THINKING & REASONING, 2009, 15 (4), 388–404

Psychology Press
Taylor & Francis Group

The law of large numbers in children's diversity-based reasoning

Fuhong Li, Bihua Cao, Yiyuan Li and Hong Li
Southwest University, ChongQing, and Ministry of Education, China

Gedeon Deák
La Jolla, CA, USA

Adults increase the certainty of their inductive inferences by observing more diverse instances. However, most young children fail to do so. The present study tested the hypothesis that children's sensitivity to instance diversity is determined by three variables: ability to discriminate among instances (*Discrimination*); an intuition that large numbers of instances increase the strength of conclusion (*Monotonicity*); ability to detect subcategories and evaluate numerical differences between the subcategories, or *Extraction*. A total of 219 Chinese children aged 6 to 11 were tested for sensitivity to diversity by means of Discrimination, Monotonicity, and Extraction. The results indicated that children at all ages were able to discriminate instances and attend to set size. However, only 9- and 11-year-olds demonstrated Extraction and sensitivity to diversity. Furthermore, among all children diversity scores increased linearly with the level of Extraction. These results suggest that the law of large numbers plays a role in children's diversity-based reasoning.

Keywords: Children; Classification; Cognitive development; Diversity; Inductive reasoning; Large numbers; Quantitative reasoning.

Induction is the process of making inferences that extend beyond the available evidence under uncertain conditions. For example, upon learning that a dog has a particular unobserved biological property, one can extend

Correspondence should be address to Hong Li, School of Psychology, Southwest University, ChongQing, China, 400715. E-mail: lihongwrm@vip.sina.com

This work was supported by Key Discipline Fund of National 211 Project, China. (NSKD08014). Thanks to Sun Hongjin for his modification of language of the initial version of this manuscript. Thanks to Tang yan and Liu Diyuan for their help in collecting data.

http://www.psypress.com/tar DOI: 10.1080/13546780903343227

this knowledge to other dogs, and possibly to other mammals. In this sense, induction allows us to generalise knowledge and make predictions about the world. Nevertheless, inductive inferences are never certain. People are more certain about inductive inference supported by diverse examples (Lopez, 1995; Rhodes, Brickman, & Gelman, 2008; Sloman & Lagnado, 2005). If we wish to examine the health of dogs in a village, for example, it would be better to inspect samples of dogs from various regions of the village than to inspect a large number of dogs from just one region. When induction involves a deliberate search for or preference for diverse examples, we call it diversity-based reasoning (Heit & Hahn, 2001).

Diversity-based reasoning has been explained in different ways. Osherson, Smith, Wilkie, Lopez, and Shafir's (1990) similarity-coverage model postulated not only similarity-based evidence selection, but also *coverage*-based selection. Consider the following example:

(1) Hippopotamuses require Vitamin K for the liver to function.
Rhinoceroses require Vitamin K for the liver to function/
So all mammals require Vitamin K for the liver to function.
(2) Hippopotamuses require Vitamin K for the liver to function.
Hamsters require Vitamin K for the liver to function/
So all mammals require Vitamin K for the liver to function.

Osherson et al. claimed that hippopotamuses and rhinoceroses are no more similar to the class "mammal" than hippopotamuses and hamsters. Thus, similarity alone cannot account for adults' preference for the latter. Also important is the comparative "coverage" of the conclusion category by different premise sets. Because small mammals (e.g., hamsters; squirrels) are not covered by the first premise, the coverage of {hippopotamuses, rhinoceroses} is narrower than {hippopotamuses, hamsters}. Consequently, the second argument is stronger. Sloman (1993) explained this diversity effect in terms of feature overlap. He suggested that when premise categories differ from each other their features have relatively little overlap, and thus cover a larger part of feature space. Conversely, when premise categories are similar their coverage of the feature space is smaller. Heit (2000), by contrast, suggested that an argument with two similar premise categories, such as cows and horses, might activate idiosyncratic common properties of these premises (e.g., having hooves; living on farms). This makes it plausible that, by analogy, the shared unknown property of cows and horses is also idiosyncratic to a subset of animals, but not general throughout the superordinate category. In contrast, two diverse premise categories, such as horses and mice, are less likely to share idiosyncratic properties that are not general throughout the superordinate category. By this account, coverage will be greater when idiosyncratic properties are less available or unknown.

