



Brief article

Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change[☆]

Marjorie Rhodes^{*}, Daniel Brickman, Susan A. Gelman

*Department of Psychology, University of Michigan, 2024 East Hall,
530 Church Street, Ann Arbor, MI 48109-1043, USA*

Received 31 May 2007; revised 29 September 2007; accepted 7 March 2008

Abstract

Evaluating whether a limited sample of evidence provides a good basis for induction is a critical cognitive task. We hypothesized that whereas adults evaluate the inductive strength of samples containing multiple pieces of evidence by attending to the relations among the exemplars (e.g., sample diversity), six-year-olds would attend to the degree to which each individual exemplar in a sample independently appears informative (e.g., premise typicality). To test these hypotheses, participants were asked to select between diverse and non-diverse samples to help them learn about basic-level animal categories. Across various between-subject conditions ($N = 133$), we varied the typicality present in the diverse and non-diverse samples. We found that adults reliably selected to examine diverse over non-diverse samples, regardless of exemplar typicality, six-year-olds preferred to examine samples containing typical exemplars, regardless of sample diversity, and nine-year-olds were somewhat in the midst of this developmental transition.

© 2008 Elsevier B.V. All rights reserved.

[☆] This research was supported by NICHD Grant HD-36043 to the third author. The first and second authors were supported by funding from the Michigan Prevention Research Training Grant (NIH Grant number T32 MH63057-03). We are grateful to the parents, teachers, and children of the Ann Arbor Public Schools for participating in this research.

^{*} Corresponding author. Tel.: +1 734 646 2360; fax: +1 734 615 0573.

E-mail address: rhodesma@umich.edu (M. Rhodes).

Keywords: Inductive reasoning; Development; Typicality; Diversity

1. Introduction

Inductive reasoning is central to human learning, as most of the knowledge we possess is acquired via inductive inferences rather than through direct instruction or observation. Thus, there has been considerable interest in understanding the basis of inductive reasoning processes throughout development. The question of how inductive reasoning skills develop relates to a major theoretical debate in the field of cognitive development regarding the relative contributions to development of knowledge enrichment and of conceptual change. Do young children generalize knowledge using the same reasoning mechanisms as adults do (sometimes arriving at different conclusions due to limitations in their knowledge base), or does young children's inductive methodology differ systematically from the adult approach (e.g., Carey, 1985; Heit, 2000; Viale & Osherson, 2002)? Within this framework, the goal of the present research was to examine developmental changes in how individuals approach a key challenge of inductive reasoning—determining whether limited evidence provides a strong sample on which to base broader generalizations.

1.1. Adults' criteria for evaluating samples

As discussed by Heit (2000), adults have different strategies for evaluating samples that contain single versus multiple pieces of evidence. For example, to evaluate whether a single bird is informative about all birds, adults usually consider how typical the given exemplar is of the category “birds,” and base generalizations on typical exemplars (e.g., robins) more than atypical exemplars (e.g., penguins, Rips, 1975). When evaluating samples containing multiple pieces of evidence, however, adults do not simply sum the typicality of each of the given exemplars; rather, they focus on group-level properties.

To illustrate this distinction, consider the following inductive problem: If a person wants to learn about a biological property of mammals, is it better to examine a sample containing a lion and a tiger, or a sample containing a whale and a llama? Lions and tigers are likely to be perceived as more typical of the category ‘mammal’ than are either whales or llamas (e.g., Barr & Caplan, 1987; Diesendruck & Gelman, 1999). Yet, as described by Osherson, Smith, Wilkie, Lopez, and Shafir (1990), the sample of a whale and llama seems relatively more informative in this case because adults evaluate multiple-exemplar samples by attending to the extent to which the given exemplars, together, *cover* the relevant inclusive category. Whales and llamas provide broad coverage of mammals in that they represent very different kinds of mammals. Viewing diverse samples as more informative is a robust phenomenon among adult populations (Heit & Feeney, 2005; Kim & Keil, 2003; Lopez, 1995; but see Medin, Coley, Storms, & Hayes, 2003, for a description of cross-cultural var-

iation) and is consistent with theorizing from the philosophy of science indicating that it is more valid to draw conclusions based on evidence obtained from diverse sources (Heit, Hahn, & Feeney, 2004).

