rhodes et al., 2010). Children showed sensitivity to sample size information when the experimenter prefamiliarized children to all evidence exemplars before sampling them (Xu & Tenenbaum, 2007), thereby specifying beforehand all possible objects that children could encounter in the task. However, learning outside of laboratory is not always aided by pragmatic cues such as knowing all possible objects one can encounter and knowing that an expert supplied the evidence.

The current studies

The four experiments reported in this article were designed with the above issues in mind. In the current studies, we examined performance in tasks that necessitate comparisons between large and small samples (Experiment 2) and those that do not (Experiments 1, 3, and 4). The need to compare large and small samples was eliminated by manipulating sample size between participants such that one group of participants received evidence about a small sample ($n = 4$) and a different group of participants received evidence about a large sample ($n = 16$). Thus, if children's difficulty in using sample size information is due to their difficulty in comparing samples, this between-participants manipulation should yield sample size effects.

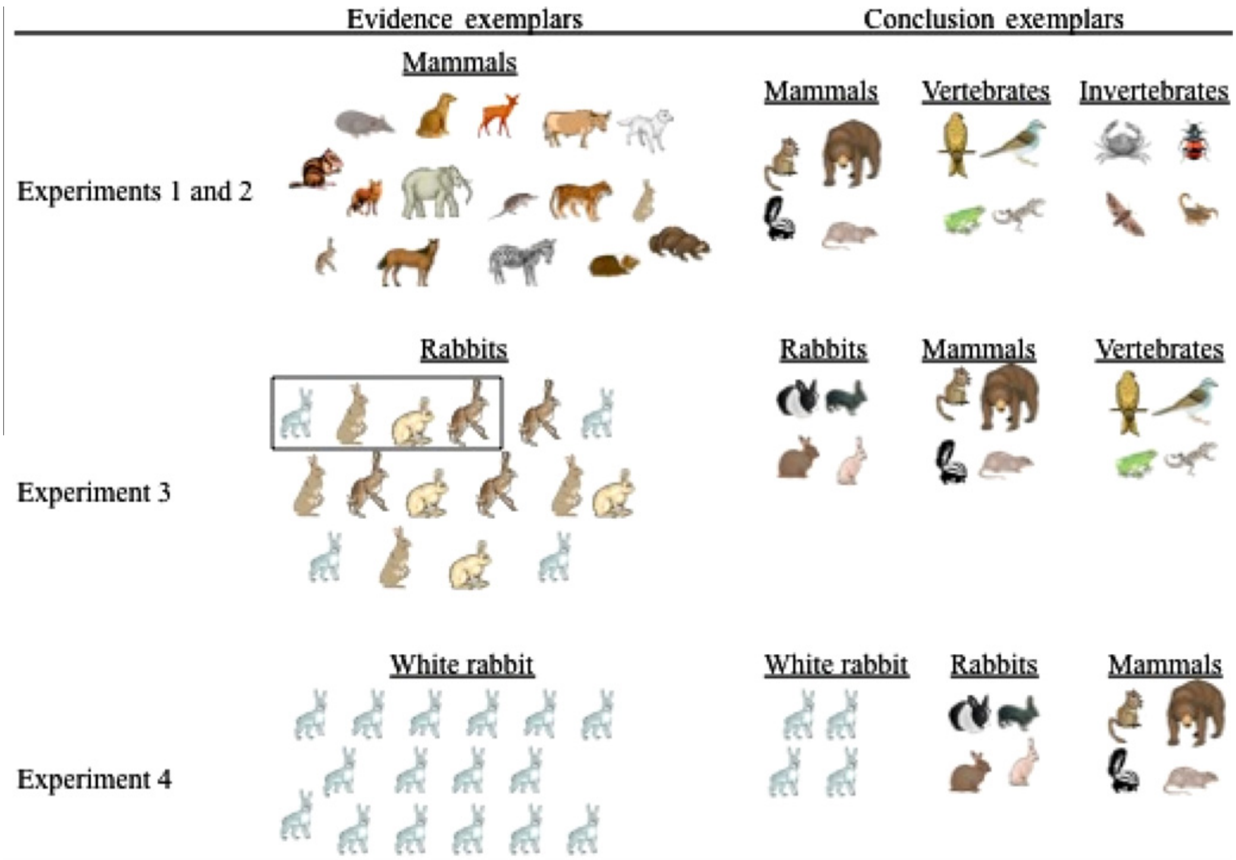
A second goal of these studies was to approximate learning in naturalistic contexts where a learner gradually accrues knowledge over repeated exposures to stimuli. To do so, we presented evidence exemplars one by one and told participants whether each animal had or did not have a property (akin to parents pointing out objects in the environment to their children and indicating that a particular item, for instance, is called a "cat" or "can meow") and presented exemplars serially (again, akin to children encountering objects in the world and making inferences about them). In addition, we presented evidence about novel properties attributed to most exemplars (75%) because probabilistic properties are more consistent with the stochastic nature of the input received by children outside of the lab (e.g., not every cat that a child encounters would meow) (Lawson & Fisher, 2008).

We manipulated sample diversity across the experiments such that participants were presented with a high diversity sample in Experiments 1 and 2, a medium diversity sample in Experiment 3, and a maximally homogeneous sample in Experiment 4 (see Table 1). We chose this approach to test the possibility that sample size effects, if present in young children, may be heightened for diverse samples.

To examine the breadth of generalization, we included a wide range of conclusion exemplars (i.e., mammals, birds, and invertebrates). We predicted that diverse samples would promote a broader range of generalizations than homogeneous samples. The basis for this prediction was twofold. First, because young children have difficulty in representing the category "covered" by the evidence in tasks that include multiple evidence exemplars (Lopez et al., 1992; Rhodes et al., 2008), children might overgeneralize properties common to diverse samples. That is, rather than following the normative pattern of generalizing only to those exemplars "covered" by the evidence category, children may infer that a particular property is broadly generalizable if the property occurs for a relatively large number of diverse items. Second, because our task involved a nonpragmatic learning context, our prediction is consistent with recent findings by Rhodes and colleagues (2010), who found that children generalized narrowly when pragmatic cues were available but generalized more broadly in the absence of pragmatic cues.

In sum, the current experiments were motivated by two goals. First, although there exists a large body of evidence on children's attention to statistical properties of available input in categorization and word learning (Colunga & Smith, 2002; Samuelson et al., 2009), less is known about how children use such properties in property induction. Furthermore, that which is known presents a mixed picture regarding children's use of statistical properties such as sample size and sample diversity. Thus, one goal of these experiments was to examine whether certain conditions support or hinder children's and adults' sensitivity to sample size and sample diversity in the course of property induction. The

Table 1
Design of Experiments 1 to 4.



Note. The same 4 exemplars were presented in the small and large sample conditions in Experiment 3. These 4 exemplars are highlighted by a black border around them.

second goal of these experiments was to explore the extent to which sample size and sample diversity influence the breadth of inductive generalizations children and adults are willing to endorse.

Experiment 1

Experiment 1 examined the influence of sample size on inductive generalizations for a diverse set of exemplars (see Table 1). Participants learned about a property that was associated with 75% of the exemplars within a sample of *mammals* (e.g., *rabbit*, *horse*, *elephant*) and then were asked to make inferences about different mammals (e.g., *cow*), exemplars drawn from a different category of vertebrates (e.g., *bird*), and exemplars that included only invertebrates (e.g., *insect*). Half of the participants were given evidence about a small sample ($n = 4$), and the other half learned about a large sample ($n = 16$).

Method

Participants

The participants were 30 adults (undergraduate students, 14 women and 16 men), 30 school-age children ($M = 7.5$ years, $SD = 0.48$, 13 girls and 17 boys), and 30 preschoolers ($M = 4.7$ years, $SD = 0.35$, 12 girls and 18 boys). In all of the experiments reported in this article, adults were recruited from psychology courses and received partial course credit for their participation, and children were recruited from local preschools and elementary schools and received a small prize for their participation. All participants were from a moderately large eastern US city. No participants were involved in other studies reported here.

Design and materials

Participants were randomly assigned to either the small sample condition or the large sample condition. In the small sample condition participants were presented with evidence about 4 mammals, whereas in the large sample condition participants were presented with evidence about 16 mammals. A set of 16 mammals was created such that all of the items were presented in the large sample condition and a subset of these items was randomly selected as the sample for each participant in the small sample condition (see Table 1). The set of mammals included an equal number of small and large mammals expected to be familiar to children.

