COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1561–1573

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01137.x

The Role of Within-Category Variability in Category-Based Induction: A Developmental Study

Marjorie Rhodes,[a] Daniel Brickman[b]

[a]*Department of Psychology, New York University*
[b]*Department of Psychology, The University of Michigan, Ann Arbor*

Received 7 December 2009; received in revised form 4 May 2010; accepted 7 June 2010

Abstract

The present studies tested the hypothesis that strong assumptions about within-category homogeneity impede children's recognition of the inductive value of diverse samples of evidence. In Study 1a, children (7-year-olds) and adults were randomly assigned to receive a prime emphasizing within-category variability, a prime emphasizing within-category similarities, or to not receive a prime. Only following the variability prime, children demonstrated a reliable preference for evaluating diverse over nondiverse samples to determine whether there is support for a category-wide generalization. Adults demonstrated a robust preference for diverse samples in all conditions. These effects extended beyond the specific categories included in the prime, as well as to multiple types of test questions. Study 1b demonstrated that priming variability leads children to select diverse samples only when doing so is informative for induction. Implications for conceptual development are discussed.

Keywords: Induction; Conceptual development; Categories; Sampling; Variability; Natural kinds

1. Introduction

Categories are incredibly powerful cognitive tools, in large part because they allow people to overlook superficial similarities and focus on the properties that individuals share. A critical function of categories is to promote inductive learning (Rips, 1975). Thus, upon learning something about an individual (e.g., that a spider bites), we can generalize this information to the category as a whole (e.g., all spiders) and use this information to guide behavior (e.g., to avoid spiders).

Correspondence should be sent to Marjorie Rhodes, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: marjorie.rhodes@nyu.edu.

Great inductive power is gained by assuming that categories are homogeneous (e.g., assuming that spiders share many observable and unobservable features; Gelman, 1988). Yet strong homogeneity assumptions can also lead us astray. For example, not all spiders bite, and those that do vary in the level of threat that their bites pose. Whereas avoiding all spiders, despite this variability, may be an acceptable strategy, there are clear cases where assuming homogeneity can have problematic consequences. For example, if a child assumes that all dogs are friendly because her pet cocker spaniel is friendly, she may put herself in a dangerous situation if she encounters a pit bull.

The goal of this article is to examine the role of within-category variability in inductive reasoning and how it changes across development. In particular, this study tests the hypothesis that young children consider within-category variability less readily than adults do, leading to systematic developmental differences in inductive reasoning.

Evidence that adults consider within-category variability in category-based induction comes from work on how adults evaluate the inductive potential of various samples. Adults do not view all samples as equivalently informative for drawing inferences about a category as a whole (Feeney & Heit, 2007; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975). Although adults attend to a variety of features for evaluating samples (Medin, Coley, Storms, & Hayes, 2003), a key criterion that they attend to robustly is sample *diversity* (Heit, 2000; Heit, Hahn, & Feeney, 2004). Adults view samples that diversely represent a category (e.g., a Chihuahua and a collie) as more informative than nondiverse samples (two Chihuahuas). This preference for diverse samples indicates an awareness that individuals within categories vary in important ways and that it is useful to sample from across that variability before generalizing to a category as a whole. Adults' preference for diverse samples has been documented in a number of tasks, including their ratings of inductive arguments (Osherson et al., 1990) and their evidence selection (Kim & Keil, 2003; Lopez, 1995; Lopez, Atran, Coley, Medin, & Smith, 1997; Rhodes, Brickman, et al., 2008; Rhodes, Gelman, et al., 2008).

In contrast, there is substantial evidence that children, before age 9, do not value samples that diversely represent categories. For example, in Lopez, Gelman, Gutheil, and Smith (1992), 6- and 8-year-olds were equally likely to extend a property to a category after learning that it was found in a diverse or a nondiverse sample (see also Carey, 1985; Gutheil & Gelman, 1997; Rhodes, Gelman, & Brickman, 2010). In Rhodes, Gelman, et al. (2008) and Rhodes, Brickman, et al. (2008) children younger than age 9 did not reliably choose to create diverse samples of evidence (e.g., they were equally likely to choose to check a robin and an eagle to see whether a property is true of birds as they were to check two robins).