The present study addresses a fourth possibility, that diversity-based reasoning increases the sufficiency of premises for projecting a property because it capitalises on a reasoning heuristic: the law of large numbers. This refers to the empirical law of large numbers, not the mathematical law of large numbers (Sedlmeier & Gigerenzer, 1997). Nisbett, Krantz, Jepson, and Kunda (1983) suggested that people's judgements are affected by subordinate category size because they use a sample size heuristic. They realise that larger samples are more representative of the population from which they are drawn than are smaller samples. A related possibility is that people infer that a more numerous subordinate category leaves a smaller proportion of the superordinate category unaccounted for than a small subordinate category; hence the larger subordinate class is a better basis for inductive inference to unknown members of the superordinate class. There is evidence that adults are biased to believe that large samples are more reliable than small samples for concept formation and generalisation (Fong & Nisbett, 1986; Jepson, Krantz, & Nisbett, 1983). Jepson et al. (1983) showed that people can use statistical heuristics such as the law of large numbers in solving particular kinds of problems in particular domains. In one problem, for example, the protagonist should judge characteristics of a lottery. As expected, the great majority of the answers were statistical answers; that is, they incorporated intuitive notions of the law of large numbers. Jepson et al. also argued that these intuitive statistical concepts are learned through repeated exposure to the law of large numbers across domains during development. Interestingly, a brief training session on this reasoning principle was shown to significantly enhance people's use of law of large numbers (Fong & Nisbett, 1986), suggesting that the statistical heuristic is sufficiently available to be triggered or reinforced by training or cueing.

The present experiment attempts to determine whether the law of large numbers plays a role in the diversity-based reasoning of children as well as adults. According to the law of large numbers, the diversity effect can be explained by the fact that the more diverse the premises, the more category members they represent, and hence the stronger the conclusion. For example, hippopotamuses and rhinoceroses are two kinds of big animals with many shared features. By contrast, hippopotamuses and hamsters are very different kinds of animals, one big and the other small, one an ungulate and the other a rodent, one wild and the other domesticated, etc. It can be inferred that hippopotamuses and rhinoceroses represent only one subcategory of animals, whereas hippopotamuses and hamsters represent two subcategories. Accordingly, a larger numbers of animals is represented, or covered, by the diverse premise (hippopotamuses and hamsters). This view is similar to the view of Osherson et al. (1990), but Osherson et al. merely implied that the diversity effect might relate to the estimated relative number of superordinate category members represented by the premises—that is, to extraction. However, this idea was not tested empirically. Therefore, one goal of the present study is to test Osherson et al.'s claim that diversity depends on extraction in children of different ages.

More generally, this study addresses the controversy over whether children use diversity-based reasoning at all for induction. Some studies suggested that younger children show diversity effects (Heit & Hahn, 2001; Lo, Sides, Rozelle, & Osherson, 2002; Shipley & Shepperson, 2006) whereas others found that children have difficulty in diversity reasoning (Carey, 1985; Gutheil & Gelman, 1997; Lo et al., 2002; Lopez, Gelman, Gutheil, & Smith, 1992; Rhodes et al., 2008). Because of this controversy, our second goal is to further test whether younger children show diversity-based reasoning, and possibly to explain children's difficulties using diversity information for induction.

We hypothesised that a factor in mature diversity-based inductive reasoning is applying the law of large numbers. The development of acquiring and using this law might be related to the development of diversity-based reasoning. Specifically, we predicted that children's diversity-based reasoning is determined by at least three variables (Figure 1). The first is the ability to discriminate perceptual and/or conceptual differences between premise objects or subclasses (*Discrimination*). We predict that this is a necessary precondition of diversity reasoning. If children ignore perceptual difference between premise objects, they will fail to detect diversity differences. The second variable is an intuition that a larger number of premise subclasses increases the strength of the conclusion (*Monotonicity*). The last variable is the ability to estimate subcategory numerosity differences between diverse premises and non-diverse premises (*Extraction*). We expect younger children to show less diversity-based

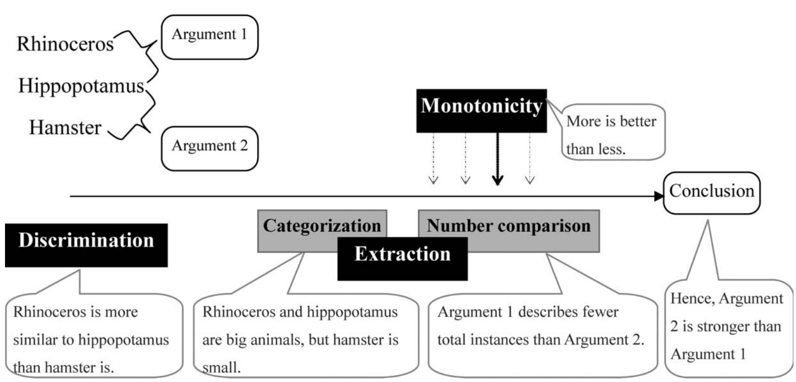

Figure 1. Variables and the proposed process of diversity-based reasoning.

reasoning than older children, and that this trend will be related to some of these three variables.