Both conditions involved property attribution at a rate of 3:1 such that 75% of the instances involved the presence of the property and 25% of the instances involved the absence of the property. The presence/absence of properties was pseudo-randomized with the constraint that the absence of a property in the large sample condition did not occur for consecutive instances. Participants were shown a picture of the target and told, "This animal has [does not have] omat inside of it". Targets appeared with an arrow pointing toward the animal referring to a small red blotch when the property was present and with the arrow pointing to a blank area when the property was absent. As is typical in induction studies, properties were described as novel internal features that could be interpreted as biological.

After the introduction of evidence, participants were asked to make predictions about a range of target items. There were 4 instances from each of three target categories: *mammals* (e.g., cow), nonmammal *vertebrates* (e.g., bird), and *invertebrates* (e.g., crab). Participants in both sample size conditions were presented with the same set of 12 conclusion exemplars (see Table 1). For each conclusion exemplar, participants were shown a picture of the animal and asked, "Do you think this animal has omat inside of it?" After children responded, the next item was introduced. All target items were presented in random order.

Procedure

Children were interviewed in a quiet location at their school or day care center. The task was administered on a laptop and designed using PsyScope. Adult participants completed the task on individual computers in a laboratory on campus. Participants were told that the task involved two parts; in the first part they would learn something about an animal, and in the second part they would be asked some questions.

The game involved a scene with a tree that children were told they needed to click in order to see the animals. When participants clicked on the tree, a single exemplar emerged and they were told about the animal (or, during the projection phase, asked to predict the property of the animal). The method of selecting and discovering the animals was designed to sustain children's interest in the task and to ensure that the items were presented individually. The task lasted approximately 10 min.

Results

Responses to the items from each target category were collapsed to yield mean projections for each participant for the mammal, vertebrate, and invertebrate targets (see Fig. 1). The main analysis involved a mixed analysis of variance (ANOVA) with age (adults, school-age children, or preschoolers) and sample size (large or small) as between-participants variables and target (mammals, vertebrates, or invertebrates) as a within-participants variable. The analysis revealed a main effect of sample size, $F(1, 84) = 18.12$, $p < .0001$, $\eta^2 = .18$. Post hoc analyses in this and all other experiments reported in this article involved Scheffé's tests and were controlled for family-wise error. Post hoc analyses indicated that there was a higher rate of projections in the large sample condition ($M = .62$) than in the small sample condition ($M = .42$), $p < .0001$.

There was also a main effect of age, $F(2, 84) = 6.26$, $p < .0001$, $\eta^2 = .13$, due to a higher rate of projections by preschoolers ($M = .61$) than by adults ($M = .41$), $p < .01$. Both of these main effects were mediated by an age by sample size interaction, $F(2, 84) = 3.83$, $p < .05$, $\eta^2 = .08$, with simple effects revealing age differences only in the large sample condition, $F(2, 42) = 7.76$, $p < .01$, $\eta^2 = .27$. Follow-up analyses indicated that preschoolers were the only age group to show the sample size effect, exhib-

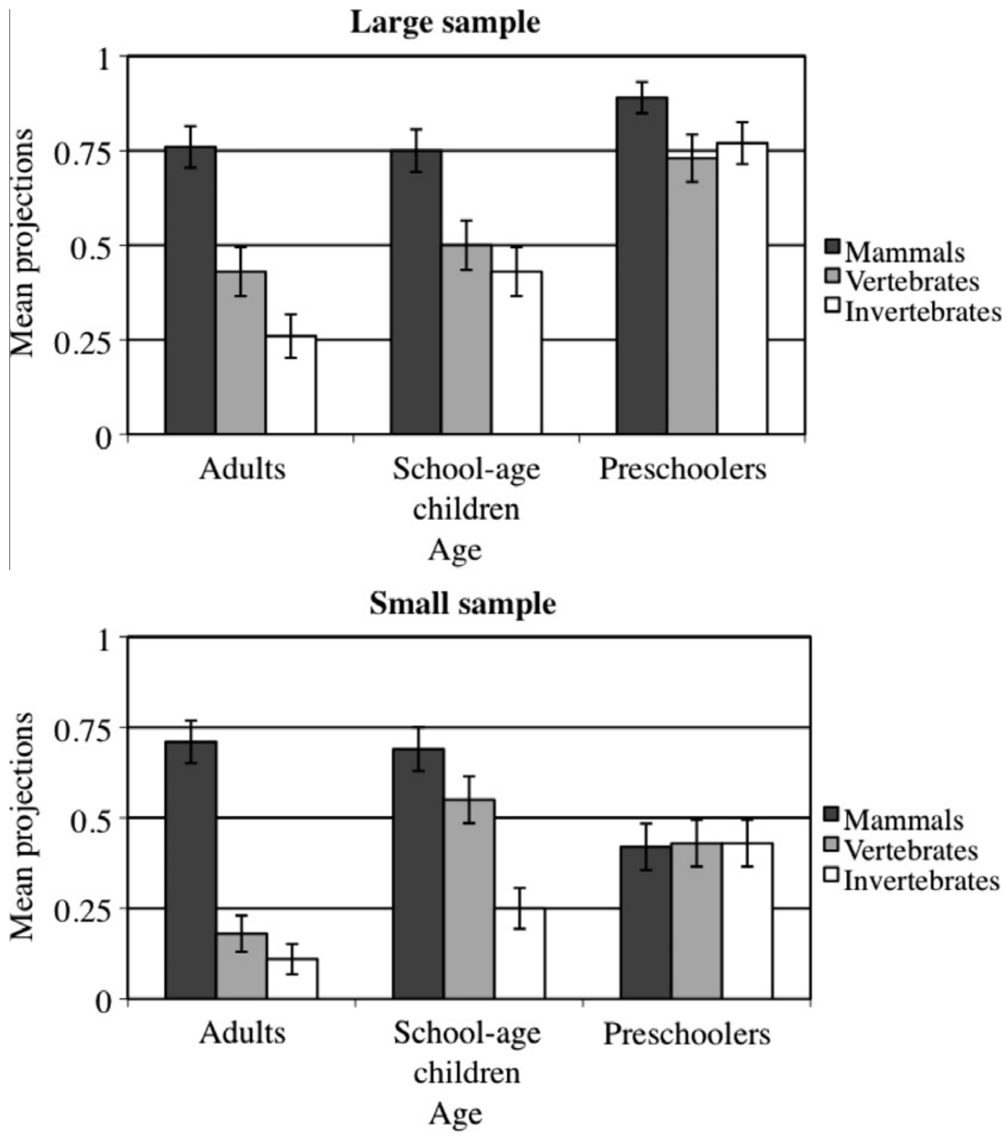


Fig. 1. Mean projections for each age group for each target in the large sample condition and in the small sample condition in Experiment 1. Bars indicate 1 standard error of the mean.

iting a higher rate of projections from large samples ($M = .79$) than from small samples ($M = .43$), $F(1, 28) = 12.81$, $p < .01$, $\eta^2 = .31$.

The main analysis also revealed an effect of target, $F(2, 164) = 59.05$, $p < .001$, and a target by age interaction, $F(4, 164) = 11.09$, $p < .001$, $\eta^2 = .40$. Separate analyses for each age group were conducted to test the hypothesis that sample size differentially affected the range of projections for older and younger participants. The analysis revealed that school-age children and adults showed the same target effect in both conditions, with rates of projections decreasing such that responses followed a pattern of projections for mammals > projections for vertebrates > projections for invertebrates, all *ps* < .005. In contrast, preschoolers showed no target effect, $F < 1$, $p > .25$, but did exhibit the sample size effect (projections from large samples > projections from small samples) for each of the three target types: mammals, $F(1, 28) = 14.33$, $p < .001$, $\eta^2 = .34$, vertebrates, $F(1, 28) = 6.51$, $p < .01$, $\eta^2 = .17$, and invertebrates, $F(1, 28) = 5.91$, $p < .01$, $\eta^2 = .17$.