Children's failure to attend to sample diversity in inductive reasoning tasks is meaningful for several reasons. First, it is clear that children's neglect of sample diversity does not relate to problems noticing or processing diversity. A control condition in Rhodes, Gelman, et al. (2008) documented that 6-year-olds could reliably distinguish diverse from nondiverse samples, yet they did not view diverse samples as more informative. Heit and Hahn (2001) also provide evidence that 5-year-olds can recognize and reason about sample diversity (e.g., that children can sort diverse vs. nondiverse samples; see also Shipley & Shepperson, 2006).

Second, children's failure to attend to sample diversity cannot be attributed to general difficulties with statistical reasoning, or in reasoning about the relation between samples and populations (Xu & Denison, 2009; Xu & Garcia, 2008).

The hypothesis considered in this work is that children's failure to recognize the value of diverse samples for induction is conceptual in nature and stems from a great emphasis on within-category homogeneity in early concepts. Put simply, if children have a strong assumption that all dogs are fundamentally the same, why should it matter to them which dogs they observe? From this perspective, what changes across development is how assumptions about category homogeneity are balanced with expectations about within-category variability.

Gelman (2003) has argued that cognitive biases to assume that categories are homogeneous play a powerful role in early conceptual development; they propel knowledge acquisition by allowing children to overlook superficial differences and focus on underlying regularities. Such biases may discourage children from incorporating within-category variability into their concepts (Gelman & Kalish, 1993). From this perspective, although children perceive within-category variability, strong homogeneity biases lead children to exclude variability from their concepts.

A number of previous developmental studies support the hypothesis that children view categories as more homogeneous than adults do. For example, young children have strong expectations that category members will behave in category-consistent manners, even in the face of contrasting individuating information (Berndt & Heller, 1986; Taylor, 1996). Also, preschool children are more likely than adults to believe that their everyday categories reflect an objective natural reality (Kalish, 1998; Rhodes & Gelman, 2009a,b) and to infer that a property observed in one individual will be true of a category (Rhodes & Gelman, 2008).

2. Method, Study 1a

The goal of this study was to test whether failure to consider within-category variability contributes to children's neglect of sample diversity in inductive reasoning. By random assignment, children completed primes that focused them on within-category variability, within-category similarities, or, in a control condition, not to complete priming activities. Next, children completed evidence selection tasks, to test whether exposure to variability increases preferences for evaluating diverse samples before making category-wide generalizations.

Test questions involved the categories that were presented as part of the primes, as well as other categories that were not included in the primes. Thus, this work tests whether the effect of priming variability operates only on the category for which variability information is introduced, or whether effects extend to other similarly structured categories within the domain. This work tested appreciation for two distinct types of diversity: diversity based on sampling locations (e.g., whether a sample of fish drawn from four different lakes is more informative than a sample drawn from a single lake for inferences about all fish) and sampling categories (e.g., whether a sample comprised of a basset hound and a collie is more informative than two collies for inferences about all dogs).

This study included 7-year-old children. Previous work has found that a reliable preference for diverse samples develops around ages 8–9; thus, 7-year-olds should not yet value sample diversity but would perhaps be close enough to developing this preference that it could be elicited given a particular experimental context. For comparison purposes, this work also included adult participants. Because adults have fairly robust preferences for sample diversity, they should reliably select diverse samples across all three conditions, whereas children should do so only following the variability prime.

2.1. Participants

Participants included 57 children (34 males, 23 females; *M* age = 7.03 years, range = 5.64–7.61 years) recruited from first grade classrooms in a public elementary school in a midsize city in the Midwestern United States. Adult participants included 48 college students, recruited from campus locations in the same midsize city, who volunteered to participate in exchange for a $5 gift card.

2.2. Procedures

Children completed the experiment during individual sessions with trained undergraduate research assistants in a quiet area of their school. Adults completed the experiment independently, using a computer program. Participants were randomly assigned to one of three conditions, each containing 19 children and 16 adults: Variability, Similarity, or Control.