1.2. Children's criteria for evaluating samples

When evaluating samples that contain single pieces of evidence, children as young as five years old share adults' preferences for typical exemplars (Lo, Sides, Rozelle, & Osherson, 2002; Lopez, Gelman, Gutheil, & Smith, 1992). There is decidedly mixed evidence, however, regarding the criteria that children use to evaluate multiple-exemplar samples.

Lopez et al. (1992) examined preferences for diverse samples developmentally by asking six-year-olds, eight-year-olds, and adults to select whether to generalize a novel property found among diverse or non-diverse sets of animals. In these experiments, adults reliably preferred to generalize properties found among diverse sets, eight-year-olds demonstrated limited sensitivity to diversity, but six-year-olds responded randomly. Gutheil and Gelman (1997) replicated these negative diversity effects among children in a series of studies focusing on basic-level animal categories (see also Rhodes, Gelman, & Brickman, 2008). Based on these findings, the authors of these studies proposed that young children do not attend to sample diversity to evaluate evidence, and therefore, that there are important developmental changes in the mechanisms that support inductive reasoning.

An alternate perspective, however, suggests that apparent developmental changes in inductive reasoning result from limitations in young children's knowledge base (e.g., Carey, 1985; Heit & Hahn, 2001). To examine this possibility, Heit and Hahn (2001) designed studies that involved items and properties thought to be more familiar to young children. They presented five-year-olds with two samples of dolls, for example, including three diverse dolls belonging to one character and three non-diverse dolls belonging to another character. Participants were then shown another different doll, and were asked to predict which character owned the doll. On these questions, children reliably responded that the target doll belonged to the character that owned the diverse set of dolls. Based on these data, Heit and Hahn (2001) suggest that young children can engage in diversity-based reasoning under simplified conditions.

These findings demonstrate that young children can recognize sample diversity and sort evidence based on diversity. In our view, however, they do not clearly indicate that young children view diverse samples as a stronger basis for induction. Instead, children may solve these problems by recognizing that diverse items better match diverse (rather than non-diverse) sets of evidence (Lo et al., 2002). Similarly, Shipley and Shepperson (2006) report that preschool children prefer to test toys from two sub-classes (e.g., one blue whistle and one red whistle) in order to determine if the whistles 'made good party favors.' Although this task also reveals that young children recognize and reason about sample diversity, because participants were not asked to make an inference about a larger set (e.g., whistles not included in either specific sub-class), this study also does not provide clear evidence that young chil-

dren use sample diversity to determine whether a sample provides a good basis for broader generalizations (i.e., about a larger category or unobserved instance).

1.3. *The present research*

In the present study, we further examine diversity-based reasoning in children and also directly examine the possibility that the mechanisms that support inductive reasoning among young children differ systematically from those that support adult induction. Although previous research has documented that young children do not consistently demonstrate preferences for diverse samples on tasks designed to elicit diversity-based reasoning, prior work has not focused on revealing the kinds of strategies that children *do* in fact apply to such problems. Identifying that children have a consistent approach to these inductive problems would support the proposal that there are meaningful developmental changes in the mechanisms that support induction.

As reviewed above, a requirement for engaging in diversity-based reasoning is that individuals evaluate samples by considering the relations among premise exemplars (e.g., how well the given exemplars, taken together, cover an inclusive category), as opposed to the inductive potential of each premise considered separately (e.g., premise typicality). We proposed that young children's difficulty evaluating multiple-exemplar samples based on diversity may result from a failure to evaluate these group-level relations; instead, we hypothesized that young children evaluate samples by focusing on the individual properties (e.g., the typicality) of each of the given exemplars. If this hypothesis is correct, young children should prefer to base generalizations on multiple-exemplar samples that contain the most informative individual exemplars, regardless of the relations among the exemplars. We tested this possibility by examining how children and adults weight the influence of properties of individual examples in a sample (i.e., typicality) and the properties of whole samples (i.e., diversity) in evaluating inductive potential.