The final analysis examined individual patterns of projections. Participants were considered to have followed a pattern of positive projections if they projected the property to at least 3 of 4 instances of each target type, a pattern of negative projections if they projected the property to 1 or 0 instances of each target type, and a pattern of random responding if they projected the property to 2 of 4 instances of each target type. Table 2 shows the number of participants who followed each

Table 2
Number of individuals following a pattern of positive projections (projecting to 3 or 4 instances of a target), random projections (projecting to 2 instances of a target), or negative projections (projecting to 1 or 0 instances of a target) in Experiment 1.

			Targets		
Age	Condition	Pattern of projections	Mammals	Vertebrates	Invertebrates
Adults	Large sample	Positive	11**	4	1
		Random	2	5	4
		Negative	2	6	10*
	Small sample	Positive	10*	1	1
		Random	4	2	2
		Negative	1	13**	12**
School-age children	Large sample	Positive	12**	6	4
		Random	1	4	4
		Negative	2	5	7
	Small sample	Positive	11**	7	2
		Random	2	4	3
		Negative	2	4	10*
Preschoolers	Large sample	Positive	13**	10*	11**
		Random	0	3	2
		Negative	2	2	2
	Small sample	Positive	6	3	7
		Random	2	5	1
		Negative	7	7	7

Note. A single asterisk (*) indicates that the number of individuals showing this pattern is greater than would be expected by chance at $p < .05$; a double asterisk (**) indicates a score that was different from chance at $p < .01$.

of these patterns. There were 15 individuals in each condition, and the probability of at least 10 of 15 participants following a pattern of positive or negative projections is $p < .02$, whereas the probability of at least 6 of 15 participants following a pattern of random responding is $p < .04$, binomial theorem.[1] The majority of adults showed a pattern of positive projections for mammals in the small sample condition (10 participants) and in the large sample condition (11 participants), a pattern of negative projections for vertebrates and invertebrates in the small sample condition (12 participants in both conditions), and a pattern of negative projections for invertebrates in the large sample condition (10 participants). Similarly, school-age children exhibited a pattern of positive projections for mammals in the small sample condition (11 participants) and in the large sample condition (13 participants) and a pattern of negative projections for invertebrates in the small sample condition (10 participants). In contrast, preschoolers followed a pattern of positive projections for all targets in the large sample condition (13 participants for mammals, 10 participants for vertebrates, and 11 participants for invertebrates). No other patterns were significantly different from chance.

Discussion

Results of Experiment 1 provide evidence that sample size influenced induction in preschoolers such that larger samples led to a higher rate of projections than smaller samples. Furthermore, the larger sample of evidence about a diverse set of exemplars led children to generalize properties broadly both to the types of animals represented in the evidence sample (i.e., mammals) and to the types of animals not represented in the evidence sample (i.e., nonmammal vertebrates and invertebrates).

[1] Because the four experiments reported here included 15 participants from each age group, the same analytical strategy and statistical criterion were used across experiments. Note that the probability of following a pattern of positive projections or negative projections was different from the probability of following a pattern of random responding because the former included two different types of responses (e.g., projections to 3 or 4 targets and projections to 0 or 1 targets, respectively; the probability of either pattern is $p = .40$), whereas the latter included only one type of response (e.g., projections to 2 targets; the probability of the pattern is $p = .20$).

Sample size did not influence older participants' projections; both school-age children and adults generalized at an equal rate from small and large samples of evidence. Both older groups followed a pattern of narrow generalization, with a high rate of projections to only the exemplars from the same category as the evidence exemplars. In fact, there was evidence of negative projections in both groups of older participants, indicating that these participants used evidence that a property was associated with mammals as evidence that the property should not be generalized to invertebrates. These results suggest that the inferences of older participants were constrained to the category covered by the exemplars in the evidence regardless of sample size. Thus, older participants, but not younger children, generalized exclusively to the exemplars from the category presented in the samples. At least for older children and adults, this pattern of narrow generalization was unaffected by whether there was a large or small sample of evidence.

The results of Experiment 1 appear to be at odds with previous research suggesting that adults favor large samples over small samples when making inductive inferences (Osherson et al., 1990). We attribute this apparent conflict to task differences; the lack of the sample size effect for older participants likely stemmed from the fact that the design of Experiment 1 did not involve comparing between two samples of evidence. Our interpretation of the current results is that a sample consisting of 4 exemplars provided older participants with sufficient information to conclude that the evidence sample covered the category of *mammals*. This interpretation is consistent with prior research indicating that adults can estimate category coverage given as few as 3 exemplars (Osherson et al., 1990; Rips, 1975). Therefore, it is possible that if we asked adults in our study to choose a sample that provided the best evidence in support of a conclusion, they would choose a larger sample over a smaller sample. We explored this possibility directly in Experiment 2 by manipulating sample size within participants (as was done in prior research), thereby necessitating comparisons between large and small samples.

Overall, results of Experiment 1 provided evidence consistent with the hypothesis that an increase in sample size for a diverse sample of evidence would lead young children to generalize properties to a broad range of exemplars. However, these results do not eliminate the possibility that sample size alone was responsible for the higher rates of projections for the large samples. Similarly, because the larger sample in Experiment 1 was also the more diverse sample, it could be argued that children were influenced by sample diversity alone rather than by the interaction of sample diversity and sample size. One goal of the following studies was to explore these different possibilities. Thus, Experiment 3 tested the possibility that the effects reported in Experiment 1 stemmed from sample diversity alone, and Experiment 4 tested the possibility that the effects reported in Experiment 1 stemmed from sample size alone.

Experiment 2

The results of Experiment 1 suggest that preschoolers, but not older participants, were sensitive to sample size in the course of induction. These results are surprising given that the literature has consistently shown sample size effects in adults (Osherson et al., 1990; Sedlmeier & Gigerenzer, 1997). Our interpretation is that this apparent conflict underscores the importance of evidential context in the course of induction. Critically, Experiment 1 manipulated sample size between participants, whereas studies that have reported sample size effects in adults (Osherson et al., 1990) manipulated sample size within participants. The goal of Experiment 2 was to confirm that adults would prefer to generalize properties from large samples when the context involved choosing between two samples.

Method

Participants

The participants were 15 adults (undergraduate students, 10 women and 5 men), 15 school-age children (M = 7.6 years, SD = 0.43, 9 girls and 6 boys), and 15 preschoolers (M = 5.1 years, SD = 0.35, 7 girls and 8 boys).

Design and materials

The materials were identical to those used in Experiment 1. Sample size was manipulated within participants, who were presented with both a large and small sample of evidence and asked to generalize properties to each of the targets. All exemplars were presented serially such that all members of one sample were presented one by one at a particular screen location (e.g., top left quadrant), followed by serial presentation of members of the other sample at a different screen location. As in Experiment 1, properties were distributed within a sample at a 3:1 ratio. One property (e.g., "omat inside") was attributed to 75% of the members of one sample, whereas the other property (e.g., "unti inside") was attributed to 75% of the members of the other sample. The items were presented on a computer screen such that one sample appeared on the top left side of the screen and the other sample appeared on the top right. We counterbalanced the order (first or second) and location (left or right) of the small and large samples as well as the property type.

The composition of each sample was manipulated in the same way as in Experiment 1. The large samples contained all 16 evidence exemplars used in Experiment 1. The small sample contained a subset of 4 randomly chosen items from the 16 evidence exemplars.

After both samples were introduced, participants were asked to generalize one of the two properties to the target exemplars. The target exemplars were the same as those used in Experiment 1. Each target was presented, and participants were asked, "Do you think this animal has omat inside or unti inside?" Targets were presented in random order.

Procedure

Participants were told that they were going to learn about some animals and then answer questions. After learning about all of the individuals within one sample, participants were told that they would learn about a different set of animals. The individuals from the other sample were then presented. Prior to the projection phase, participants were told that they would be asked about some animals. The task lasted approximately 15 min.

Results

Analyses examined the proportion of large sample selections for each of the targets for each age group (see Table 3). The first analysis examined whether the rate at which participants chose the large samples was different from what would be expected by chance ($M = .50$). The mammal target was the only target to elicit a significant effect, and adults were the only age group to select the large sample for mammals at a rate significantly above chance, $t(14) = 5.92$, $p < .0001$.