In both the Variability and Similarity conditions, the experiment began with exposure to information about an animal category (birds or dogs). In both conditions, children were shown a colorful picture that contained 36 diverse exemplars drawn from the category. Identical stimuli were used across conditions. They then completed an activity with the experimenter, which was designed to draw their attention to all of the exemplars in the set. Children were asked to point to the same number of exemplars in both conditions, and primes were presented with similar levels of enthusiasm by the experimenter.

2.2.1. Primes

2.2.1.1. Variability: In the Variability condition, for the category *birds*, children were told, ‘‘Look at all these birds! See how different they all are? Look. There are a lot of different colors. Point to some that have different colors. Some have big beaks and some have small beaks. Can you point to some with big beaks? And with small beaks? Some fly and some don’t. Can you find one that flies? And one that doesn’t fly? Some are very big and some are very small. Where is one that is very big? Where is one that is very small? Some of them eat small things like worms, and some of them eat big things like mice. Do you see some that eat small things? How about birds that eat big things? Do you see anything else?’’ The activity was comparable for the category *dogs.*

Table 1
Summary of categories, properties, and samples for the test questions

Task Type	Category	Property	Diverse Sample	Nondiverse Sample
Sampling locations	Birds	Have hollow bones	One bird from each of four mountains	Four birds from one mountain
	Dogs	Have four-chamber hearts	One dog from each of four towns	Four dogs from one town
	Monkeys	Have spleens inside	One monkey from each of four jungles	Four monkeys from one jungle
	Fish	Have opercula inside	One fish from each of four lakes	Four fish from one lake
Sampling categories	Frogs	Have a carapace inside	One green tree frog; one orange frog	Two green tree frogs
	Turtles	Have fused tailbones	One box turtle; one large turtle	Two box turtles
	Cats	Have papillae	One black cat; one orange cat	Two black cats
	Pigs	Have prenasal bones	One pink pig; one pink/black pig	Two pink pigs
	Dogs	Are dichromatic	One golden retriever; one black Labrador	Two golden retrievers
	Birds	Are endothermic	One cardinal; one blue jay	Two cardinals

2.2.1.2. Similarity: In the Similarity condition, for the category *birds*, children were told, ''Look at all these birds! See how many there are? Look. They all have feathers. Can you point to some feathers? They all have beaks. Point to their beaks! They all have babies by laying eggs. Can you point to some that have babies by laying eggs? They all have little scales on their feet. Do you see the scales on their feet? And they all feed their babies by putting food right in their mouths! Point to some that feed their babies that way. Do you see anything else?'' The activity was comparable for the category *dogs*.

2.2.2. Order and test questions

In both the Variability and Similarity conditions, children first completed a prime (for either the category *dog* or the category *bird*). Then, they were asked a sampling-locations task (see below) about the category presented in the prime, followed by a sampling-locations task about one new animal category (either *monkeys* or *fish*; see Table 1). Subsequently, they completed another priming activity for whichever category they had not completed initially, followed by a sampling-locations task about the category in this prime and a sampling-locations task about the other new category. Next, children completed six sampling-categories questions (see Table 1), including four new categories (cats, pigs, frogs, and turtles) and the two primed categories (dogs and birds). In this block, the questions about the primed categories were always asked last, so that response strategies used for these categories would not influence how children responded to the new categories. The following factors were counter-balanced across participants: whether the first priming activity was for *birds* or *dogs*, whether the first new sampling-locations question was for *monkeys* or *fish*, the order of the four sampling-categories questions involving new animal categories, and the order of the sampling-categories questions for the primed categories.

2.2.2.1. Sampling locations: These questions followed Rhodes, Gelman, et al. (2008). Children were shown four scenes (e.g., four mountain tops), and were told, for example, ''Let's pretend that you are a scientist who is trying to find out something new about birds. You are trying to find out if birds have hollow bones. But you can't look at all the birds in the world to find out if birds have hollow bones; you can only look at four birds—just four. Here are some mountains. There are birds that live on each of these mountains. So, to find out if birds have hollow bones, do you want to look at four birds from one mountain, or one bird from each mountain? Which birds should you look at to find out if birds have hollow bones?'' The order of the answer choices was counterbalanced across participants.