To examine participants' evaluation of different kinds of samples, we asked participants to select pairs of animals that they would like to examine to help them learn about a basic-level animal category. In all conditions, participants were given choices between diverse and non-diverse samples of animals (e.g., two dogs of different or the same type) to help them learn about basic-level categories (e.g., dogs). This task explicitly required participants to select the sample they believe is the stronger one, and therefore, is a direct test of evidence evaluation (Lopez, Atran, Coley, Medin, & Smith, 1997; Rhodes et al., 2008; Shipley & Shepperson, 2006). We asked participants to select samples to support a general conclusion, as opposed to a specific conclusion, to preclude a strategy based on similarity matching.

To test our hypothesis about the differential influence of typicality and diversity for each age-group, we constructed four sets of stimuli corresponding to a between-subjects design with four conditions. Across the conditions, we varied the amount of typicality present in the diverse and non-diverse sets (see Table 1). We included six-year-olds and adults, as well as a sample of nine-year-olds in order to better chart the course of developmental change. As detailed in Table 1, we hypothesized that when typicality was held constant (all exemplars were typical or atypical)

Table 1
Summary of conditions and hypotheses

Condition	Non-diverse set	Diverse set	Hypotheses
All-Typical	Typical	Typical	Adults favor diverse Nine-year-olds favor diverse Six-year-olds respond at chance
All-Atypical	Atypical	Atypical	Adults favor diverse Nine-year-olds favor diverse Six-year-olds respond at chance
Diverse-Typical	Atypical	Typical	Adults favor diverse Nine-year-olds favor diverse Six-year-olds favor diverse
Diverse-Atypical	Typical	Atypical	Adults favor diverse Nine-year-olds respond at chance Six-year-olds favor non-diverse

nine-year-olds and adults would select diverse samples and six-year-olds would choose randomly. When typicality varied across the two samples, however, we hypothesized that six-year-olds would choose to examine the typical samples, regardless of whether they contained diverse or non-diverse exemplars, but that adults would reliably select to examine diverse sets. When typicality and diversity were directly pitted against each other (e.g., a strategy based on diversity would favor the diverse set and a strategy based on typicality would favor the non-diverse set), we expected nine-year-olds to demonstrate chance-level responding, reflecting the relative fragility of their preference for diversity (e.g., Gutheil & Gelman, 1997; Lopez et al., 1992; Rhodes et al., 2008).

2. Method

2.1. Participants

Participants ($N = 133$) included 43 six-year-olds (20 male, 23 female, M age = 6.4; range: 5.5–7.2), 42 nine-year-olds (15 male, 27 female, M age = 9.2; range: 8.2–10.4), and 48 college students (19 male, 29 female, M age = 18.9; range: 18.2–19.9). An additional 10 six-year-olds and 14 college students participated in pre-testing of the stimuli (see below). Children were recruited from a public elementary school in a Midwestern university town via letters sent home to all children in participating classrooms; only children who returned signed permission letters were included. College students were recruited from an introductory psychology subject pool and received partial course credit for participating.

2.2. Stimulus development

To test these hypotheses it was necessary to create sets of stimuli that varied as desired in terms of typicality. To do so, we collected color photographs of 20–30 ani-

imals from each of five basic-level animal categories: birds, cats, dogs, fish, and pigs. First, a group of 14 college students rated the typicality of each exemplar with respect to its category on a seven-point scale, following the “goodness of example” procedure developed by Rosch (1975). The task was presented via computer. The items were blocked by category, with the categories, and items within each category, presented in a separate random order for each participant. Based on these ratings, we selected two species of each category that were perceived as highly typical and two species that were perceived as less typical. Analyses confirmed that the average typicality ratings of the species selected as typical and atypical exemplars significantly differed (see Table 2).

To confirm that young children also perceived the typicality of the selected examples as desired, we conducted a pre-test with a group of 10 six-year-olds. Children were interviewed individually. The task was introduced as follows, “I am making a book about different animals. For each kind of animal, I have to pick the best picture to put in the book.”