Additional analyses examined potential age differences. Responses were analyzed using a mixed ANOVA with age (adults, school-age children, or preschoolers) as a between-participants factor and target (mammal, vertebrate, or invertebrate) as a within-participants factor. This analysis revealed a main effect of target, $F(1, 42) = 10.35$, $p = .002$, $\eta^2 = .19$, which was mediated by a target by age interaction, $F(2, 42) = 3.39$, $p = .04$, $\eta^2 = .12$. Simple effects analysis revealed significant age differences for the mammal targets such that adults selected the large sample more often than preschoolers, $F(1, 28) = 11.05$, $p = .003$, $\eta^2 = .28$. The responses of adults and preschoolers were not statistically different

Table 3
Proportion of selections of the large sample for each target for all three age groups in Experiment 2.

	Targets		
	Mammal	Vertebrate	Invertebrate
Adults	.75* (.16)	.45 (.25)	.42 (31)
School-age children	.63 (.26)	.42 (.29)	.35 (.28)
Preschoolers	.49 (.28)	.56 (.24)	.52 (.15)

Note. Standard deviations are in parentheses.
* $p < .01$ (one-sample t test).

Table 4
Number of individuals following a pattern of positive projections (projecting to 3 or 4 instances of a target), random projections (projecting to 2 instances of a target), or negative projections (projecting to 1 or 0 instances of a target) in Experiment 2.

		Targets		
Age	Pattern of projections	Mammals	Vertebrates	Invertebrates
Adults	Positive	12**	2	5
	Random	3	9*	2
	Negative	0	4	8
School-age children	Positive	8	4	2
	Random	4	6*	8*
	Negative	3	5	5
Preschoolers	Positive	6	6	2
	Random	6*	5	10*
	Negative	3	4	3

Note. A single asterisk (*) indicates that the number of individuals showing this pattern is greater than would be expected by chance at $p < .05$; a double asterisk (**) indicates a score that was different from chance at $p < .01$.

from the responses of school-age children, both *F*s < 2.00, *p*s > .15. There were no age differences for responses to vertebrates and invertebrates.

Individual patterns examined the number of participants who showed a pattern of positive, negative, and random projections for the property associated with the large sample. Table 4 shows the number of participants following each of these patterns. A significant number of adults showed a pattern of positive projections for mammals (12 participants) and showed a random pattern of projections for vertebrates (9 participants). Children consistently showed a random pattern of projections; school-age children showed a random pattern for vertebrates (6 participants) and invertebrates (8 participants), and preschoolers showed a random pattern for mammals (6 participants) and invertebrates (10 participants). No other patterns were significant.

Discussion

The results of Experiment 2 are consistent with the hypothesis that adults would prefer to generalize a property from a larger and more diverse sample rather than from a smaller and less diverse sample when the task involves selecting between two samples. Thus, these results are consistent with evidence that adults are sensitive to sample size and sample diversity when making inductive generalizations in forced choice tasks (Osherson et al., 1990). Moreover, the results are consistent with other work showing that young children do not prefer to generalize from a larger sample of evidence when the task involves choosing between two samples (i.e., when sample size is manipulated within participants) (Li et al., 2009; Lopez et al., 1992). Thus, these results support the interpretation that the responses of adults in Experiment 1 were not due to their inability to distinguish the greater inductive potential of large and diverse samples over small and less diverse samples but rather due to the context of the task.

Finally, it is worth noting that the sample size effect was restricted to mammals. The limited range of the effect is consistent with the view that adults interpreted the evidence as relevant to a specific "coverage" category (Osherson et al., 1990).

Experiment 3

Recall that one suggestion from Experiment 1 was that the broader range (and higher rate) of projections for large samples, at least for young children, was mediated by the diversity of exemplars within the sample. However, there are alternative interpretations of the results from Experiment 1. First, young children may prefer to generalize from larger samples regardless of sample diversity. Second, it could also be argued that whereas in Experiment 1 the exemplars in the small and large samples all were instances of the mammal category, the greater number of exemplars provided a better

coverage of the mammal category and so made the large sample more diverse. Thus, it is also possible that children's sensitivity to diversity information (alone) was responsible for the results from Experiment 1. Experiment 3 was designed, in part, to address these alternatives.

Experiment 3 had the same design as Experiment 1 with the exception of the exemplars used (see Table 1). Rather than learning about the properties of *mammals*, participants learned about a collection of *rabbits*. The target items included a novel set of rabbits, mammals (e.g., *cows*), and nonmammal vertebrates (e.g., *birds*) that were used in Experiment 1. Critically, the evidence sample was modified such that the same 4 *rabbits* were presented in both conditions, each occurring once in the small sample condition and 4 times in the large sample condition. We introduced only the same exemplars in both samples to ensure that the diversity of each sample was consistent across both sample size conditions.

If in Experiment 1 young children simply overgeneralized evidence from large samples, then in the large sample condition of Experiment 3 preschoolers should make projections to all target types—similar to their pattern of projections in Experiment 1. However, if sample size interacts with sample diversity, preschoolers could exhibit a narrower pattern of generalizations in Experiment 3 than in Experiment 1 because the evidence sample in Experiment 3 (i.e., *rabbits*) is less diverse than the evidence sample in Experiment 1 (i.e., *mammals*). Finally, if diversity (alone) caused children to make a broad range of generalizations for the large sample but not for the small sample in Experiment 1, we should observe no differences between samples given that the same 4 exemplars were used in both sample conditions. As in Experiment 1, we expected no sample size effects for adults and older children due to the between-participants manipulation of sample size; we expected that older participants would generalize properties to only rabbits for large and small samples.

Method

Participants

The participants were 30 adults (undergraduates, 14 women and 16 men), 30 school-age children (M = 7.5 years, SD = 0.60, 17 girls and 13 boys), and 30 preschoolers (M = 4.9 years, SD = 0.52, 14 girls and 16 boys).

Design, materials, and procedure

All facets of the design were the same as in Experiment 1 with the exception of the exemplars used. The evidence exemplars included a set of *rabbits*. The conclusion targets were the same as in Experiment 1 except that the invertebrates were eliminated and a category of novel *rabbits* was included. We chose this strategy in order to keep the amount of conclusion items constant across the experiments. Thus, the conclusion categories were *rabbits*, *mammals*, and nonmammal *vertebrates*. In all other respects, the design, materials and procedures were identical to Experiment 1.

Results

Responses to each of the items were averaged to yield mean projections for the rabbits, mammals, and vertebrates (see Fig. 2). Mean projections were submitted to an age (adults, school-age children, or preschoolers) by sample size (large or small) by target (rabbit, mammal, or vertebrates) mixed ANOVA. There was a main effect of sample size, $F(1, 82) = 3.89$, $p < .05$, $\eta^2 = .05$, with higher rates of projections in the large sample condition ($M = .55$) than in the small sample condition ($M = .46$), and a main effect of age, $F(2, 84) = 17.46$, $p < .0001$, $\eta^2 = .26$, due to a higher rate of projections among children (school-age: $M = .54$; preschoolers: $M = .61$) than among adults ($M = .36$), both $ps < .001$, $\eta^2 = .34$. Both of these factors were mediated by an age by sample size interaction, $F(2, 84) = 3.93$, $p < .05$, $\eta^2 = .04$. The age effect for the large sample condition, $F(2, 42) = 17.02$, $p < .0001$, $\eta^2 = .35$, indicated the following developmental pattern: Preschoolers ($M = .71$) made higher rates of projections than school-age children ($M = .54$), who responded at higher rates than adults ($M = .36$), all $ps < .01$. The age effect for the small sample condition, $F(2, 42) = 4.56$, $p < .05$, $\eta^2 = .20$, stemmed from both groups of children (school-age: $M = .53$; preschoolers: $M = .51$) making higher rates of projections than adults ($M = .36$), both $ps < .01$. As in Experiment 1, preschoolers were the only age group to make projections