2.2.2.2. Sampling categories: These questions followed Rhodes, Brickman, et al. (2008) (see Table 1). Children were shown two sets of animals from the same basic-level category: one diverse sample (e.g., a black Labrador and a golden retriever), and one nondiverse sample (e.g., two golden retrievers). Children were asked, for example, ''Pretend you're a scientist studying dogs. Your job is to find out if dogs are dichromatic. But you can't look at all of the dogs in the world to find out about dogs. You can only look at two dogs, just two. Which dogs do you want to look at to find out about dogs?'' Children responded by pointing to the diverse or nondiverse set. The lateral position of the diverse sample was counterbalanced across questions.

3. Results

Analyses were conducted using a generalized linear model, with a binomial probability distribution and a logit link function. The dependent variable was the number of diverse-sample selections out of the total possible for each type of question. Age and condition were entered as fixed factors, and analyses test for main effects of each variable, as well as for an interaction. Wald chi-square statistics were generated as indicators of the significance of these predictor variables. Descriptive statistics are presented as proportions of diverse-sample selections. These data, with 95% Wald confidence intervals, are presented in Figs. 1 and 2. In these figures, values with error bars that do not overlap the line marking .50 (the proportion of diverse responses expected by chance) indicate that average responses differed from equal probability responding.

3.1. Sampling locations

Examining the four sampling-locations questions revealed an effect of Age, $\chi^2(1) = 41.81$, $p < .001$, and an interaction between Age and Condition, $\chi^2(2) = 7.61$, $p = .02$. Children selected diverse samples more often in the Variability condition than in either the Similarity or Control conditions (compared to similarity, $p < .001$, compared to control, $p = .01$), whereas adults' responses did not vary by condition (see Fig. 1A,B). Relative to the Control condition, the variability prime doubled the likelihood of selecting a

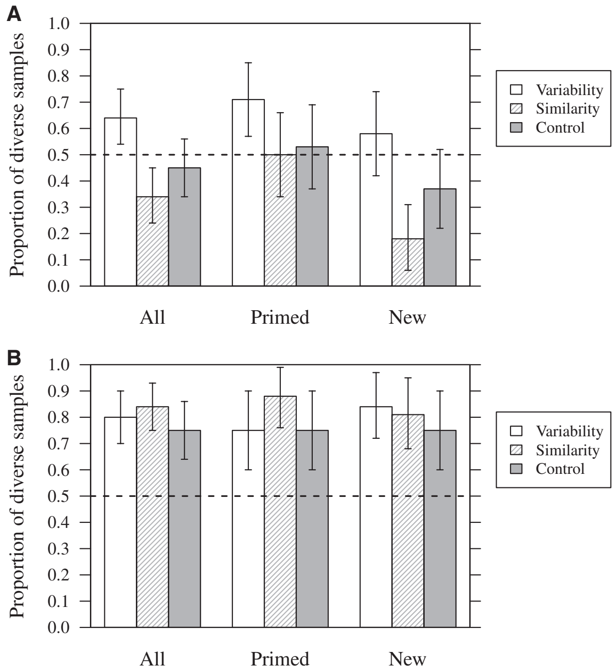

Fig. 1. Proportions of questions on which children (A) and adults (B) selected diverse samples on sampling-location questions. *Note.* Error bars represent Wald 95% confidence intervals. Error bars that do not overlap with the line marking .50 indicate that the mean for that condition is reliably different from the proportion expected by chance.

diverse sample (OR = 2.24; 95% CI: 1.17, 4.30). The same pattern was found for the categories included in the prime and the new categories; for both sets of categories, children selected diverse samples more often in the Variability condition than in either of the other two conditions (for some effects, these comparisons were marginally significant, perhaps due to the small number of questions asked when these items are examined separately, all *p*s < .09).

