Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
J. Experimental Child Psychology 86 (2003) 253–276
Journal of Experimental Child Psychology
www.elsevier.com/locate/jecp

The role of diverse instruction in conceptual change

Brett K. Hayes,[a,*] Alison Goodhew,[b] Evan Heit,[c] and Joanna Gillan[a]

[a] *School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia*
[b] *University of Newcastle, Newcastle, NSW 2308, Australia*
[c] *University of Warwick, Coventry CV4 7AL, UK*

Received 3 June 2003; revised 10 September 2003

Abstract

This study examined how a fundamental principle of induction and scientific reasoning, information diversity, could be used to promote change in children's mental models of the earth's shape. Six-year-old children ($N = 132$) were randomly allocated to a control or to one of two training conditions. Some training groups received instruction that simultaneously challenged children's beliefs concerning (a) why the earth appears flat to a surface observer and (b) the role of gravity. Others received instruction that repeatedly challenged only one of these beliefs. An adaptation of the Vosniadou and Brewer (1992, *Cognitive Psychology 24*, 535–585) protocol for identifying mental models of the earth was administered before and after instruction. Both instruction methods produced increases in factual knowledge. Only children receiving instruction about two core beliefs, however, showed an increased rate of acceptance of a spherical earth model at posttest. The findings show that instruction that challenges diverse aspects of children's naïve scientific beliefs is more likely to produce conceptual change.

Keywords: Conceptual change; Framework theories; Concepts

* Corresponding author. Fax: +61-2-9385-3641.
E-mail address: B.Hayes@unsw.edu.au (B.K. Hayes).

0022-0965/$ - see front matter
doi:10.1016/j.jecp.2003.09.002

Introduction

A substantial body of evidence indicates that children often construct conceptual frameworks or intuitive theories in an attempt to explain the natural phenomena that they observe (see Wellman & Gelman, 1998, for a review). Although there is still debate about how such conceptual systems are organized (cf. Carey, 1985; di Sessa, 1993; Hammer, 1996; Vosniadou & Brewer, 1992), most models of children's naïve scientific beliefs assume that they constrain the acquisition of new concepts and are resistant to revision. Mere exposure to concrete experiences that contradict intuitive theories is unlikely to bring about significant conceptual change (Chinn & Brewer, 1993; Kuhn, 1989).

These attributes are well illustrated in studies of children's understanding of an aspect of elementary astronomy, the shape of the earth. Nussbaum and Novak (1976) showed that although second-grade children claim that the earth is round, more detailed questioning elicited responses that were consistent with a belief in a flat earth or a variety of alternative nonspherical models. Vosniadou and Brewer (1992) developed a more explicit system for identifying the mental models that children use to answer a variety of factual and inferential questions about the earth's shape. When this system was applied to the interview responses of school-age children, it was found that only a small minority of 6-year-olds (15%) used a spherical model, with the proportion increasing gradually with age to 40% of 9-year-olds and 60% of 11-year-olds. This pattern of developmental change has been widely replicated with children in Israel (Nussbaum, 1979), Nepal (Mali & Howe, 1979), India (Samarapungavan, Vosniadou, & Brewer, 1996), and Greece and Samoa (Vosniadou, 1994) and with Native American children (Diakidoy, Vosniadou, & Hawks, 1997).

Such models have been shown to be highly resistant to change by conventional instruction methods, particularly in young children (e.g., Chinn & Brewer, 1993; Kikas, 1998; but see Diakidoy & Kendeou, 2001, and Sneider & Ohadi, 1998, for more positive training outcomes with older children). Nussbaum and Novak (1976), for example, found that presenting pictures of the earth as seen from space and encouraging children to interact with globe models of the earth had little impact on 7-year-olds' beliefs about the earth's shape. When a revised instructional program was administered to Israeli 7-year-olds, some improvement in children's understanding of the spherical nature of the earth was noted (Nussbaum & Sharoni-Dagan, 1983), but the implications of this finding are unclear because it was not based on an explicit system for identifying children's mental models from interview responses and was not supported by any statistical analysis of the magnitude of model change. In a related domain, Vosniadou (1991) found that following presentation of text explaining the day–night cycle, third-grade children were able to report additional facts about the astronomical motions of the sun and earth but did not alter their preexisting explanations of the cycle.

Carey (2000) has suggested two reasons that most previous attempts to modify young children's scientific concepts have failed. The first is that many researchers have failed to take account of the structure of children's naïve scientific beliefs. Vosniadou and Brewer (1992) have shown that children's erroneous beliefs about the shape of the earth are based upon two more general misconceptions. First, young children have difficulty resolving the apparent contradiction between the spherical shape of the earth and its appearance as flat to an observer on the ground. Second, children fail to appreciate the influence of gravity on objects on different parts of the earth's surface. The degree to which individual children endorse each of these beliefs constrains the type of model that they construct when making inferences about the earth's shape. Hence, a belief that the ground extends along a flat plane will prevent children from constructing a spherical model. When the belief that unsupported things fall is naïvely applied to the earth's surface, children are likely to construct a disk model of the earth or endorse a "dual earth" model with a round earth located in space coexisting with a flat earth where people live.