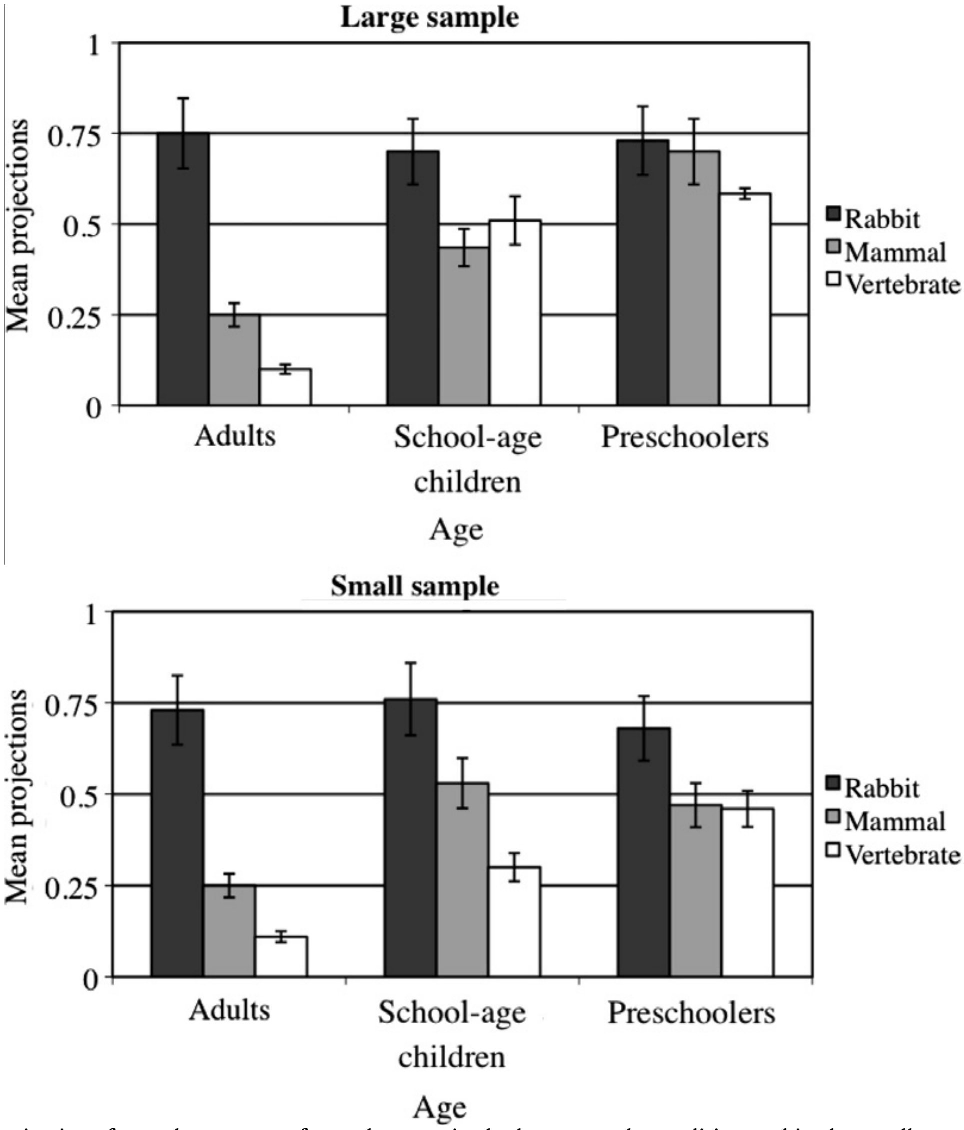


Fig. 2. Mean projections for each age group for each target in the large sample condition and in the small sample condition in Experiment 3. Bars indicate 1 standard error of the mean.

for large samples ($M = .71$) at a higher rate than for small samples ($M = .50$), $F(1, 28) = 5.91$, $p < .001$, $\eta^2 = .14$.

The analysis also showed a main effect of target, $F(2, 164) = 66.13$, $p < .001$, $\eta^2 = .48$, which was mediated by an interaction with age, $F(4, 164) = 6.90$, $p < .001$, $\eta^2 = .17$. Separate analyses revealed that adults showed the same target effect in both conditions, with rates of projections decreasing such that responses followed a pattern of projections for rabbits > projections for mammals > projections for vertebrates, all $ps < .05$. School-age children made a higher rate of projections to vertebrates in the large sample condition than in the small sample condition, $F(1, 28) = 7.39$, $p < .01$, $\eta^2 = .21$. Preschoolers made a higher rate of projections for large samples than for small samples for mammals, $F(1, 28) = 4.80$, $p < .05$, $\eta^2 = .16$, and vertebrates, $F(1, 28) = 5.51$, $p < .05$, $\eta^2 = .17$. No participants showed a sample size effect for projections to rabbits.

Analysis of the individual patterns of responses used the same analytical strategy as in Experiment 1. Table 5 presents the number of participants categorized as following positive, negative, or random patterns of projections. In both sample size conditions, participants in each age group showed a pattern of positive projections to rabbits (10 adults, 11 school-age children, and 10 preschoolers in the small sample condition; 10 participants in each age group in the large sample condition). In both sample size conditions, adults showed a pattern of negative projections to the mammal targets (11 participants in the small sample condition; 10 participants in the large sample condition) and vertebrates

Table 5
Number of individuals following a pattern of positive projections (projecting to 3 or 4 instances of a target), random projections (projecting to 2 instances of a target), or negative projections (projecting to 1 or 0 instances of a target) in Experiment 3.

Age	Condition	Pattern of projections	Targets		
			Rabbits	Mammals	Vertebrates
Adults	Large sample	Positive	10*	1	0
		Random	1	5	2
		Negative	4	10*	13**
	Small sample	Positive	10*	2	0
		Random	3	2	0
		Negative	2	11**	15**
School-age children	Large sample	Positive	10*	2	4
		Random	3	7*	8**
		Negative	2	6	3
	Small sample	Positive	11**	5	0
		Random	4	7*	6
		Negative	0	3	9
Preschoolers	Large sample	Positive	10*	10*	5
		Random	4	4	5
		Negative	1	1	5
	Small sample	Positive	10*	4	2
		Random	3	5	4
		Negative	2	6	9

Note. A single asterisk (*) indicates that the number of individuals showing this pattern is greater than would be expected by chance at $p < .05$; a double asterisk (**) indicates a score that was different from chance at $p < .01$.

(all 15 participants in the small sample condition; 13 participants in the large sample condition). A significant number of school-age children showed a pattern of random responding for mammals (7 each in the small sample and large sample conditions) and for vertebrates in the small sample condition (8 participants). Preschoolers showed a pattern of positive projections for mammals in the large sample condition (11 participants). No other patterns were significantly different from chance.

Breadth of generalization across Experiments 1 and 3

To examine possible changes in the breadth of generalization as a function of sample diversity, we compared children's performances in Experiment 1 and 3 using a subset of projection items present in both experiments (i.e., *mammals* and *vertebrates*). We recognize the need for caution when performing between-experiment comparisons; at the same time, we believe that such a comparison is justified in this case given the similarity in methods and populations of participants across the two experiments. We restricted these analyses to preschoolers because this was the only group to consistently show the sample size effect in Experiments 1 and 3.

To examine whether young children are more likely to broadly generalize properties that occur in relatively large and diverse samples, we compared the patterns of projections in preschoolers in the large sample condition of Experiment 1 and Experiment 3 for a subset of target items that were identical for the two experiments (i.e., *mammals* and *vertebrates*). Toward this goal, we performed a two-way ANOVA with sample diversity (high/Experiment 1 or medium/Experiment 3) and target type (mammals or vertebrates) as between-participants factors. The analysis revealed a main effect of sample diversity, $F(1, 28) = 8.66$, $p < .01$, $\eta^2 = .24$, and separate analyses for each target confirmed that preschoolers made a higher rate of projections for the high diversity sample (Experiment 1) than for the medium diversity sample (Experiment 3) for both sets of targets: mammal, $F(1, 28) = 7.22$, $p < .05$, and vertebrates, $F(1, 28) = 5.32$, $p < .05$.

Discussion

The main findings of Experiment 3 replicated those reported in Experiment 1: Only preschoolers consistently made a higher rate of projections from large samples than from small samples. Preschoolers exhibited the sample size effect when making projections to mammal and vertebrate targets. In

contrast, adults consistently generalized properties only to exemplars from the category represented in the evidence (i.e., other rabbits) regardless of whether a sample included a large or small number of exemplars. Older children showed a mixture of both patterns; they generated the same rate of projections for rabbits and mammals when presented with large and small samples, but they generated a higher rate of projections for vertebrates in the large sample condition than in the small sample condition. Thus, like Experiment 1, these results show that young children endorse a broader, more unconstrained range of projections from large samples than from small samples of evidence, whereas adults generalized properties more narrowly to the exemplars from the category represented in the evidence sample regardless of sample size. Moreover, because the same exemplars were presented in the small sample and the large sample in Experiment 3, the pattern of broader generalization for larger samples was observed when diversity across large and small samples was held constant.