As shown in Fig. 1A, for sampling-locations questions, children selected diverse samples more often than expected by chance in the Variability condition, for the analysis of the primed categories and total categories. For the new categories, children's selections of diverse samples did not exceed the level expected by chance, and children in the Similarity condition reliably selected nondiverse samples. Children's responses in the Control condition did not differ from chance for any question type.

3.2. Sampling categories

For responses to the six sampling-categories questions, there were significant effects of Age, $\chi^2(1) = 64.69$, $p < .001$, Condition, $\chi^2(2) = 5.92$, $p = .05$, and an interaction between Age and Condition, $\chi^2(2) = 12.65$, $p = .002$. As shown in Fig. 2A,B, children selected diverse samples more often in the Variability condition than in either the Similarity or Control conditions, *p*s < .001, whereas adults' responses did not vary by condition. Relative to the Control condition, the variability prime increased the likelihood of selecting a diverse sample by 4.12 (CI: 2.34, 7.28). The same pattern was found for the categories included in the prime as for the new categories; in each case, children selected diverse samples more often in the variability condition than in either other condition, all *p*s < .01.

For each type of category-sampling question, children selected diverse samples more often than expected by chance only in the Variability condition, whereas adults reliably selected diverse samples in each condition. Children reliably preferred nondiverse samples in the Similarity condition for the primed categories; no other pattern of responses reliably differed from chance.

3.3. Individual response patterns

Analyses of individual response patterns revealed similar patterns as were found for the group means; these analyses are presented in the Appendix.

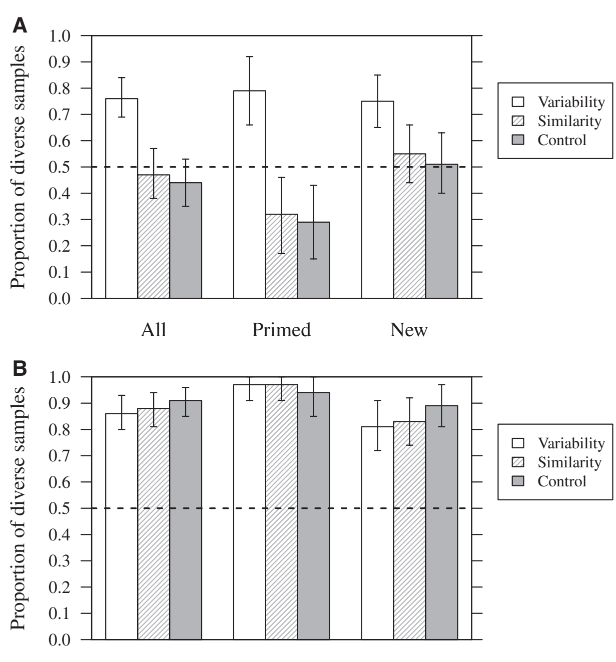

Fig. 2. Proportions of questions on which children (A) and adults (B) selected diverse samples on sampling-categories questions. *Note.* Error bars represent Wald 95% confidence intervals. Error bars that do not overlap with the line marking .50 indicate that the mean for that condition is reliably different from the proportion expected by chance.

4. Study 1b

The results from Study 1a may indicate that, as hypothesized, the variability prime led children to recognize the informative value of diverse samples. Another possibility, however, is that the variability prime led to a generalized belief that diverse samples are more interesting or important, thus accounting for increased selections of diverse samples without a particular recognition of their inductive value.

To evaluate these possibilities, in Study 1b, a new sample of children ($n = 20$) completed the two variability primes from Study 1a. Subsequently, children were asked six questions identical to the sampling-categories questions asked in Study 1a (see Table 1), with the exception that half of the children were asked to find out whether properties applied to specific subtypes that were included in the samples (e.g., an additional picture of a golden retriever was shown, and children were asked ''to find out if *this kind* of dog is dichromatic''), whereas the other half were asked questions in an identical manner to Study 1a (e.g., ''to find out if *dogs* are dichromatic''). In both conditions, children could choose between the two samples offered in the sampling-categories questions of Study 1a (see Table 1). If children reliably select diverse samples for the entire category questions, but not the subtype questions, this would suggest that the variability prime leads children to use diversity in a systematic manner—only when it makes a sample more informative for the question at hand.