According to Vosniadou and Brewer (1992) normal developmental progression toward a spherical earth model involves the gradual revision of these two underlying beliefs. It follows that procedures that aim to promote conceptual change in mental models of the earth should target these beliefs. Previous failures in promoting model change may be due, in part, to a failure to provide information that challenges these beliefs. One of the aims of the present study therefore was to promote change in children's earth concepts through the presentation of information that directly challenged the misconceptions that underlie a nonspherical conception of the earth's shape.

The diversity principle in scientific reasoning

A second problem with many previous attempts to modify children's scientific beliefs is that they have not been guided by a coherent model of how scientific beliefs are modified in the light of new data (Carey, 2000). Although there is still considerable debate about the role of new data in the construction and modification of scientific theories (cf. Lakatos, 1970; Kuhn, 1970), there is broad agreement on some of the characteristics that define salient or persuasive scientific evidence. One principle that has long been influential in the philosophy of science is that the superior way to test a scientific theory is by conducting diverse experiments that assess different aspects of the theory (e.g., Bacon, 1620/1898; Hempel, 1966; Nagel, 1939; but see Wayne, 1995; Lo, Sides, Rozelle, & Osherson, 2002, for critiques). According to the diversity principle, hypotheses are considered to have greater confirmation when supported by diverse rather than by similar data sets and when confirming data are obtained using diverse methods. When a hypothesis is being tested, additional observations are sought from a different source or by using a different methodology that can provide converging evidence for the hypothesis.

A considerable body of evidence suggests that adults and children without formal scientific training make use of the diversity principle in their intuitive reasoning about observed phenomena. Osherson, Smith, Wilkie, López, and Shafir (1990), for example, asked adults to rate the relative strength of inductive arguments with multiple premises that varied in their diversity. Consider, for example, the following pairs of arguments: (1) "Lions have an ulnar artery and giraffes have an ulnar artery. Therefore rabbits have an ulnar artery." (2) "Lions have an ulnar artery and tigers have an ulnar artery. Therefore rabbits have an ulnar artery." Osherson et al. showed that adults find arguments like (1), which contain a more diverse set of premises, to be more persuasive than (2). The same pattern is found with arguments that have more general conclusion categories (e.g., "Lions have an ulnar artery and giraffes have an ulnar artery. Therefore all mammals have an ulnar artery"). Hence, the use of more diverse premises seems to promote a broader range of inductive inferences.

Lopez (1995) reports an even clearer parallel between this intuitive use of the diversity principle and formal scientific reasoning. Adults were presented with a novel property of a single premise category (e.g., "lions") and asked to seek further evidence to test whether this property generalized to a conclusion category (e.g., "all mammals"). The adults consistently sought diverse evidence about animals that were dissimilar to the premise animals (e.g. "rabbits") rather than about animals that were seen as similar to the premise (e.g., "tigers") (also see Spellman, Lopez, & Smith, 1999).

Early investigations of young children's inductive reasoning (e.g., Carey, 1985; Gutheil & Gelman, 1997; Lopez, Gelman, Gutheil, & Smith, 1992) reported that children younger than 9 years of age had some difficulty in applying the diversity principle in inductive reasoning tasks (see also Heit, 2000, for a review of other exceptions to the diversity principle). There are reasons to believe, however, that these negative developmental findings may reflect age differences in knowledge of taxonomic relations and performance components that limit children's ability to express a preference for diverse information. One problem with these studies is that tasks involving reasoning about diverse premises were presented within the same testing session with tasks that involved different inductive principles (e.g., choosing between arguments with single premises that varied in similarity to the conclusion). Hence, for children to display sensitivity to the diversity principle, they would have to shift their inductive strategies across different items.

Heit and Hahn (2001) tried to minimize the impact of such performance factors by presenting children between 5 and 9 years of age with inductive arguments that differed only in the diversity of their premises. The objects used in their induction task were pictures of everyday objects likely to be familiar to young children (e.g., footballs, dolls, flowers). Under these conditions Heit and Hahn found robust sensitivity to premise diversity in children as young as 5 years. Subsequent studies (Heit & Hahn, 2002; Lo et al., 2002) have shown that young children are capable of using the diversity principle as a basis for inductive inferences about both observable properties (e.g., "has a doll") and nonobservable properties (e.g., "have T cells").

The diversity principle not only is relevant to reasoning involving category hierarchies but also has implications for causal reasoning. Kim and Keil (2003) have shown that adults' judgments about the causes of observed medical symptoms are influenced by the diversity of these symptoms within a causal hierarchy. Symptoms that arise from different causal mechanisms (e.g., burning of skin, destruction of white blood cells), which in turn share a common cause (e.g., exposure to radiation), are seen as more diagnostic than symptoms that arise from a single mechanism. Hence, providing information about multiple branches of a causal hierarchy leads to stronger conclusions than providing additional information about a single branch.

The diversity principle and conceptual change

This work shows that, like scientists, lay adults and children draw stronger inductive inferences from information that impacts on diverse aspects of their underlying beliefs. These findings have considerable implications for the understanding of conceptual change. They suggest that shifts in naïve models of natural phenomena are more likely to occur when people encounter new information that challenges several features or assumptions of these models. Such diverse instruction is likely to produce more marked change in mental models than the accumulation of new data that repeatedly challenges a single assumption of the model.