At the same time, for large samples, preschoolers exhibited a narrower range of projections in Experiment 3 than in Experiment 1. Because the principal difference between Experiment 1 and Experiment 3 is the diversity of the evidence sample, this finding supports the idea that children broadly generalize those properties that consistently occur in large samples of diverse items. A straightforward prediction, therefore, is that the pattern of projections in preschoolers should become further narrowed down as diversity in the evidence sample is further decreased. Experiment 4 was designed to test this prediction.

Experiment 4

Experiment 4 examined the influence of sample size when evidence involved a homogeneous sample. Thus, participants were given either a small sample ($n = 4$) or a large sample ($n = 16$) of evidence about the same white rabbit. To ensure that the samples were maximally homogeneous, the small and large samples were represented by multiple instances of the same individual. The target categories included the *same individual, other rabbits,* and *other mammals.* Results of Experiment 3 undermined the possibility that in young children large samples always lead to a broad pattern of projections regardless of sample diversity. Experiment 4 was designed to provide a further test of this possibility. If sample size interacts with sample diversity to produce a pattern of broad generalization in young children, we might expect the sample size effect (higher level of projections in the large sample condition vs. the small sample condition) to diminish or disappear in Experiment 4, which included maximally homogeneous evidence samples.

Method

Participants

The participants were 30 adults (undergraduates, 14 women and 16 men), 30 school-age children ($M = 7.8$ years, $SD = 0.58$, 12 girls and 18 boys), and 30 preschoolers ($M = 5.0$ years, $SD = 0.55$, 16 girls and 14 boys).

Design, materials, and procedure

All facets of the design were the same as in the above experiments with the exception of the exemplars used. The evidence consisted of a single white rabbit presented 4 times (in the small sample condition) or 16 times (in the large sample condition). Target exemplars included 4 items from each of the following categories: the individual white rabbit presented as the evidence exemplar, novel rabbits (same as Experiment 3), and mammals (same as Experiments 1 and 3). In all other respects, the design was identical to Experiment 1. Table 1 depicts the items used in Experiment 4.

Results

Mean projections for each of the three targets in each condition are presented in Fig. 3. A mixed ANOVA with age (adults, school-aged children, or preschoolers) and sample size (large or small) as between-participants factors and target (same individual, rabbit, or mammal) as a within-participants factor revealed a main effect of age, $F(2, 84) = 6.60$, $p < .01$, $\eta^2 = .12$; both groups of children generated

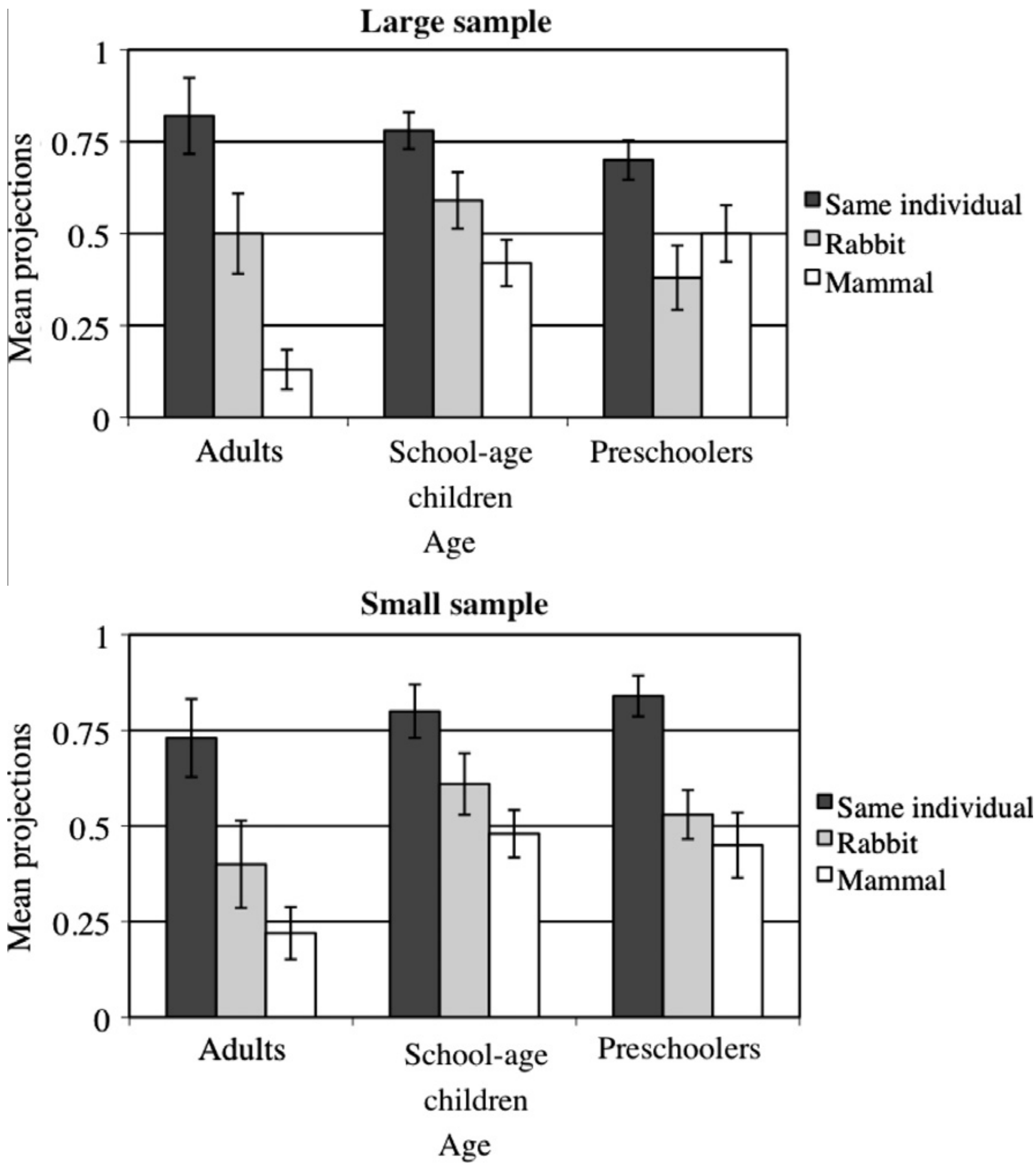


Fig. 3. Mean projections for each age group for each target in the large sample condition and in the small sample condition in Experiment 4. Bars indicate 1 standard error of the mean.

a higher rate of projections than adults, both *p*s < .05. The target by age interaction was not significant, $F < 1$, *ns.* As predicted, there was no effect of sample size, $F < 1$, *ns.* Further tests confirmed that participants in all age groups made projections to the same individual, novel rabbits, and mammals at an equal rate in the large and small sample conditions, all *F*s < 1.3, *p*s > .17.

We next explored individual patterns. All participants followed a pattern of positive projections for the same individual (white rabbit) in both sample conditions (12 adults, 10 school-age children, and 12 preschoolers in the small sample condition; 12 adults, 12 school-age children, and 11 preschoolers in the large sample condition). A significant number of adults followed a negative pattern of projections for mammals (10 participants in the small sample condition; 14 participants in the large sample condition). No other patterns were significantly different from chance. Table 6 presents the number of participants following each pattern of projection.

Breadth of generalization across Experiments 3 and 4

To examine whether the breadth of generalizations is influenced by sample diversity for relatively large samples, we compared the patterns of projections in preschoolers in the large sample condition of Experiments 3 and 4 for a subset of projection items that were identical for the two experiments

Table 6
Number of individuals following a pattern of positive projections (projecting to 3 or 4 instances of a target), random projections (projecting to 2 instances of a target), or negative projections (projecting to 1 or 0 instances of a target) in Experiment 4.

Age	Condition	Pattern of projections	Targets		
			White rabbit	Rabbits	Mammals
Adults	Large sample	Positive	12*	7	0
		Random	1	2	2
		Negative	2	6	13*
	Small sample	Positive	11*	5	1
		Random	0	2	0
		Negative	3	8	14*
School-age children	Large sample	Positive	11*	9	1
		Random	3	3	8*
		Negative	0	3	6
	Small sample	Positive	10*	9	1
		Random	3	3	10**
		Negative	1	3	4
Preschoolers	Large sample	Positive	10*	3	2
		Random	3	4	10**
		Negative	1	8	3
	Small sample	Positive	11*	5	3
		Random	3	7*	4
		Negative	0	3	8

Note. A single asterisk (*) indicates that the number of individuals showing this pattern is greater than would be expected by chance at $p < .05$; a double asterisk (**) indicates a score that was different from chance at $p < .01$.