Finally, we predicted that diversity-based reasoning is different for basic and superordinate category levels, because children more easily learn basic-level categories and know more about these categories than superordinate categories (Gelman & O'Reilly, 1988; Mervis & Crisafi, 1982; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). For this reason, there might be differences in diversity reasoning for basic- versus superordinate-level categories. Moreover, these differences might be due to discrimination or extraction.[1] That is, if children are better able to differentiate, and/or estimate the numerosity of, subcategories of basic-level categories than superordinate-level categories, these differences might contribute to differences in diversity-based reasoning about basic- or superordinate-level categories.

METHOD

Participants

Preschool and elementary school children were recruited from a city kindergarten and a primary school in ShengZhen, a middle-sized city in GuangDong province of China. There were 46 six-year-olds ($M = 6$ years, 1 month (6;1), range = 5;3 to 6;6), 52 seven-year-olds ($M = 7;1$, range = 6;6 to 8;0), 56 nine-year-olds ($M = 9;2$, range = 8;1 to 10;2), and 65 eleven-year-olds ($M = 11;3$, range = 10;5 to 13;0). Approximately equal numbers of boys and girls participated in each age group (112 boys, 107 girls).

Design

Children were exposed to (a) four problems testing their response to Diversity, (b) four problems testing their ability of Discrimination, (c) four problems testing their intuition of Monotonicity, (d) four problems testing their ability of Extraction. Thus children answered a total of 16 questions. Question order was random except that Diversity problems were followed by corresponding Discrimination problems.

There were two conditions for the Diversity and Extraction problems. In one condition the conclusion object was at the basic category level, such as "dog" (condition B). In the other condition the conclusion object was at the superordinate category level, such as "animal" (condition S). Participants

[1]Note that we did not compare basic- versus superordinate-level differences in monotonicity, because that ability is based simply on comparing the number of *given* arguments in each subcategory.

were randomly assigned, half to each condition. The same Monotonicity problems were used in each condition (Table 1).

Materials

Colour pictures were used as visual stimulus for each animal, person, or object mentioned in an argument. The relative sizes of the pictured objects were proportional to the sizes of the real objects. For example, the picture of a tiger was much bigger than the picture of a rabbit.

Each problem (except for the Discrimination problems) was presented as a brief story, consisting of a pair of contrasting arguments: one supposedly strong and one supposedly weak (see Figures 2–5). The materials for the Discrimination test were the same as those of the Diversity test. For example, if the premises of the Diversity test were (cow, horse) and (cow, squirrel), then the prior Discrimination question was: "Please look at these three animals. Which animal do you think is similar to the cow? Is the horse [more] like the cow, or is the squirrel like the cow?" The materials in the Diversity, Monotonicity, and Extraction tests were different from each other, in order to minimise interference effects. Otherwise, the correlation between the extraction and diversity scores might have been artificially increased due to consistent responding to the same context.

Participants were provided with booklets describing a detective and showing a picture of the detective. After the task was explained to children they completed two familiarisation problems to ensure that they understood the task. Familiarisation trial data were not included in the analyses. Children were then given the test problems. After hearing each problem they responded by circling their preferred answer.

Procedure

Prior to the experiment, 10 adults were asked to answer the 16 questions individually. Their answers were consistent with our expectations. Most adults (90%) chose the more diverse premise in the Diversity and Extraction test; all of them were able to discriminate the perceptual difference of premise objects. All chose the premise with larger sample size in the Monotonicity test.

Two experimenters tested children in groups of six to ten in 20–30-minute sessions. The children were told that they would take part in an interesting test consisting of stories. They were asked to listen to each story and answer questions about it. They were first shown a picture corresponding to each story and asked to name the depicted animals or objects. If they could not name it, or named it incorrectly, the experimenter corrected it and had the children repeat the correct name. Then they were tested formally.