A major aim of this study therefore was to examine the role of diverse instruction in conceptual change in the context of children's understanding of the earth's shape. The diversity principle suggests that, in this domain, conceptual change will be more likely if we use a few examples that challenge multiple assumptions (e.g., flatness of the earth, the role of gravity) than if a larger number of examples challenging just one of these assumptions are presented. Presenting the same or similar kinds of evidence repeatedly should lead to progressively smaller changes in beliefs, as the evidence becomes less surprising or informative each time it is presented. In contrast, the presentation of a more varied or diverse set of evidence, relative to two different core beliefs, should lead to greater changes in beliefs overall.

This prediction about the impact of diverse instruction on conceptions of the earth's shape was examined here with 6-year-old children. According to previous surveys of children's earth concepts (e.g., Nussbaum & Novak, 1976; Vosniadou & Brewer, 1992), the majority of children of this age would not be expected to have achieved a spherical model of the earth. This was confirmed by administering an adaptation of the Vosniadou and Brewer (1992) interview protocol and assessing children's earth concepts both in terms of response to specific factual and inferential questions and in terms of children's current mental models. Children were then provided with training that challenged beliefs about the relative size of the earth or the role of gravity (single-belief condition) or an equivalent number of training instances that challenged both of these beliefs (dual-belief condition). At the end of training, children's understanding of the shape of the earth was reassessed. Both groups receiving instruction were expected to show greater change than controls who received no instruction. However, if the diverse instruction does promote conceptual change, then more children should show a shift from a nonspherical to a spherical model in the dual-belief condition than in the single-belief condition.

Method

Participants

The participants were 132 first-grade children (60 boys, 72 girls) from one public and three private schools in a middle-income metropolitan area in Southeast Australia. Children were randomly assigned to the three experimental conditions, with the constraint that one of the schools involved in testing inadvertently recruited more students in the dual-belief condition. There were 40 children in each of the control and single-belief conditions, and 52 in the dual-belief condition. Children's ages ranged from 6 years 2 months to 7 years 5 months ($M = 6$ years, 7 months). A review of the grade curriculum and an interview with class teachers confirmed that no formal instruction about the earth's shape had been provided prior to participation.

Materials

A structured interview to assess children's current mental models of the earth was adapted from the protocol used by Vosniadou and Brewer (1992). The interview schedule was identical to that used by Vosniadou and Brewer (1992) except that one item from the Vosniadou and Brewer version ("Now I want you to show me where Champaign is. Where is China?") was omitted. Parallel forms of Items 8 and 9 were created for use in pre- and posttraining assessments (see Appendix A). The interview consisted of 14 items that required verbal (e.g., "What is the shape of the earth?") or drawing responses (e.g., "Can you draw a picture of the earth?"). A combination of factual (e.g., "Show me where the moon and stars go") and inferential questions (e.g., "If you walked for many days in a straight line, where would you end up?") was used. Where necessary, protocol items were followed by prompting questions, such as "Can you tell me a little more about that?" Computer-generated pictures (of a house in a flat landscape or of a children's playground in a flat landscape) were used in conjunction with Item 9 in both interview forms.

Three instructional videos were produced for use in the respective training conditions using Avid Cinema v1.1.2 on a Macintosh Power PC 6500. In the videos scientifically accepted information about the earth was presented in four episodes. In these episodes information about the earth was presented via spoken narration accompanied by animated cartoons, live action video clips, and still photographs (see Appendix B for a video summary). The episodes included demonstrations of factual information (e.g., views of the earth from space), analogies (e.g., gravity acts like a magnet), and causal explanations (e.g., the earth appears flat because we are observing only a small part of the surface). For children in the single-belief condition all four video episodes focused on either the relative size of the earth or the effects of gravity. For those in the dual-belief condition the video contained two episodes focusing on the size of the earth and two focusing on gravity. The order of presentation of these two sets of episodes was counterbalanced across participants in this group. The selection of specific episodes was also counterbalanced within this group so that each of the four possible video episodes illustrating size or gravity issues was viewed an equal number of times. The duration of all instructional videos was approximately 6 min.

Procedure

All participants were first administered an adaptation of the Vosniadou and Brewer (1992) interview schedule before training commenced. Approximately 2 h after the completion of the pretest interviews, training with the instructional videos commenced as outlined below. Seven days later the experimenter returned and administered the alternate form of the schedule. The assignment of each interview form to pre- and posttesting was counterbalanced across participants. All interviews were carried out individually by the second and fourth authors and lasted approximately 15 min.

Children assigned to the two training conditions viewed the relevant instructional video in groups of 3 or 4. The only difference between the procedures for the single- and the dual-belief conditions was the type of instructional video presented (four episodes containing information on the size of the earth *or* gravity or 2×2 episodes containing information of the size of the earth *and* gravity, respectively). In the single-belief condition, 20 children were randomly assigned a video that contained information about the size of the earth, and 20 were assigned a video containing information about gravity. The control group was not presented a video and had no further contact with the experimenter until the posttest interview.

Scoring of interview protocols

All interviews were scored in two ways, first at the item level and then at the model level. Interviews were scored according to the template devised by Vosniadou and Brewer (1992), which gives a comprehensive list of possible responses to each interview question. At the item level, responses to each of the 14 interview items were scored as "correct" (1) or "incorrect" (0) on the basis of a consensual scientific understanding of the shape of the earth, and correct responses were tallied across items.