(i.e., *rabbits* and *mammals*). We restricted this analysis to preschoolers because this was the only group to consistently show the sample size effect in Experiment 3. Toward this goal, we conducted an ANOVA with sample diversity (medium diversity/Experiment 3 or low diversity/Experiment 4) and target type (rabbits or mammals) as between-participants factors. The analysis revealed a main effect of sample diversity, $F(1, 28) = 7.72$, $p < .05$, $\eta^2 = .22$, and separate analyses for each target confirmed that preschoolers made a higher rate of projections for the medium diversity sample (Experiment 3) than for the low diversity sample (Experiment 4) for both sets of targets: rabbits, $F(1, 28) = 5.45$, $p < .05$, and mammals, $F(1, 28) = 6.39$, $p < .05$.

Sample size effect as a function of sample diversity: Comparisons among Experiments 1, 3, and 4

A final set of analyses explored differences in sample size effects across experiments. To further test the possibility that the observed sample size effects were more pronounced for more diverse evidence samples, we compared responses for the target that was used in all three experiments (mammal). All three age groups were included in this final analysis. We created a difference score by subtracting the mean projections in the small sample condition from each score in the large sample condition for the mammal targets in Experiments 1, 3, and 4. Fig. 4 presents these difference scores for each age group across these three experiments. We then conducted an ANOVA for each age group to determine whether the magnitude of the sample size effect was different for each experiment. There were no differences for both older groups, all Fs > 4, ps > .10. However, there was a significant effect for preschoolers due to a greater sample size effect in Experiment 1 than in Experiment 3, $F(1, 29) = 5.97$, $p = .02$, $\eta^2 = .16$, and a greater sample size effect in Experiment 1 than in Experiment 4, $F(1, 29) = 15.35$, $p < .001$, $\eta^2 = .33$. The differences in sample size effects in Experiments 3 and 4 did not reach significance, $F < 4$, $p = .08$.

Discussion

Results of Experiment 4 revealed that sample size had no effect on projections in any of the three age groups when participants were presented with a maximally homogeneous sample. Participants in all age groups made positive projections about the same individual for large and small samples. Unlike Experiments 1 and 3, preschoolers in Experiment 4 did not use larger samples to generalize to a broad

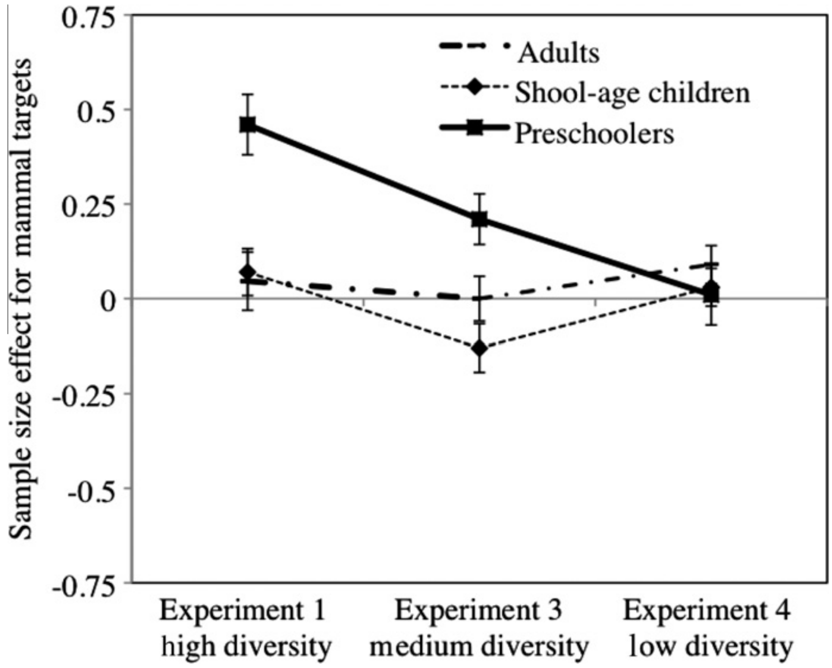


Fig. 4. Sample size effect (i.e., difference scores between large and small samples) for the mammal targets for each age group in Experiments 1, 3, and 4.

range of exemplars. Therefore, these results support the hypothesis that sample size has the greatest influence on projections from samples in which there is diversity among the exemplars.

The analysis across Experiments 3 and 4 showed that young children made a broader range of projections when the evidence sample included different rabbits (Experiment 3) than when the evidence sample included the same white rabbit (Experiment 4). Finally, the analysis of the sample size effect for the common item used in Experiments 1, 3, and 4 indicates that the sample size effect was greatest for the diverse samples. These findings provide further evidence that—at least for young children—the breadth of generalization was influenced by the diversity of exemplars within a sample.

General discussion

The studies reported in this article point to two novel findings. First, children and adults were highly sensitive to task context in the course of induction. When the task involved comparisons between samples (Experiment 2), the sample size effect was observed for adults but not for children younger than 8 years old—a finding consistent with prior research (Gutheil & Gelman, 1997; Lopez et al., 1992; Osherson et al., 1990). However, when the task did not involve comparisons between samples (i.e., sample size was manipulated between participants rather than within participants in Experiment 1), the opposite pattern of results was observed; preschoolers were sensitive to sample size, whereas adults were not. The latter finding was replicated in Experiment 3 with a set of evidence exemplars that was different from that in Experiment 1.

Second, preschoolers were the only age group sensitive to the interaction of sample size and sample diversity when the task did not involve comparisons between samples. In particular, preschoolers made very broad generalizations when presented with relatively large and diverse samples; for instance, children generalized properties attributed to a set of mammals to other mammals as well as to vertebrates and invertebrates (Experiment 1). At the same time, children's projections became progressively narrower as the diversity of the evidence sample decreased (Experiments 3 and 4). These findings shed light on one of the central questions in the study of induction: How do children decide that some properties can be generalized across a range of exemplars? Evidence presented in this article suggests that preschoolers infer that a property is broadly generalizable if it occurs for a large number of diverse exemplars. The remainder of this discussion considers the implications of these results for current theories of induction.

Sensitivity to sample size and diversity in young children's induction

Some research has shown that prior to 6 years of age, children consider sample size when making inferences (Jacobs & Narloch, 2001; Xu & Tenenbaum, 2007), whereas other research has shown that this ability appears later in development (Li et al., 2009; Lopez et al., 1992). To us, the most parsimonious interpretation of these conflicting findings is that some contexts support the use of sample size by young children, whereas other contexts do not. The design of the current studies offers several suggestions regarding the factors that contribute to young children's use of sample size information to support inductive generalizations.

Prior work showed that before 7 or 8 years of age, children do not consider sample size in their inductive generalizations when the task involves a forced choice between a large sample and a small sample (Gutheil & Gelman, 1997; Li et al., 2009), but that children do consider sample size when the task involves generalizing from a single sample (Jacobs & Narloch, 2001; Xu & Tenenbaum, 2007). The current research established this finding within the same study that involved similar stimuli and methods: In Experiments 1 and 3, preschoolers showed a sample size effect when generalizing from a single sample (i.e., when sample size was manipulated between participants), whereas children did not show a sample size effect when choosing between two samples (i.e., when sample size was manipulated within participants in Experiment 2). The current results support the conclusion that young children are sensitive to sample size but that the context of comparison—attending to each sample, assessing sample–conclusion relations independently, and then choosing which set of evidence provides the best support for the conclusion—may place too many demands on children.

In the current studies, sample size effects were pronounced for the most diverse sample (Experiment 1) and were absent for the most homogeneous sample (Experiment 4). Use of homogeneous exemplars in some of the previous research (e.g., Gutheil & Gelman, 1997) could be another reason why previous research on the use of sample size information led to null results. We exercise some caution in applying these results more generally to literature on the emergence of diversity-based reasoning. Diversity-based reasoning has been described in terms of normative induction (Heit et al., 2005) to suggest that people use diversity to constrain inferences to the most likely conclusion. In the current studies, children's responses violated the diversity principle; rather than constraining inferences to a narrow set of conclusions, young children generalized more broadly from the more diverse evidence. Young children may have endorsed a broader range of inferences because they were unable to evaluate the category covered by a diverse set of evidence (e.g., Gutheil & Gelman, 1997; Lopez et al., 1992). Older participants appear to have recognized that a sample was representative of a particular category and so constrained their inferences to only members of the evidence category. In fact, in many cases adults used evidence that members of a category have a certain property to support the inference that members of a different category would lack this property.