Table 1
Items used across tasks and conditions

		Items			
Task	*Condition*	*Test 1*	*Test 2*	*Test 3*	*Test 4*
Diversity & Discrimination	Basic	Bigger orange	Red candy	Plum in round shape	Red chocolate
		Biggest orange	Shallow red candy	Plum in oval shape	Carmine chocolate
		Small orange	Yellow candy	Shrunken plum	Yellow chocolate
	Superordinate	Cow	Fugui bird	Cat	Eagle
		Horse	Sparrow	Dog	Vulture
		Squirrel	Eagle	Elephant	Blue sparrow
Extraction	Basic	Red apple	Pear in round shape	Red bag	Brown toy
		Fuchsia apple	Pear in oval shape	Shallow bag	Shallow brown toy
		Yellow apple	Puffy pear	Yellow bag	Yellow toy
	Superordinate	Tiger	Bailing bird	Eagle	Dog
		Leopard	Mountain sparrow	Vulture	Wolf
		Rabbit	Mountain eagle	Sparrow	Buffalo
Monotonicity	–	5 children	5 dogs	5 toy rabbits	5 birds
		2 children	2dogs	2 toy rabbits	2 birds

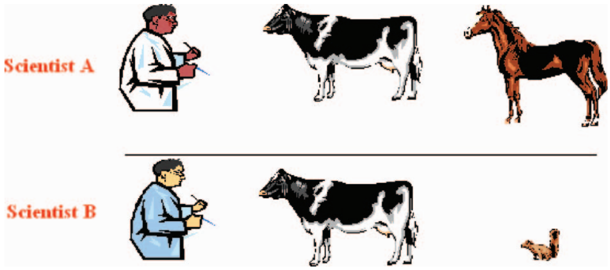

Figure 2. Sample test problem for Diversity. Translation of text: "There are lots of animals in a far-away zoo. Scientist A and Scientist B want to know what substance is in the bones of all animals. Scientist A checked the bones of a cow and a horse and found Tincide, so Scientist A concluded that all the animals have Tincide in their bones. Scientist B checked the bones of a cow and a squirrel and found Tincide, so Scientist B concluded that all the animals have Tincide in their bones. Which scientist do you believe more? [Is it] scientist A or scientist B?" (See Appendix for Pingyin version).

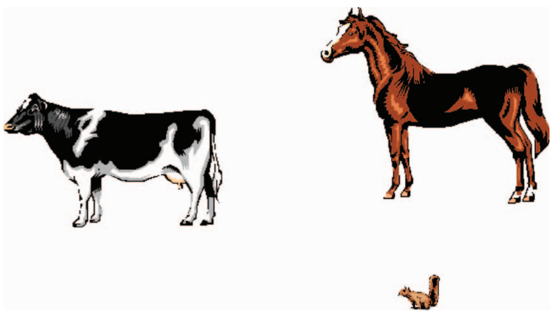

Figure 3. Sample test problem for Discrimination. Translation of text: "Please look at these three animals. Which animal do you think is similar to a/the cow? Does a/the horse seems like a/the cow, or does a/the squirrel seems like a/the cow?" (See Appendix for Pingyin version).

Each child got four scores (range = 0 to 4) for the Diversity, Discrimination, Monotonicity, and Extraction tests. A score of 2 would be expected by chance.

For each diversity problem the children got one point for choosing the argument with a more diverse premise. For each Discrimination problem the children got one point for choosing the test object that was more similar to the target object. For Monotonicity, the children got one point for choosing the argument with much premise items. For Extraction, the children got one point for choosing the argument with a more diverse premise.

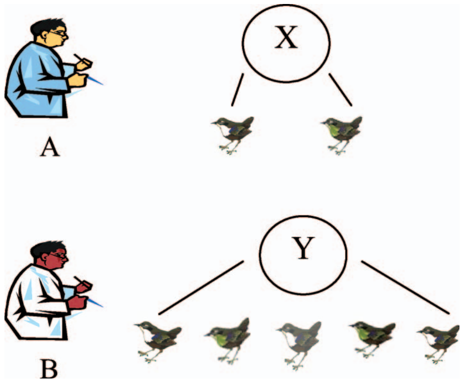

Figure 4. Sample test problem for Monotonicity. Translation of text: "There are lots of birds in a far-away zoo. Scientist A and Scientist B want to know what substance is inside the body of birds. Scientist A checked two birds and found X, so Scientist A concluded that all the birds have X inside their body. Scientist B checked five birds and found Y, so Scientist B concluded that all the birds have Y inside their body. Which scientist do you believe more? Is it scientist A or scientist B?" (See Appendix for Pingyin version).

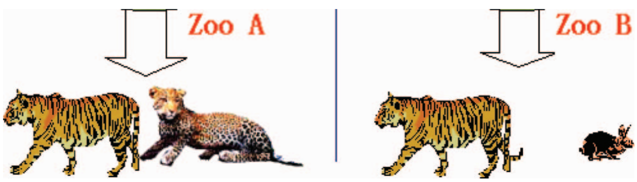

Figure 5. Sample test problem for Extraction. Translation of text: "The kingdom will hold a meeting. The king asked each zoo send two animals for meeting. Zoo A sent a tiger and a leopard, Zoo B sent a tiger and a rabbit. Which two animals represented more members of their zoo? Was it the tiger-leopard pair or tiger-rabbit pair? (See Appendix for Pingyin version).