Results indicated that children in the subtype condition showed no reliable preference for diverse samples ($M = 0.53$, CI: 0.41, 0.66), whereas consistent with Study 1a, children in the whole-category condition did ($M = 0.70$, CI: 0.58, 0.82). These data suggest that the variability prime led children to view diverse samples as more informative for category-wide inferences, not to a generalized belief that diverse samples are more interesting or important.

5. Discussion

In these studies, priming 7-year-old children with within-category variability increased their appreciation of the inductive power of diverse samples. Thus, whether individuals account for within-category variability importantly influences inductive reasoning. Interestingly, priming within-category homogeneity did not influence children's responses (as children in the Similarity and Control conditions did not differ), supporting the proposal that children's everyday concepts emphasize underlying regularities. These findings support the proposals that young children's emphasis on within-category homogeneity impedes their recognition of the value of diverse samples for induction, and that what develops across childhood is increased incorporation of within-category variability into children's concepts.

Priming within-category variability increased selections of diverse samples across two types of questions (sampling locations and sampling categories) targeted toward category-wide generalization, but not for questions targeted toward learning about a specific subtype. Thus, priming within-category variability led children to value diverse samples in a systematic and coherent manner. Although similar effects were found across the two sampling tasks, effects were larger for sampling-categories than sampling-locations questions. This could be either due to the order of the test trials, or perhaps to simpler response demands for the sampling-categories questions (e.g., sampling-categories questions asked children to point to one of two visually presented samples, whereas sampling-locations questions asked children to point to each of four locations or four times to one location).

There are at least two processes by which the primes presented in this study could have led to increased appreciation for sample diversity. First, the primes could have functioned to increase children's knowledge. In this manner, the primes could be considered informational, such that children were unaware of the variability that existed among birds, for example, and the prime makes them aware of it. From this perspective, what is acquired across development would be an increase in children's specific knowledge about animal categories (see Proffitt, Coley, & Medin, 2000). Alternately, the primes could have served to challenge children's abstract beliefs (or their ''framework theories'') about the structure of natural kind categories. From this perspective, exposure to meaningful variability within the category *bird* challenges children's general belief that animal categories are homogeneous, such that the effect of the prime is on the conceptual structure of the domain of animals.

There are several reasons to favor this second explanation. In these data, children's preference for diverse samples extended beyond categories included in the prime. Importantly, this study was designed to prevent the possibility that children could recognize the value of diverse samples for only the primed categories, and then carry over their response strategy to the new categories. On the sampling-categories questions, which provided the strongest evidence of generalization to new categories, the new categories were always presented first, before questions about the primed categories. The sampling-categories questions were in a different format, used different stimuli, and required different forms of responses than the sampling-locations questions or the activities in the prime. Thus, children's valuing of sample diversity on these items could not result from a carryover of a response strategy developed for the primed categories.

Furthermore, previous work suggests that children's neglect of within-category variability does not relate to ignorance about variation. Instead, children appear to acknowledge variation, but they view it as inconsequential for their understanding of the category as a whole. For example, in Gelman and Raman (2003), 4-year-olds were shown a picture of two penguins and asked, ''Do *these birds* fly?'' Children responded with ''no,'' indicating that they were aware of this category anomaly. When children were shown the same picture, however, and asked, ''Do *birds* fly?'' they reliably responded with ''yes,'' indicating that their awareness of the category anomalies did not influence their beliefs about the category as a whole. Thus, what appears to change across development is not only increased knowledge of variation, but increased appreciation that this variability has important implications. As prior work suggests that children are on the cusp of developing this appreciation in middle childhood (ages 7–9), an open question is whether priming variability would also increase preferences for diverse samples among younger children.