It is also possible that young children followed a non-normative pattern because the evidence lacked strong pragmatic cues. Pragmatic cues support the assumption that evidence is meant to be useful or helpful and so was provided with relevant constraints in mind (Rhodes et al., 2010). The absence of strong pragmatic cues may cause young children to rely strictly on the statistical properties—the amount of evidence and variability of the exemplars within the sample—as a basis for generalization.

Finally, in the current studies, participants were presented with probabilistic properties, whereas other studies have used stable properties. Children may attend to sample size and sample diversity when they are primed to think about the stochastic nature of the evidence. However, note that null effects in Experiment 4 suggest that this explanation alone does not account for the sample size effects observed in Experiments 1 and 3.

Implications for existing theories of induction

There has been considerable interest in understanding the conditions under which children generalize properties from known to novel exemplars. For example, prior work documented that preschoolers are willing to broadly generalize some properties (i.e., *has important internal parts*) but not others (e.g., *is dusty*) (Keil et al., 1998). One explanation for this distinction is that children generalize those

properties that fit within their naive theories about which properties are associated with a specific category of exemplars (Gelman, 2003; Keil et al., 1998). From this perspective, children will generalize properties such as *has important internal parts* more broadly than properties such as *is dusty* because they understand that having important parts is a property that is representative of all members of the same biological category, whereas being dusty is an incidental property not related to category membership.

Another theoretical perspective suggests that category-based induction proposed by the naive theory approach is characteristic of mature generalization but not of early generalization. Within this approach, generalization early in development is based on low-level mechanisms of perception, attention, and memory (Fisher, 2010; Fisher & Sloutsky, 2005; Rakison & Lupyan, 2008; Sloutsky, 2003; Sloutsky & Fisher, 2005). In general, this perspective is consistent with a large body of evidence suggesting that young learners are highly sensitive to statistical features of available evidence—such as feature distributions and feature correlations—in the course of language and category learning (Colunga & Smith, 2002; Rakison & Hahn, 2004; Samuelson et al., 2009).

Results presented in this article are inconsistent with the approaches that (a) grant young children the ability to perform spontaneous category-based generalization and (b) predict no developmental differences in the mechanisms of generalization between children and adults (Gelman, 2003; Keil et al., 1998; Xu & Tenenbaum, 2007). Contrary to the predictions of the category-based approaches, there were large differences in the patterns of projections between younger and older participants. These observed age differences are consistent with the notion that mechanisms of induction change in the course of development (Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2004, 2005; Sloutsky, Kloos, & Fisher, 2007) and that young children rely on statistical properties of evidence as a basis for reasoning about the properties of exemplars (Lawson & Fisher, 2008; Rakison & Hahn, 2004).

This study raises questions about the mechanisms underlying the reported effects. One interpretation of these results is that the observed age differences were due to differences in category knowledge. It is possible that young children generalized properties from the large sample more broadly in Experiments 1 and 3 because they do not have the sufficient knowledge of—or familiarity with—the categories of *mammals* (Experiment 1) and *rabbits* (Experiment 3). Although this interpretation might explain why young children fail to constrain their inferences to a specific category (Experiments 1 and 3), it does not explain why children broadly generalize properties associated with a large and diverse sample (Experiment 1) but not properties associated with a large and nondiverse sample (Experiment 4).

Sample size differences may also be the result of developmental differences in memory processes. For instance, older participants may have the capacity to remember the range of exemplars presented in a large sample of evidence, which would support a common pattern of inferences in the large and small samples. Younger children, on the other hand, might not be able to recall all of the exemplars from a large sample of evidence and so rely on a different strategy to generalize properties from large samples than the small samples. This possibility remains to be explored in future research.

The current studies explored children's use of sample size information when making generalizations in the domain of living kinds. However, young children's sensitivity to sample size has also been demonstrated in the social cognition literature, particularly in studies of person perception (Boseovski & Lee, 2006), stereotyping (Klaczynski & Aneja, 2002), and social inferences (Jacobs and Narloch, 2001). Although there are a number of methodological differences that make it difficult to compare studies directly, we believe that findings from these studies and the current experiments converge to support the conclusion that young children are able to rely on statistical features of data—specifically, the amount of available information—to make inductive generalizations.

Conclusions

Taken together with the findings of Lawson and Fisher (2008), the current findings suggest an answer to the question posed at the beginning of this article. Specifically, for a property to become broadly generalizable among living kinds, properties need to satisfy three conditions, each of which is necessary but none of which is sufficient: The property needs to occur (a) with high regularity,

(b) within a diverse set of exemplars, and (c) for a relatively large set of items. Overall, these results indicate that young children are highly sensitive to regularities in the available evidence during the process of inductive generalization.

References

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*, 500–513.

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

Colunga, E., & Smith, L. B. (2002). What makes a word? In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 214–219). Mahweh, NJ: Lawrence Earlbaum Associates, Inc..

Fisher, A. F., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development, 76*, 583–597.

Fisher, A. V. (2010). Mechanisms of induction early in development. In M. Banich & D. Caccamise (Eds.), *Generalization of knowledge: Multidisciplinary perspectives* (pp. 89–112). New York: Psychology Press.

Gelman, S. A. (1988). Development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65–96.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64*, 159–174.

Hayes, B. K. (2007). The development of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 25–54). New York: Cambridge University Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chate (Eds.), *Rational models of cognition* (pp. 248–274). Oxford University Press.

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 47*, 243–273.

Heit, E., Hahn, U., & Feeney, A. (2005). Defending diversity. In W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside of the laboratory: Essays in honor of Douglas L. Medin* (pp. 87–90). Washington, DC: American Psychological Association.

Jacobs, J. E., & Narloch, R. H. (2001). Children's use of sample size and variability to make social inferences. *Applied Developmental Psychology, 22*, 311–331.

Keil, F. C., Smith, C. S., Simons, D., & Levin, D. (1998). Two dogmas of conceptual empiricism. *Cognition, 65*, 103–135.

Klaczynski, P. A., & Aneja, A. (2002). Development of quantitative reasoning and gender biases. *Developmental Psychology, 38*, 208–221.

Lawson, C. A., & Fisher, A. V. (2008). Children's attention to property likelihood as a guide to property projection. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual meeting of the Cognitive Science Society* (pp. 1557–1561). Mahweh, NJ: Lawrence Earlbaum Associates, Inc..

Li, F., Cao, B., Li, Y., Li, H., & Deák, G. (2009). The law of large numbers in children's diversity-based reasoning. *Thinking and Reasoning, 15*, 388–404.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 26*, 181–206.

Lopez, A. (1995). The diversity principle in the testing of arguments. *Memory and Cognition, 23*, 374–382.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63*, 1070–1090.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.

Rakison, D. H., & Hahn, E. (2004). The mechanisms of early categorisation and induction: Smart or dumb infants? In R. Kail (Ed.), *Advances in child development and behavior (Vol. 32* (pp. 281–322). San Diego: Academic Press.

Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development, 73*(1), 1–110.

Rhodes, M., Gelman, S. A., & Brickman, D. (2008). Developmental changes in the consideration of sample diversity in inductive reasoning. *Journal of Cognition and Development, 9*, 112–143.

Rhodes, M., Gelman, S. A., & Brickman, D. (2010). Children's attention to sample composition in learning, teaching, and discovery. *Developmental Science, 13*, 421–429.

Rips, L. J. (1975). Induction about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*, 665–681.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Samuelson, L. K., Schutte, A. R., & Horst, J. S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalizations. *Cognition, 110*, 322–345.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*, 33–51.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*, 246–251.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in preschool-aged children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166–188.

Sloutsky, V. M., & Fisher, A. V. (2005). Similarity, induction, naming, and categorization (SINC): Generalization or verbal inductive reasoning? Response to Heit and Hayes. *Journal of Experimental Psychology: General, 134*, 606–611.

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development, 79*, 639–651.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*, 179–185.

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