However, the number of correctly answered items does not necessarily reveal whether a child has formed a coherent model of the shape of the earth, nor does it reveal what the model is. Therefore it was crucial to also assess responses at the model level. The pattern of verbal and drawing responses across items was used to identify each child's current mental model of the shape of the earth. A model-level scoring key was adapted from that developed by Vosniadou and Brewer (1992). The scoring key described the pattern of expected responses to interview items for each mental model. Two changes were made to Vosniadou and Brewer's (1992) scoring key. In both cases the changes were motivated by our observation that some children showed a response pattern that suggested a reasonable grasp of the spherical shape of the earth, but because they failed or misinterpreted one item, would have been unfairly categorized as having a more primitive model. First, in our scoring system it was possible for a child to answer Item 11 ("Is there an end or an edge to the earth?") in the affirmative and still be classified as having a flattened sphere model. Second, if a child answered "ground" or "water" to Item 12 ("What is below the earth?"), they could still be classified as having a flattened sphere or sphere model if their responses to all other items were consistent with the expected patterns for these models. These scoring changes meant that 4 children at pretest and 5 at posttest who would have been classified as having a disk earth model using the Vosniadou and Brewer criteria were instead assigned to the flattened sphere category. In addition, 7 children at pretest and 12 at posttest who would have been classified as having a disk earth or mixed model by Vosniadou and Brewer's criteria were judged to have a sphere model.

Results

Reliability of scoring

The interrater reliability of item-level and model-level scoring was examined by randomly selecting a sample ($n = 56$) of pre- and posttraining assessments carried out by the second and fourth authors and having these scored by an independent rater. This rater was given a written copy of the scoring rules and had the opportunity to question the experimenters about rule interpretation, but was blind to the group membership of each child being assessed. For item-level scoring, exact agreement on the total number of correct answers was obtained for both pre- and posttraining assessments in 55 of a possible 56 cases. In one case there was a one-point difference between the scores awarded by the raters on a pretest.

High interrater reliability was also found in assignment of interview protocols to different categories of mental models, with the raters agreeing on the pre- and post-test assignment of 53 of 56 assessments, Cohen's $\kappa = .94$, $z = 18.17$, $p < .001$. These disagreements were resolved by discussion between the raters.

Item-level performance

Because unequal proportions of children from different schools were allocated to the three experimental conditions, we carried out a preliminary analysis to compare item-level performance across schools. A one-way ANOVA showed that the total number of correct responses to interview questions across pre- and posttests did not differ as a function of school, $F(3, 129) = 0.88$, $p = .45$. In all subsequent analyses, therefore, data were collapsed across participating schools.

The main research question was concerned with differences in the effects on children's mental models of training involving one or two underlying beliefs. It was important, therefore, to establish first whether the two types of single belief training (i.e., size or gravity) had similar impacts on children's performance. To this end a preliminary analysis was carried out to check whether the type of belief targeted for change affected accuracy in the single-belief training condition. A one-way ANOVA showed that the total number of correct responses to interview questions across pre- and posttests for participants trained only with examples illustrating the relative size of the earth ($M = 6.4$, $SD = 1.7$) did not differ significantly from the number of correct responses given by those trained only with examples illustrating the effects of gravity ($M = 6.15$, $SD = 1.8$), $F(1, 38) = .12$, $p = .73$. In subsequent analyses of overall group performance, therefore, the results for the single-belief condition were collapsed across participants exposed to the "size" and "gravity" videos. Within the dual-belief condition we also examined whether there were differences in performance between the various combinations of size and gravity training episodes.

A one-way analysis of variance found no differences between these conditions in the number of correct responses across pre- and posttests.

The total number of correct responses, out of a maximum score of 14, at each test occasion is given in Table 1. The data were entered into a 3 (training group) $\times$ (2) (test occasion) mixed ANOVA with repeated measures on the second factor. Two contrasts were planned. The first contrast compared the performance of control children with that of children in the single-belief and dual-belief conditions. The second compared performance in the two training conditions. The respective weights assigned to the control, single-belief, and dual belief conditions were +2, −1, −1 for the first contrast and 0, 1, −1 for the second contrast. The ANOVA revealed that across groups, there was a significant increase in accuracy from pretest ($M = 6.51$) to posttests ($M = 7.42$), $F(1, 129) = 45.26$, $MSE = 52.75$, $p < .001$. This effect, however, was qualified by an interaction with the contrast comparing the performance of the training and control conditions, $F(1, 129) = 7.78$, $MSE = 9.09$, $p < .01$. The increase in children's correct responding from pre- to posttesting was significantly greater for children given some form of training ($M_{\text{Posttest}-\text{Pretest}} = 1.19$) than for controls ($M_{\text{Posttest}-\text{Pretest}} = 0.35$). Although the increase in accuracy from pre- to posttests was somewhat greater for the dual-belief than for the single-belief condition, this contrast did not reach significance.