Overall, the present findings and previous work are consistent with the interpretation that exposure to within-category variability challenges children's abstract beliefs about the nature of animal categories. Previous work has demonstrated that children hold such abstract beliefs about animal categories that extend beyond their knowledge of specific animals.

Rhodes and Gelman (2009b) found that children expect animal categories to have absolute boundaries (in contrast to the boundaries of artifact categories, which they view as graded), and that they applied these expectations to both novel and familiar animals. Additionally, Brandone and Gelman (2009) found that children generated generic noun phrases to describe fictional animal, but not artifact, categories (e.g., after exposure to a single new ‘‘modie’’ with many eyes, children generated sentences like ‘‘Modies have a lot of eyes’’). Thus, they assumed that new animal categories would have high levels of coherence, even in the absence of any previous specific knowledge.

The present proposals indicate that children should more readily appreciate sample diversity for artifact categories than for natural kinds, due to domain differences in their assumptions about within-category homogeneity. Interestingly, two previous sets of studies examining diversity-based reasoning with artifact categories both produced more positive assessments of children’s abilities (Heit & Hahn, 2001; Shipley & Shepperson, 2006). However, the methods used in these studies were quite different from methods used to examine reasoning with natural kinds, and they may not tap the same component of diversity-based reasoning (see discussion in Rhodes, Gelman, et al., 2008). Thus, future work should directly compare diversity-based induction with animal and artifact categories.

The present findings contribute to our understanding of developmental changes in induction across childhood. Whereas children demonstrate many adult-like strategies in their category-based induction (e.g., basing inference on conceptual categories over superficial perceptual properties, Gelman & Markman, 1986), it may be that the same early-emerging biases that enable the early development of category-based induction (e.g., biases to focus on underlying regularities) also serve to prevent children from recognizing the value of sample diversity. Future work should examine the developmental trajectory of how children incorporate within-category variability into their concepts across development.

Acknowledgments

We thank Susan Gelman for her input into the study design; the schools, teachers, and families who participated; and Sam Kolkey, Amanda Karp, and Joseph Dietzel for their assistance with data collection. Funding was provided by a Rackham Predoctoral Fellowship from the University of Michigan and by New York University.

References

Berndt, T. J., & Heller, K. A. (1986). Gender stereotypes and social inferences: A developmental study. *Journal of Personality and Social Psychology*, *50*, 889–898.

Brandone, A., & Gelman, S. A. (2009). Differences in preschoolers’ and adults’ use of generics about novel animals and artifacts. A window onto a conceptual divide. *Cognition*, *110*, 1–22.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.

Feeney, A., & Heit, E. (2007). *Inductive reasoning: Experimental, developmental, and computational approaches*. New York: Cambridge University Press.

Gelman, S. A. (1988). The 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 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*, *24*, 308–325.

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.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin and Review*, *7*, 569–592.

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.

Kalish, C. W. (1998). Natural and artificial kinds: Are children realists or relativists about categories? *Developmental Psychology*, *34*, 376–391.

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory and Cognition*, *31*, 155–165.

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., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition*, *108*, 543–556.

Rhodes, M., & Gelman, S. A. (2008). Categories influence predictions about individual consistency. *Child Development*, *79*, 1271–1288.

Rhodes, M., & Gelman, S. A. (2009a). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*, 294–274.

Rhodes, M., & Gelman, S. A. (2009b). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review*, *16*, 920–924.

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). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681.

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.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*, 97–104.
Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 5012–5015.

Appendix

Individual response patterns, by age and condition, Study 1

	Diverse	Nondiverse	Other
Children			
Variability	8	0	11
Similarity	2	0	17
Control	1	2	16
Adults			
Variability	8	0	8
Similarity	8	0	8
Control	8	0	8

Note. According to the binomial theorem, consistently selecting diverse or nondiverse samples across all six questions would be unlikely to result from chance alone, $p = .03$. Thus, we identified the number of participants displaying a consistent response strategy across all trials, and all other participants were coded as ''other.'' Fisher's exact tests confirmed that the distribution of codes varied by condition for children, $p = .01$, but not for adults.