Children were asked to choose pictures for each animal category. For example, children were presented with four pictures of dogs, including two from the atypical set and two from the typical set, and were asked, “We want to pick the picture that best shows DOG, so that anyone who looked at our book would understand what DOG really means. What is the best picture to show DOG?” After their initial choice, participants were asked to select the “next best” two additional times. Thus, for each set, we obtained a ranking of how well children thought each exemplar represented the animal category. The animal categories, and the items within each category, were presented in a separate random order for each participant. As predicted, at least nine out of the ten children chose the two typical exemplars as their first two choices of ‘best examples’ for each animal category; significantly more often than

Table 2
Animal species selected as typical and atypical exemplars of each basic-level category

Category	Mean (<i>SD</i>) typicality ratings				<i>t</i> ^a
	Typical		Atypical		
Dog	Golden retriever	7.0 (.00)	German spitz mittel	5.6 (2.1)	2.35*
	Black labrador	6.8 (.40)	Hairless Chinese crested	5.6 (1.9)	
Fish	Goldfish	6.9 (.45)	Yellow boxfish	5.0 (2.3)	3.96**
	Blue angelfish	6.6 (.55)	Pufferfish	3.1 (0.9)	
Pigs	Yorkshire	6.8 (.45)	Warthog	3.4 (1.0)	7.45**
	Hampshire	6.6 (.55)	Javelina	3.4 (0.9)	
Cats	Orange tabby	6.8 (.45)	Checkered ragdoll	5.6 (2.1)	3.23**
	Grey and white longhair	7.0 (.00)	Red cameo Persian longhair	5.6 (1.3)	
Birds	Cardinal	6.8 (.45)	Owl	4.8 (2.4)	4.27**
	Blue jay	6.8 (.45)	Variable oystercatcher	5.6 (1.7)	

^a Paired-sample *t*-test comparing mean typicality ratings for items selected as typical and atypical exemplars with 13 degrees of freedom.

* *p* < .05.

** *p* < .01.

expected by chance (sign tests, $ps < .05$). Based on these findings, we constructed four sets of stimuli, one for each of the conditions described in Table 1. These stimuli are summarized in Table 2.

2.3. Main study procedures

Children were tested individually in a quiet area of their elementary school. Children were given the following instructions, “We’re going to pretend that you are a scientist who is trying to learn about animals. Your job is to pick the best set of animals to look at to help you learn about each kind of animal. For each question, look at each pair of animals carefully, and pick the best set to look at to help you learn new things about animals.”

Children were then asked a series of five questions. For each question, children were shown two sets of animals, a diverse and a non-diverse set, each containing color photographs (9×13 cm) of two animals. Each pair of animals was mounted on a piece of white paper. For example, children were shown one diverse (e.g., a black lab and a golden retriever) and one non-diverse (e.g., two different golden retrievers) set of dogs, and told, “Here are two sets of dogs. You are a scientist who wants to find out if dogs have ulnar arteries. Which set of dogs do you want to look at to learn about dogs?” Children pointed to indicate their responses. Adults completed a questionnaire version of the task. All questions referred to categories using generic nouns “e.g., to learn about DOGS.” This language choice, as well as the instructions, which referred to learning about *kinds* of animals, insured that children understood the task as an attempt to learn about a category as a whole, not just the given exemplars, as even preschoolers understand generic noun phrases as referring to categories, not only pictured exemplars or a subset (e.g., Gelman & Raman, 2003; Hollander, Gelman, & Star, 2002). Questions were presented in a separate random order for each participant. Presentation of the diverse set on the participants’ right or left was counter-balanced across questions and participants.

Participants of each age-group were randomly assigned to one of four conditions. All participants received identical instructions and questions. Across the conditions, we varied the typicality of the diverse and non-diverse sets to form four conditions (see Table 3).

Table 3
Sample stimuli for category DOG for each condition

	All-Typical	All-Atypical	Diverse-Typical	Diverse-Atypical
Diverse	Black labrador and golden retriever	Hairless Chinese crested and German spitz mittel	Black labrador and golden retriever	Hairless Chinese crested and German spitz mittel
Non-diverse ^a	Two golden retrievers	Two hairless Chinese crested	Two hairless Chinese crested	Two golden retrievers

^a Although the non-diverse sets included animals of the same type, the two exemplars in each set were easily identifiable as different individuals (e.g., they were slightly but noticeably different shades of color, size, etc.).