It is also important to discover whether the size and gravity components of the training conditions had different kinds of impact on children's responses at the item level, and whether these effects were additive in the dual-belief condition. To this end we computed pre–post difference scores for all 14 items in the control and dual-belief conditions and in the size and gravity versions of the single-belief condition (see Table 2). Tukey's HSD tests were then used to examine whether the change in performance for each item differed in the various training conditions compared to the control. As can be seen from Table 2, in the single-belief condition "size" and "gravity" training impacted somewhat different aspects of the understanding of the earth's shape. Training with the size version of single belief led to significantly larger improvements, relative to baseline, on Item 9 ("How come the earth [in the picture] is flat, but before you made it round?"), Item 10 ("If you walked for many days in a straight line, where would you end up?"), and Item 14 ("What is below the earth?"). Training with the gravity version of single belief led to larger improvements on Item 9, Item 14, and Item 8 ("Show me where the people live [on your drawing]"). In the dual-belief condition performance also improved on Items 9 and 14. Notably, however, improvements were also found on some items that were unaffected by single-belief training, including Items 11 ("Would you ever reach the end/edge of the earth?"), 12 ("Can you fall off that edge?"), and 13 ("Where would you fall?").

Table 1
Means (and standard deviations) for item-level accuracy at each test occasion

Training condition	Pretest	Posttest
Control	6.88 (2.2)	7.23 (2.21)
Single belief	6.00 (1.87)	7.12 (1.73)
Dual belief	6.66 (2.09)	7.91 (1.87)

Table 2
Means (and standard deviations) for pre–post difference scores for each interview item

	No training control		Single belief (size)		Single belief (gravity)		Dual belief	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Item 1	1.08	(0.80)	1.15	(0.53)	1.56	(0.39)	1.46	(0.48)
Item 2	0.54	(0.98)	1.15	(0.39)	1.00	(0.57)	1.04	(0.66)
Item 3	0.13	(0.80)	1.15	(0.39)	1.15	(0.39)	1.04	(0.39)
Item 4	0.51	(0.72)	0.58	(0.53)	0.75	(0.69)	0.84	(0.62)
Item 5	1.08	(0.62)	1.15	(0.69)	1.26	(0.81)	1.31	(0.72)
Item 6	0.38	(0.46)	0.58	(0.39)	0.86	(0.39)	0.38	(0.23)
Item 7	0.51	(0.80)	1.15	(0.39)	0.29	(0.57)	0.67	(0.47)
Item 8	−0.51	(0.46)	1.15	(0.39)	1.56*	(0.39)	1.04	(0.39)
Item 9	−0.13	(0.46)	1.77*	(0.69)	1.15*	(0.69)	1.79*	(0.57)
Item 10	0.76	(0.46)	2.19*	(0.53)	1.05	(1.07)	1.04	(0.74)
Item 11	0.00	(0.22)	0.42	(0.53)	0.32	(0.48)	1.79*	(0.50)
Item 12	−0.13	(0.80)	0.00	(0.39)	1.15	(0.77)	1.35*	(0.38)
Item 13	0.76	(0.62)	0.42	(0.53)	1.15	(0.39)	1.32*	(0.23)
Item 14	−0.13	(0.62)	2.61*	(0.53)	2.72*	(0.60)	2.44*	(0.56)

* Significantly different from corresponding item in No Training Control by Tukey's HSD test, $p < .05$.

Stability and change in mental models

We now turn to the crucial analyses, those at the model level. Table 3 shows the frequency of pre- and posttraining mental model classifications for all participants. An inspection of Table 3 reveals some important trends. When a change in children's models did occur across testing occasions, it was most often a positive change from a nonsphere to a sphere model (15% of all participants) or a change from one nonsphere model to another nonsphere model (13% of all participants). Negative changes from a sphere to a nonsphere model were rare (2% of all participants).

Table 3
Pre- and posttest frequencies (and within-group percentages) of each mental model in each training condition

Model	Control f (% within group)		Single belief f (% within group)		Dual belief f (% within group)	
	Pre	Post	Pre	Post	Pre	Post
Sphere	10 (0.25)	10 (0.25)	7 (0.18)	11 (0.28)	12 (0.23)	25 (0.48)
Flattened sphere	4 (0.10)	5 (0.13)	4 (0.10)	6 (0.15)	4 (0.08)	3 (0.03)
Hollow sphere	5 (0.13)	7 (0.18)	9 (0.23)	8 (0.20)	10 (0.19)	7 (0.13)
Disk earth	4 (0.10)	4 (0.10)	3 (0.08)	1 (0.03)	3 (0.06)	1 (0.02)
Dual earth	11 (0.28)	10 (0.25)	15 (0.38)	13 (0.33)	17 (0.33)	14 (0.27)
Rectangular	2 (0.05)	1 (0.03)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Mixed	4 (0.1)	3 (0.08)	2 (0.05)	1 (0.03)	6 (0.12)	2 (0.04)

To compare patterns of stability and change in children's mental models in the three experimental conditions, the model category data for each testing occasion were collapsed into the two categories of "spherical" or "nonspherical." Each participant was then classified in terms of the models held at the pre- and posttests, yielding the joint frequency distribution in Table 4. The key result is that in the dual-belief condition, 40 children initially had nonspherical models, and 14 (35%) of them changed to spherical models after training. In comparison, in the control condition, only 2 of 30 children (6.7%) changed from nonspherical to spherical models, and in the single-belief condition, only 4 of 33 children (12.1%) changed from nonspherical to spherical models. Preliminary analyses showed that the proportion of children who initially had a nonspherical model and then changed to a spherical model was significantly higher in the dual-belief condition than in either the control condition, $\chi^2(1) = 7.81$, $p < .01$, or the single-belief condition, $\chi^2(1) = 5.39$, $p < .02$.