RESULTS

The data of the adult participants are not presented together with those of the children because the adult scores were at ceiling in all conditions, and because the children were assigned to only one condition whereas adults completed both the basic and superordinate conditions. Consequently, the analyses below include only the children's results.

The average scores of each test for each age group across category levels are shown in Table 2 and Figure 6. The average Discrimination and Monotonicity scores were significantly above chance in each age group. However, the Diversity and Extraction scores were significantly above chance only for 9- and 11-year-olds. Furthermore, the age-related trend in

Table 2
Mean scores (*SD*) and *t*-test results (2-tailed) for all four tests, averaging over category level (basic and superordinate)

Score type		*6-year-olds* ($N = 46$)	*7-year-olds* ($N = 52$)	*9-year-olds* ($N = 65$)	*11-year-olds* ($N = 56$)
Discrimination	*M* (*SD*)	3.78 (.51)	3.69 (.61)	3.75 (.53)	3.63 (.65)
	t-value	23.58	19.95	26.62	18.75
	p	.000	.000	.000	.000
Diversity	*M* (*SD*)	2.22 (1.03)	1.54 (.87)	2.49 (1.12)	3.04 (1.13)
	t-value	1.43	−3.81	3.55	6.87
	p	.156	.000	.001	.000
Extraction	*M*(*SD*)	2.00 (1.26)	1.88 (1.26)	2.52 (1.03)	2.59 (1.01)
	t-value	.000	−.66	4.09	4.39
	p	1.000	.511	.000	.000
Monotonicity	*M*(*SD*)	2.80 (1.05)	2.88 (1.08)	3.62 (.65)	3.46 (.83)
	t-value	5.22	5.92	19.91	13.20
	p	.000	.000	.000	.000

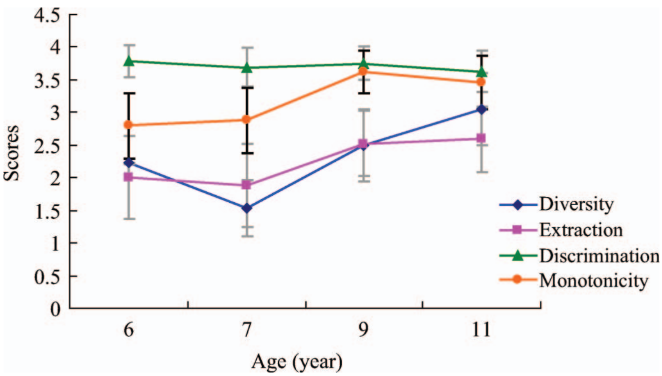

Figure 6. The mean scores on all tests for children of each age. Error bars are *SD*s.

diversity score was paralleled by Extraction scores, and absolute scores were similar at each age. The Spearman product moment correlation between Diversity and Extraction scores was low-moderate but significant, $r = .251$, $p < .001$. Diversity scores were not significantly correlated with either discrimination scores ($r = .034$) or monotonicity scores ($r = .126$). Even when age was controlled, the correlation between diversity and extraction remained significant, $r = .205$, $p = .002$.

Multivariate linear stepwise regression analyses were performed, with the diversity score as the dependent variable. Explanatory variables were Extraction, Discrimination, and Monotonicity scores. In the first step, only

Extraction was added as an explanatory variable, and explained 7.0 % of the variance. The beta of Extraction was 0.26. Adjusted $R^2 = 0.07$; $F(1, 217) = 16.23$, $p < .001$.

The average scores for each category level (basic; superordinate) at each age group are shown in Table 3. Mean Diversity scores for basic-level categories were significantly greater than chance in 9- and 11-year-olds. However, scores for superordinate-level categories were marginally greater than chance only in 11-year-olds. Mean Extraction scores for basic-level categories were significantly greater than chance in 9- and 11-year-olds, but for superordinate-level categories they did not differ from chance in any age group. Monotonicity scores were significantly greater than chance at both category levels in all age groups.