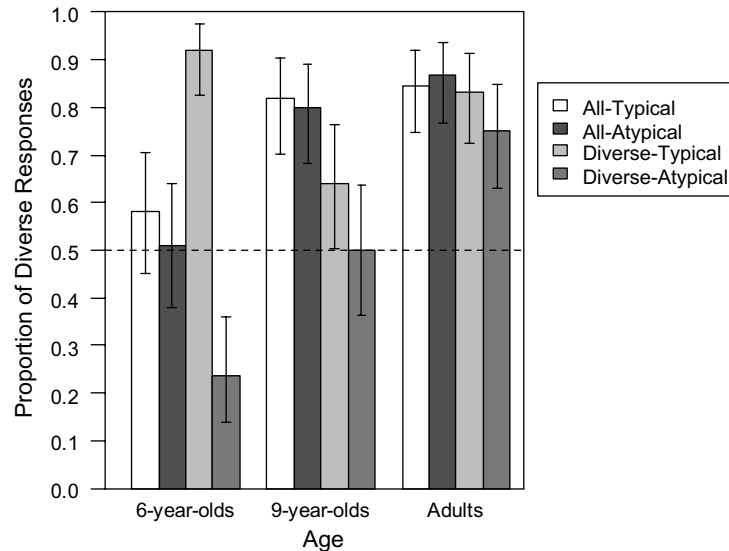


Fig. 1. Proportion of diverse responses by age and condition.

3. Results

Selections of the diverse samples were scored as ‘1’; selections of non-diverse samples were scored as ‘0.’ In order to account for the binomial structure of the data, we conducted a series of binomial regressions, using the χ^2 distributed deviance-based test (Faraway, 2006).¹ Total mean proportions by age-group and condition are presented in Fig. 1; mean proportions for each category set are presented in Table 4.

As hypothesized, there was an overall significant interaction between age-group and condition, $\chi^2(7) = 55.96$, $p < .001$. Among adults, there was no effect of condition $\chi^2(3) = 3.14$, $p > .3$. Adults selected diverse samples significantly more often than expected by chance in all conditions.

Among nine-year-olds, there was a significant effect of condition, $\chi^2(3) = 16.38$, $p < .001$. Comparisons among the conditions indicated that nine-year-olds selected the diverse samples more often in the All-Typical and All-Atypical conditions than in the Diverse-Atypical condition (Diverse-Atypical compared to All-Typical, $\chi^2(1) = 12.20$, $p < .001$, and All-Atypical, $\chi^2(1) = 10.65$, $p < .01$). Nine-year-olds chose the diverse samples more often than expected by chance in all conditions, except for the Diverse-Atypical condition.

¹ Although this analysis plan is the appropriate approach given the structure of our data, we also analyzed these data through a 3 (age-group) \times 4 (condition) analysis of variance with total proportion of diverse responses as the dependent variable. These analyses revealed an identical pattern of significant results.

Table 4
Proportion of diverse responses for each stimulus set by age-group and condition

	Six-year-olds				Nine-year-olds				Adults			
	All-Typical	All-Atypical	Diverse-Typical	Diverse-Atypical	All-Typical	All-Atypical	Diverse-Typical	Diverse-Atypical	All-Typical	All-Atypical	Diverse-Typical	Diverse-Atypical
Dog	.55	.55	1**	.36	.82*	1**	.80	.40	.85*	.83*	.92*	.82*
Pig	.55	.45	1**	.09*	.64	.73	.50	.50	.85*	.92*	.92*	.36
Bird	.64	.45	1**	.00**	.91**	.73	.60	.70	.85*	.92*	.92*	1**
Cat	.55	.63	.90*	.09*	.82*	.73	.90*	.50	.85*	.83*	.83*	.73
Fish	.64	.45	.70	.64	.91**	.82*	.40	.40	.85*	.83**	1**	.73

Note. Comparisons to proportion expected by chance using binomial regression.
* $p \leq .05$.
** $p < .01$.

Among six-year-olds, we also found a significant effect of condition, $\chi^2(3) = 56.56, p < .001$. Comparisons among conditions indicated that when typicality differed between the two samples, six-year-olds reliably chose to examine the sets that contained the typical exemplars, regardless of diversity. Thus, in the Diverse-Typical condition, six-year-olds chose the diverse set more often than in any other condition ($ps < .001$), and did so more often than expected by chance. But, in the Diverse-Atypical condition, they chose the non-diverse (i.e., typical) set more often than in any other condition ($ps < .01$), and chose the diverse set significantly less often (i.e., they chose the non-diverse set more often) than expected by chance. When the level of typicality was equal across the two samples, six-year-olds demonstrated chance-level responding.