To examine patterns of model change across the whole data set, the joint frequency distribution in Table 4 was entered into a repeated-measures log-linear analysis (Von Eye & Niedermeier, 1999). This analysis examined whether the distribution of children across the sphere and nonsphere categories varied systematically as a result of the training group and test occasion factors. It also allows the testing of specific comparisons between training groups, in a manner analogous to contrast testing in analysis of variance. The first stage of this analysis involved the iterative fitting of models to the observed frequency distribution to find a model that accurately described these data. The most parsimonious model to give a good fit to the data was one containing the test occasion factor and group contrasts comparing the control group with the single-belief and dual-belief conditions, respectively, $\chi^2(4) = 3.76$, $p = .44$. In the second stage of the log-linear analysis we examined whether individual model parameters (denoted by λ) made a significant contribution to the explanation of change in the proportion of children showing sphere and nonsphere models. Three model parameters were found to be significant. The first two

Table 4
Joint frequency distribution of pre- and posttest sphere and nonsphere mental models

Group	Pretest model	Posttest model	Frequency
Control	Nonsphere	Nonsphere	28
	Nonsphere	Sphere	2
	Sphere	Nonsphere	2
	Sphere	Sphere	8
Single belief	Nonsphere	Nonsphere	29
	Nonsphere	Sphere	4
	Sphere	Nonsphere	0
	Sphere	Sphere	7
Dual belief	Nonsphere	Nonsphere	26
	Nonsphere	Sphere	14
	Sphere	Nonsphere	1
	Sphere	Sphere	11

parameters were not directly connected with the experimental hypotheses and reflected global trends in the frequency data. Across groups, it was found that a greater proportion of children had nonsphere models (78.8% of participants) than sphere models at pretest (21.2% of participants), $\lambda = -1.39$, $z = 4.03$, $p < .01$. In addition, more children showed model stability (82.58% of participants) than model change (17.42%) across testing occasions, $\lambda = -0.97$, $z = -5.62$, $p < .001$. The third significant parameter, however, was crucial to the evaluation of our hypotheses. The proportion of children showing change from a nonsphere to a sphere model was greater in the dual-belief group (26.9% of that group) than in the control group (5% of that group), $\lambda = 1.05$, $z = 2.88$, $p < .01$. In contrast, the parameter comparing change from nonsphere to sphere models in the single-belief (10% that group) and control conditions did not reach significance, $\lambda = .15$, $z = .34$, $p = .63$. Hence, children trained in the dual-belief condition, but not those in the single-belief condition, showed a significantly greater shift from nonsphere to sphere models than children in the control group.

In a second log-linear analysis we used a broader definition of "positive" conceptual change that included not only shifts from nonsphere to spherical models, but also change from less to more sophisticated nonsphere models. Vosniadou and Brewer (1992) draw a distinction between two types of nonspherical earth models that differ in their level of scientific sophistication. "Initial models" of the earth (rectangular earth, disk earth) are based entirely on naïve observations of earth's apparent flatness. "Synthetic models" (dual earth, hollow earth, flattened sphere) represent an attempt by children to reconcile their naïve observations with information they receive from adult culture that the earth is a sphere. A shift from an initial to a synthetic model could also be seen as evidence of positive conceptual change. In this analysis, therefore, children's mental models before and after training were classified as initial, synthetic, or spherical. Positive change was defined as a shift from any nonsphere model before training to a sphere model after training or a shift from an initial to a synthetic model. These categorical data were entered into a log-linear analysis containing parameters representing the occasions of measurement and the same group contrasts used in the previous log-linear analysis. This model produced a good overall fit to the data, $\chi^2(16) = 14.45$, $p = .56$. In terms of the specific parameters generated by the model that were relevant to the hypotheses, the crucial finding was that the parameter representing the comparison of positive change in the control (7.5% of children in that group) and dual-belief conditions (30.77% of children in that group) was significant, $\lambda = 0.51$, $z = 3.09$, $p < .01$, but the parameter comparing positive change in the control and single-belief conditions (12.5% of children in that group) was not, $\lambda = -0.18$, $z = -0.86$, $p = .2$. Again, the level of positive change relative to controls was found to be greater in the dual-belief condition than in the single-belief condition.

Relation between item-level and model-level performance

To examine the relation between performance on individual items and children's mental models we calculated the difference in item-level performance across pre- and posttests for each participant. These difference scores were then correlated whether or not children showed a positive change in their mental models (as defined above) following training. A modest but significant positive correlation was found, $r(130) = 0.31$, $p < .001$. We also examined correlations between pre–post difference scores and positive model change for each item in the questionnaire. Across groups, improvements on four items were found to be positively correlated with the probability of shifting to a spherical model: Item 2 ("Which way we do we look to see the earth?"), $r(130) = 0.27$; Item 9 ("How come the earth [in the picture] is flat, but before you made it round?"), $r(130) = 0.3$; Item 11/11a ("Would you ever reach the end/edge of the earth?"), $r(130) = 0.44$; and Item 12 ("Can you fall off the edge?"), $r(130) = 0.32$, all p's $< .001$.