Two-way ANOVAs involving age (6, 7, 9, and 11 years) and category level (basic versus superordinate) were performed for Diversity. Main effects

Table 3
Mean scores (*SD*) and *t*-test results (2-tailed) for all four tests and age groups, separated by category level (basic; superordinate)

Test		*6-year-olds*	*7-year-olds*	*9-year-olds*	*11-year-olds*
Basic category level					
Discrimination	*M*(*SD*)	3.86 (.47)	3.89 (.32)	3.75 (.51)	3.82 (.39)
	t-value	18.70	30.65	19.49	24.71
	p	.000	.000	.000	.000
Diversity	*M(SD)*	2.23 (1.07)	1.41 (.84)	2.75 (1.16)	3.64 (.78)
	t-value	1.00	−3.65	3.65	11.15
	p	.323	.001	.001	.000
Extraction	*M*(*SD*)	2.27 (1.39)	2.30 (1.46)	3.06 (.91)	2.89 (.88)
	t-value	.92	1.05	6.58	5.40
	p	.361	.297	.000	.000
Monotonicity	*M*(*SD*)	2.73 (1.03)	3.15 (.95)	3.47 (.72)	3.61 (.69)
	t-value	3.31	6.29	11.58	12.41
	p	.002	.000	.000	.000
Superordinate category level					
Discrimination	*M*(*SD*)	3.71 (.55)	3.48 (.77)	3.76 (.56)	3.43 (.79)
	t-value	15.22	9.61	18.01	9.57
	P	.000	.000	.000	.000
Diversity	*M*(*SD*)	2.21 (1.02)	1.68 (.90)	2.24 (1.03)	2.42 (1.10)
	t-value	1.00	−1.78	1.35	2.06
	P	.323	.082	.182	.045
Extraction	*M*(*SD*)	1.75 (1.12)	1.44 (.82)	2.00 (.87)	2.29 (1.05)
	t-value	−1.10	−3.41	.00	1.44
	P	.277	.001	1.000	.155
Monotonicity	*M*(*SD*)	2.88 (1.08)	2.60 (1.15)	3.76 (.56)	3.32 (.94)
	t-value	3.98	2.60	18.01	7.40
	P	.000	.012	.000	.000

of age and category level on Diversity were found. Older children had higher diversity scores than younger children, $F(3, 211) = 20.75$, $p < .001$. Also, Diversity scores at the superordinate level were lower than at the basic category level, $F(1, 211) = 7.25$, $p < .01$. There was a significant interaction between age and category level, $F(3, 211) = 5.67$, $p = .001$. As shown in Table 3, the interaction was due to 7-year-olds scoring slightly lower on superordinate- than basic-level problems, whereas 9-year-olds showed the opposite trend, and this trend was stronger in 11-year-olds who scored significantly higher on basic- than superordinate-level problems.

A similar ANOVA was performed for Extraction scores. A main effect of age confirms that older children had higher Extraction scores than young children, $F(3, 211) = 6.30$, $p < .001$. Category level also showed a significant main effect, suggesting that Extraction at the superordinate level was lower than at the basic level, $F(1) = 27.39$, $p < .001$. There was no interaction between age and category level, $F(3, 211) = 0.74$, $p = .527$.

DISCUSSION

The predictive variable in children's diversity-based reasoning

The overall high accuracies of Discrimination and Monotonicity suggested that all children from 6 to 11 years old could understand the tasks and questions. However, only 9- and 11-year-old children answered Extraction and Diversity questions correctly. It was hypothesised that diversity-based reasoning is determined by three variables: ability to differentiate the premise categories (Discrimination), tendency to correlate numerosity of premises (or example) with strength of the conclusion (Monotonicity), and sensitivity to numerosity differences between the premise subclasses (Extraction).

The results indicated that only Extraction showed similar accuracy to, and was correlated with, diversity-based reasoning. Children who assigned greater strength of conclusion to the more numerous of two basic-level categories were more likely to show consistent diversity-based reasoning. Thus, ability to do numerical comparison of subcategories was a valid predictor of diversity-based reasoning.

These results are consistent with our proposal that the law of large numbers underlies diversity reasoning. According to the law of large numbers, the more diverse a premise, the more category members it represents, and correspondingly the stronger the conclusion. By this description, diversity is similar to representativeness.

"In making predictions and judgments under uncertainty, people ... rely on a limited number of heuristics, which sometimes yield reasonable judgments" (Kahneman & Tversky, 1973, p. 237). Representativeness is a heuristic used to estimate the probability of uncertain events by relying on the degree to which a sample or event is typical. Beyond probabilistic reasoning, other phenomena of categorisation, comparison, and inference are influenced by the representativeness bias (Mervis & Rosch, 1981; Osherson et al., 1990). However, a principled account of representativeness has not been easy to come by (Tenenbaum & Griffiths, 2001). Its proponents (Kahneman & Tversky, 1996; Mervis & Rosch, 1981) have asserted that representativeness should be defined only operationally in terms of people's judgements. The current results suggest that, by late childhood, representativeness of diverse premises is at least partly based on the inferred size (number) of the premise subcategories. However, the specificity of these inferences depends on the child's knowledge of the subcategories: for superordinate-level judgements, knowledge of subcategory might be too sparse to estimate subcategory size and thereby influence judgements of coverage (Gutheil & Gelman, 1997; Markman & Wisniewski, 1997). This can explain why we did not see systematic extraction- or diversity-based responses for superordinate categories, even in the older age groups.