4. Discussion

A fundamental challenge of cognitive development is to establish criteria for evaluating when samples of limited evidence provide strong bases for inductive inferences. The goal of the present research was to examine developmental changes in how people evaluate whether limited samples of evidence provide a good basis for induction. In this study, young children used different criteria than adults for making such judgments. Six-year-olds valued item typicality, whereas older children and adults valued overall sample diversity. These findings support our proposal that young children focus on properties of individual exemplars, whereas adults focus on group-level properties involving relations among exemplars. Nine-year-olds were somewhat in the midst of this transition; they favored diverse samples, except when doing so required selecting atypical exemplars *instead of* typical exemplars.

4.1. Children's evaluation of multiple-exemplar samples

When typicality differed between the diverse and non-diverse samples, six-year-olds chose typical sets (regardless of sample diversity). Based on an examination of only the conditions in which typicality varied across samples, one possible explanation for these findings could be that young children and adults have different approaches to integrating information about typicality and diversity. In other words, individuals of all ages may recognize the value of both typicality and diversity, but when the two factors are pitted against each other, young children are drawn toward typicality (e.g., Heit et al., 2004; see also Osherson et al., 1990). Another related possibility is that young children value typical samples more than adults do because they fail to recognize atypical examples as category members (e.g., Bjorklund & Thompson, 1983), and therefore, do not recognize diverse samples containing atypical members as valuable.

Considering young children's performance across all four of the conditions included in this experiment, however, undermines the possibility that our findings relate only to developmental differences in how individuals integrate sample diversity and typicality. Young children did not select diverse samples even when the level of

typicality was held constant across the two samples. Thus, children appear not to recognize the value of diverse evidence. Across conditions, young children's performance is predicted only by sample typicality; when typicality varies, young children make reliable predictions; when typicality is held constant, they demonstrate chance-level responding.

4.2. Adults' evaluation of multiple-exemplar samples

These findings also highlight an important characteristic of adult inductive reasoning. Specifically, the ability to distinguish the inductive potential of individual exemplars from the inductive potential of a set of exemplars appears to be a critical component underlying adult cognition. Although this distinction has been discussed in prior theoretical work (Heit, 2000), we are aware of no previous study that systematically examined this ability by varying the inductive potential of the individual exemplars in a sample. Thus, this pattern of findings demonstrates that examining cognitive development is a critical method for isolating the processes that contribute to adult cognition (Heit & Hahn, 2001; Lopez et al., 1992).

Nonetheless, these findings do not imply that typicality is unimportant within adults' inductive reasoning. As described by Osherson et al. (1990), adults evaluate the relations among individual pieces of evidence in order to identify the extent to which a given sample covers the relevant category. Often, the sample containing the most diverse examples provides the greatest coverage and is therefore viewed as the strongest sample, as was found in the present study. We selected atypical items that were only moderately atypical, however, and Osherson et al. (1990) indicate that sometimes adults will conclude that two non-diverse but highly typical items provide better coverage than two diverse but highly atypical items. Thus, in general, adults approach sample evaluation by considering how the exemplars in a category relate to one another and the category as a whole.

4.3. The present findings in the context of prior work

We have suggested that differences in methodology underlie the inconsistencies across prior reports with respect to children's diversity-based reasoning. Children succeed on tasks that require matching diverse items to diverse sets (Heit & Hahn, 2001) or reasoning about a limited class of items (Shipley & Shepperson, 2006), but have more difficulty when they are required to evaluate samples for the purpose of making broader generalizations. Future work more directly comparing children's performance on different types of reasoning tasks would help to clarify these discrepancies. Also, given the present findings, it is important to consider that preferences for typical information may have biased young children toward diverse sets in some prior work. For example, Lo et al. (2002) found some evidence of diversity effects among preschool children; however, an informal examination of their items suggests that more typical items were often included in the diverse sets (e.g., children rated a ladybug and an ant as stronger evidence than a ladybug and a beetle; we suspect that the first sample is more diverse *and* more typical than the second sample).