Discussion

Our aim was to evaluate the effectiveness of a novel approach to promoting conceptual change in young children. The training intervention targeted the two core misconceptions thought to underlie erroneous mental models of the earth's shape (Vosniadou & Brewer, 1992). Following models of inductive reasoning that emphasize the importance of information diversity (e.g., Heit, 2000; Osherson et al., 1990), we predicted that challenging two core beliefs would produce a greater shift toward the consensual spherical earth model than exposure to an equivalent number of training episodes that challenged just a single core belief.

Children's understanding of the earth's shape was assessed before and after training in two ways. Item-level assessment involved scoring factual and inferential responses to each interview question as correct or incorrect. Children in the single- and dual-belief training conditions showed a greater increase in correct responses from pre- to posttests than children in the control group.

Assessing performance only in terms of the number of correct items, however, does not always give a clear indication of the mental models that underlie a child's understanding of a domain (e.g., Solomon & Cassimatis, 1999; Vosniadou, 1991; Vosniadou & Brewer, 1992). This was true in the current study where there was only a modest relation between increases in item-level performance and positive shifts in children's mental models.

Overall our brief intervention had only modest success in changing children's mental models. Over two-thirds of children in the training conditions maintained nonspherical models at the posttest. Nevertheless the pattern of change was consistent with our predictions, with a significant shift toward use of a spherical model found in the dual-belief condition relative to controls, but not in the single-belief condition. Dual-belief training was also advantaged relative to the other groups when we examined change from less complex to more sophisticated nonsphere models.

These results show that the diversity principle not only is important in young children's judgments about property generalisation (e.g., Heit & Hahn, 2001) but also can affect more complex belief structures. The consideration of evidence that challenges diverse aspects of a belief system is seen as a characteristic of both normative systems of induction (e.g., Horwich, 1982) and as sound scientific practice (e.g., Hempel, 1966). In the current study children's responses showed some parallels to those of formally trained scientists in that instruction about diverse aspects of the earth's shape was linked to greater levels of conceptual change. These findings could be seen as additional support for the "little scientist" hypothesis, which holds that there are strong parallels between the reasoning processes of children and scientists (e.g., Brewer, Chinn, & Samarapungavan, 2000; Carey, 1985; Gopnik & Meltzoff, 1994; Keil, 1989). In particular, our results suggest that one of the key principles thought to guide theory development and change in science also influences conceptual change in children.

We also need to be clear about the limitations of this "child as scientist" analogy. Our findings support the notion that, at an individual level, children and scientists may use similar kinds of inductive principles to evaluate their beliefs about natural phenomena. These data, however, do not address the more contentious issues of whether children form "theories" that resemble the theories of science, or whether conceptual change in children is analogous to shifts in theoretical stance that have been found in the history of the scientific disciplines. Our minimal assumptions about the internal structure of children's mental models of the earth were that (a) these models were constrained by two sets of beliefs related to children's everyday observations concerning the apparent flatness of the earth's surface and the function of gravity and (b) these constraints would produce internal consistency in the responses of individual children to the various interview items. Like Vosniadou and Brewer (1992) we see "mental models" as being generated from these underlying beliefs when children are required to answer questions such as those in our interview or to solve specific problems. By presenting diverse information that challenged core beliefs, we led some children toward a new pattern of responding, indicative of a new synthetic mental model. This interpretation does not require the strong assumption that the internal structure of children's beliefs before or after training resembles that of a formal scientific theory of the earth's shape, nor do we assume that the process of change involves replacing one "theory" with another. Indeed, we note that in other domains, children's beliefs about natural phenomena often lack critical features of scientific theories such as the use of formalisms and the generation of novel predictions (Brewer et al., 2000; Harris, 1996; Kuhn, 1989).

One way of viewing our results might be as a relatively straightforward demonstration that in attempting to change naïve beliefs that are based on two underlying misconceptions, challenging both assumptions is more likely to succeed than challenging just one. This view implies that conceptual change may simply be a cumulative function of the number of misconceptions that are challenged by novel evidence. Our data, however, suggest a different conclusion. Simultaneously challenging children's assumptions about both the earth's size and gravity led to outcomes that were not equivalent to a simple addition of the effects of challenging assumptions about each of the two components in isolation. Following dual-belief training children performed better on inferences relating to what happens when you approach the edge of the earth (Items 11–13). No such improvement was noted in either the "size" or the "gravity" version of single-belief training. This result was found over and above improvements in item-level performance that were found following training in both single- and dual-belief conditions (i.e., Items 9 and 14).

A further indication that it is not simply the number of assumptions that are challenged during training but also the way that this evidence is presented that is important for conceptual change comes from a comparison of our results with those of previous studies. Others have challenged young children's misconceptions about both relative size and gravity (see Nussbaum & Novak, 1976; Nussbaum & Sharoni-Dagan, 1983; Vosniadou, 1991). A possible explanation for the failure of these approaches in changing children's beliefs is that they challenged misconceptions in a serial fashion rather than presenting multiple examples that challenged a number of different misconceptions within a single training session. Nussbaum and Sharoni-Dagan (1983), for example, presented several instructional sessions containing demonstrations of the size of the earth and its effects on the perspective of a surface observer some weeks before sessions dealing with the effects of gravity.