Children's use of diversity information for inductive inference

Because most children performed well on tests of Discrimination and Monotonicity, it seems that by 6 years they were approaching a mature use of these conceptual precursors of diversity-based reasoning. This is consistent with previous findings (Gutheil & Gelman, 1997; Lo et al., 2002). It is also noteworthy that 25% of children got fewer than three Monotonicity questions correct, and these children were relatively less likely to get most diversity questions correct: only 25% did so, compared to 47% of children who answered most monotonicity questions correctly. Thus it is possible that children younger than 6 years tend to use monotonicity information less consistently for induction inference.

Only the oldest children, 9- and 11-year-olds, showed above-chance sensitivity to numerical differences between diverse and non-diverse premises at the basic level, and only these children, as a group, performed significantly better than chance in the basic-level diversity test.

Compared with the results of some past studies (Heit & Hahn, 2001; Lo et al., 2001; Lopez et al., 1992; Shipley & Shepperson, 2006), the present study paints a more negative picture of children's diversity-based reasoning, and implicates a limited ability to detect subcategory numerical difference between diverse and non-diverse premises. This was true especially for superordinate categories. It is possible that failure to consider subcategory size at the superordinate level relates to the greater heterogeneity and/or complexity of these categories compared to basic-level categories. Relational complexity has been showed to influence children's inductive reasoning (Li, Zheng, Gao, Gao, & Ling, 2005), and it is possible that complexity interferes with category size judgements.

Lopez et al. (1992) also suggested that kindergartners are insensitive to the number and diversity of premise categories; however, they found that 9-year-olds utilise diversity for inferences about superordinate categories. The present results are therefore inconsistent with Lopez et al.'s results. One possible explanation is that our material differed from Lopez et al.'s. The differences between subcategories in non-diverse premises were somewhat more obvious in the present study. For example:

Non-diverse premise in the present study: cow and horse
Non-diverse premise in Lopez et al: cow and buffalo

In fact, cow and buffalo belong to the same ox family although the buffalo has horns and a shaggy coat. The difference between non-diverse premises in Lopez's study was rather subtle. If this difference had been more pronounced, their 9-year-olds might have detected premise diversity differences and shown more diversity-based reasoning. It has been shown that if the "difference in differences" between diverse and non-diverse premises increases, diversity scores also increase (Chen, Feng, & Gao, 2005).

More generally, different circumstances alter how children perform in diversity reasoning tests. Lopez et al. (1992) and Gutheil and Gelman (1997) reported that 9-year-olds did not show diversity effects for stimulus sets with specific conclusion categories. However, Heit and Hahn (2001) used a different paradigm with family-related categories, and found that 5-year-olds could use diversity cues to make basic-level inductive inferences. Future studies should address the developmental or cultural characteristics of children's diversity-based reasoning by using the same paradigm and material across samples with different cultures, languages, or domain knowledge.

Why don't young children use diversity-based reasoning?

Carey (1985) argued that there are changes with age in knowledge systems of, for example, concepts of animals. These changes include knowledge of taxonomic relations. Children should not be able to use diversity-based reasoning about categories if they do not know the taxonomic relations between categories (e.g., basic- and superordinate-level animal categories). Lopez et al. (1992) suggested another developmental change: younger children might not use all the same reasoning processes as older children and adults.

The present results suggest that younger children's failure to use extraction explains, in some part, why they seldom show diversity-based reasoning. This failure might be due to difficulty in inferring the relative numerosity of the premise subcategories. There is other evidence that children's skill at making relative numerosity judgements about subclasses influences their ability to draw valid deductive class-inclusion inferences (Trabasso et al., 1978). The current results imply that the same quantification skills also play a role in inductive inferences. In short, older children might consider which premises encompass a greater proportion of all members of the superordinate category. This suggests a sort of "folk statistics" similar to the representativeness heuristic. This heuristic might influence children's inferences about the size of subsets within large groups (such as demographic subpopulations within a larger community). It would therefore be interesting to examine how this sensitivity influences the development of children's reasoning about political and economic phenomena.