An additional factor that will be important to consider in resolving the discrepancies across reports is content domain. In the present research, we focused on animal kinds for two reasons: to be consistent with much prior research on induction, and because biological kinds are so inductively rich that understanding how young children reason about such categories is central to understanding how they learn about the world. Other studies of diversity-based reasoning in children have focused on artifact categories (Heit & Hahn, 2001; Shipley & Shepperson, 2006), which may be more accessible to young children. However, young children generally demonstrate at least as sophisticated inductive reasoning skills when considering the hidden properties of animals as when considering the properties of artifacts (see Gelman, 2003 for a review). Therefore, we doubt that domain differences are primarily responsible for children's successes and failures on tasks related to diversity-based reasoning. Nonetheless, in future research, it would be helpful to directly examine possible domain effects.

4.4. Why might young children focus on individual exemplars?

The present findings, which document some limitations in young children's inductive reasoning, may seem at odds with other research in cognitive development that has revealed an impressive level of sophistication in young children's category-based induction. For example, young children systematically extend information that they learn about individual category members to other category members, even when confronted with conflicting perceptual information (Gelman & Markman, 1986). Thus, young children consider two sources of information that could be inductively relevant, and systematically select to base their inferences on category membership.

We suggest that young children's strong beliefs about category homogeneity may underlie both their success on these category-based induction tasks and their difficulty with sample diversity. On a number of tasks, young children have been documented to view categories as more strongly determinative of properties than adults do (e.g., Diesendruck & haLevi, 2006; Gelman & Kalish, 1993; Taylor, 1996; Waxman, Medin, & Ross, 2007), such that they predict that category members will share category-linked properties even when adults rely on individuating information. In this way, children appear to ignore within-category variability that adults find important, focusing instead on within-category commonalities. An overall tendency to overlook within-category variability could also explain children's failure to use diversity (e.g., Gutheil & Gelman, 1997). In both cases, young children focus on certain individual exemplars as standing in for the category as a whole. Over development, as children increasingly consider within-category variability, they may begin to recognize the value of sampling from diverse category members before drawing conclusions about a category as a whole.

In sum, the present findings contribute to an ongoing broad research program aimed at identifying the extent to which cognitive development is driven by both knowledge enrichment and conceptual reorganization. Both processes clearly contribute to development, as knowledge differences have been documented to influence children's reasoning on other tasks (e.g., Carey, 1985), as well as adults' tendency to

rely on diversity to guide induction (Medin et al., 2003; Proffitt, Coley, & Medin, 2000). The present work suggests, however, that the mechanisms that support inductive reasoning about natural kinds also undergo reorganization across the elementary school years.

References

- Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory and Cognition*, 15, 397–418.
- Bjorklund, D. F., & Thompson, B. E. (1983). Category typicality effects in children's memory performance: Qualitative and quantitative differences in the processing of category information. *Journal of Experimental Child Psychology*, 35, 329–344.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.
- Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review*, 6, 338–346.
- Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development*, 77, 539–553.
- Faraway, J. J. (2006). *Extending the linear model with R: Generalized linear, mixed effects, and nonparametric regression models*. Boca Raton, FL: Chapman and Hall/CRC.
- 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.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday life*. New York: Oxford University Press.
- Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development* (pp. 3–32). New York: Springer-Verlag.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183–209.
- Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, 74, 308–325.
- Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin and Review*, 7, 569–592.
- Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin and Review*, 12, 340–344.
- Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology*, 43, 243–273.
- Heit, E., Hahn, U., & Feeney, A. (2004). Defending diversity. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Festschrift in honor of Douglas L. Medin* (pp. 87–99). Washington, DC: American Psychological Association.
- Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, 38, 883–894.
- Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory and Cognition*, 31, 155–165.
- 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., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, 32, 251–295.
- Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development*, 63, 1070–1090.
- Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. L. (2003). A relevance theory of induction. *Psychonomic Bulletin and Review*, 3, 317–332.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185–200.

- Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 811–828.
- 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.
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665–681.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192–233.
- Shipley, E. F., & Shepperson, B. (2006). Test sample selection by preschool children: Honoring diversity. *Memory and Cognition*, 34, 1444–1451.
- Taylor, M. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, 67, 1555–1571.
- Viale, R., & Osherson, D. (2002). Cognitive development, culture, and inductive judgment. In R. Viale, D. Andler, & L. A. Hirschfeld (Eds.), *Biological and cultural bases of human inference* (pp. 38–48). Mahwah, NJ: Lawrence Erlbaum Associates.
- Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, 43, 294–308.