Summary

Children in the present study did not use extraction as a diversity cue until 9 to 11 years of age. This is consistent with other evidence of the late development of the ability to use category size judgements for deductive and inductive inferences. As children gain sensitivity to diversity, they are also learning to use subclass size—the law of large numbers—for inductive reasoning. The ability to extract the numerosity of subcategories was found to predict children's diversity reasoning. However, it must be noted that while correlational analyses can provide confirmatory support for a certain hypothesis, they cannot support causal hypotheses. Future experimental studies will be needed to determine whether extraction is a necessary condition of children's diversity reasoning.

Manuscript received 4 February 2009
Revised manuscript received 26 August 2009

REFERENCES

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

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language*, *3*, 167–181.

Chen, A., Li H., Feng T., & Gao X. (2005). The Diversity Effect in inductive reasoning under segment manipulation of complex cognition. *Science in China: Series C-Life Sciences*, *48*(6), 658–668.

Fong K., & Nisbett. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*, 253–292.

Gelman, S. A., & O'Reilly, A. W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*, *59*, 876–887.

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 & Review*, *7*, 569–592.

Heit, E., & Hahn U. (2001). Diversity-based reasoning in children. *Cognitive Psychology*, *43*, 243–273.

Jepson, C., Krantz, D. H., & Nisbett, R. E. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences*, *6*, 494–501.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.

Li, H., Zheng, C., Gao, X., Gao, S., & Lin, C. (2005). The influence of complexity and reasoning direction on children's causal reasoning. *Cognitive Development*, *20*, 87–101.

Lo, Y., Sides, A., Rozelle J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science*, *16*, 181–206.

Lopez, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition*, *23*(3), 374–382.

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

Markman, A. B., & Wisniewski, E. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 54–70.

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate level categories. *Child Development*, *53*, 258–266.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 32). Palo Alto, CA: Annual Reviews Inc.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339–363.

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

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

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

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

Shipley, E. F., & Shepperson, B. (2006). Test sample selection by preschool children: Honoring diversity. *Memory and Cognition*, *34*, 1444–1451.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, *25*, 231–280.

Sloman, S. A., & Lagnado, D. (2005). The problem of induction. In R. Morrison & K. Holyoak (Eds.), *Cambridge handbook of thinking & reasoning* (pp. 95–116). New York: Cambridge University Press.

Tenenbaum, J. B., & Griffiths T. L. (2001). The rational basis of representativeness. *23rd Annual Conference of the Cognitive Science Society* (pp. 1036–1041). Available from http://web.mit.edu/cocosci/Papers/cogsci01_final.pdf

Trabasso, T., Isen, A., Dolecki, P., McLanahan, A., Riley, C., & Tucker, T. (1978). How do children solve class inclusion problems? In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

APPENDIX

The sample instructions in Pingyin for Figure 2

[Pingyin] Dong wu yuan li you hen duo dong wu. Liang ge ke xue jia xiang yan jiu zhe xie dong wu de gu tou li han you shen me wu zhi? A ke xue jia yan jiu liao nai niu yu ma zhe liang zhong dong wu, fa xian gu tou li you ting,yu shi A shuo suo you de dong wu gu tou li han you ting.B ke xue jia yan jiu liao nai niu yu song shu fa xian gu tou li you ting, yu shi B shuo suo you de dong wu gu tou li han you ting. Qing wen ni geng xiang xin na yi ge ke xue jia shuo de hua?

The sample instructions in Pingyin for Figure 3

[Pingyin] Qing kan zhe san zhong dong wu, ni ren wei na zhong dong wu yu nai niu xiang si? Shi ma geng xiang nai niu hai shi song shu geng xiang nai niu?

The sample instructions in Pingyin for Figure 4

[Pingyin] Er tong jie dao liao, lao shi yao dai ban shang de 7 ge xiao peng you chu qu wan.Ta wen xiao peng you qu na li wan. You 5 ge xiao peng you xiang qu kan dian ying, ling wai liang ge xiao peng you xiang qu dong wu yuan. Lao shi shuo, suo you de xiao peng you dou zhi neng qu yi ge di fang. Ni men zai shang liang yi xia dao di qu na li. Qing wen:xiao peng you shang liang hou hui qu na li? Shi kan dian ying hai shi qu dong wu yuan?

The sample instructions in Pingyin for Figure 5

[Pingyin] Dong wu wang guo yao kai da hui. Guo wang qing mei ge dong wu yuan pai liang ge dong wu can jia da hui 。A dong wu yuan pai qu de shi yi tou lao hu he yi tou bao zi 。B dong wu yuan pai qu de shi yi tou lao hu he yi zhi tu zi 。Qing ni xiang yi xiang na ge dong wu yuan de dong wu ke neng yao duo yi xie ? Shi A dong wu yuan hai shi B dong wu yuan?