Another factor that may have contributed to the success of dual-belief training is that it capitalizes on the interdependence between core beliefs (cf. Kim & Keil, 2003). It is likely that acquiring an understanding of the size of the earth could support children's understanding of gravity, and vice versa. As children begin to accept that individual perspective of the earth's shape is constrained by the fact that they are observing only a small part of a very large sphere, their attention may be directed to the problem of how other observers located on other parts of the sphere remain attached to the surface. Providing demonstrations and explanations that are relevant to each of these issues may therefore have a synergistic effect on children's understanding of the general concept of the earth's shape.

The current findings highlight the need for the careful assessment of performance at both the item level and model level when studying conceptual change. The different patterns of change found across these two measures reflect, in part, the mechanics of the two scoring systems. In the scoring of understanding at the item level, a child's responses to each item were treated as independent. In contrast, in the procedure used for identifying mental models, the critical consideration was the pattern or configuration of responses across items. Hence, a child who responded correctly when asked to draw a picture of the earth and show "where the moon and stars go" would receive full credit for the response in item-level scoring. Whether or not their underlying model was judged to be spherical, however, depended on their responses to questions dealing with other aspects of the earth's shape.

These different approaches to assessing children's knowledge are based on the assumption that mental models represent more than a simple collection of facts and isolated inferences. They involve some form of coherent organization of this knowledge. In support of this view it was found that, in general, only improvements in performance on items that required some form of explanation of the child's view about the earth's shape (Item 9) or some form of inference (Items 11 and 12) were correlated with positive conceptual change. With one exception (i.e., Item 2) improvements on factual questions were not associated with a change in children's mental models.

Recently there has been some discussion of whether young children's beliefs about the earth's shape should be characterized as coherent mental models (cf. Nobes et al., 2003; Siegal, Butterworth, & Newcombe, 2003). Arguing against this view, Nobes et al. (2003) found relatively little correspondence between children's responses to forced choice questions concerning issues such as whether you can "fall off" the earth and whether people live all over the earth's surface.

A number of aspects of our data, however, support the view that there is some degree of coherent structure within children's beliefs about the earth's shape and that such beliefs represent more than collections of fragmented facts. First, the majority of children (93% of the total sample) were classified at pre- or posttest as using a single mental model rather than giving a mixed pattern of responses, albeit with a scoring system that allowed for more "acceptable deviations" from the expected response pattern than the system used by Vosniadou and Brewer (1992). Second, when we examined correlations between children's correct or incorrect responses across the 14 items (see Appendix C) we found a modest degree of coherence. At both pre- and posttests, for example, the accuracy of children's drawings of the earth (Item 6) was moderately correlated with an understanding of "what is above the earth" (Item 3), where people live on the surface (Item 8) with answers to inferential questions about whether the earth has an edge (Item 10). This result is discrepant with the low levels of correspondence between item responses reported by Nobes et al. (2003). This discrepancy may be due to differences in the size of the samples used to examine coherence in the respective studies (132 children in this study compared with 24 to 28 children per sample in Nobes et al.). Most notably, the patterns of model change that we obtained suggest a systematic reorganization of children's beliefs about the earth rather than the simple accrual of new factual information. Positive model change typically involved a shift from a more primitive nonsphere model to a more sophisticated "synthetic" nonsphere model (16.7% of children who showed positive change) or from a synthetic model to a spherical model (79% of children who showed positive change). Shifts from primitive models to a sphere model following training were very rare.

In considering the wider implications of our results, a number of issues still need to be clarified. One question relates to the nature of the information that was presented during training. The training episodes contained new scientific facts, analogies illustrating the size of the earth and the action of gravity, and explicit causal explanations. An important goal issue for the future, therefore, is to assess the relative importance of these different types of information in promoting conceptual change and determine whether the diversity principle applies equally to each.

Three other questions regarding the generalizability of the current results also deserve some attention. The first is whether our findings concerning change in the earth concepts of Australian children would be equally applicable to children in other cultures. Siegal et al., (2003) found that Australian preschool and elementary school children showed a better understanding of the shape of the earth than age peers from a Northern Hemisphere country (England). This difference was attributed to unique aspects of Australian culture, including the fact that from a young age children are made aware of their relatively isolated geographic position below the equator, as well as their close cultural links with the distant countries of the Northern Hemisphere. By extension it might be argued that Australian children are more likely than Northern Hemisphere children to respond in a positive way to the kinds of instruction that we provided.

We acknowledge that such cultural differences in children's initial state understanding of domains like the shape of the earth may have some influence on how quickly children shift their naïve beliefs in response to novel information. However, we would also argue that the promotion of belief change via challenging multiple aspects of the belief is a general feature of human induction that is likely to apply across many cultural contexts. In support we note that the influences of diversity on inductive reasoning have been demonstrated across a range of cultural groups including North American (Osherson et al., 1990) and Korean students (Choi, Nisbett, & Smith, 1998) and, in some domains, the Itzaj Mayans of Central America (Lopez, Atran, Coley, Medin, & Smith, 1997).

A related issue concerns recent criticisms of the interview methods developed by Vosniadou and Brewer (1992) to assess children's understanding of the earth's shape (e.g., Nobes et al., 2003; Siegal et al., 2003; Schoultz, Säljö, & Wyndhamn, 2001). Subsequent studies have introduced a number of modifications to the Vosniadou and Brewer interview, including the use of less ambiguous questions, a forced choice rather than open-ended response format, and the provision of realistic 3D models of the earth. Under such conditions children aged 6 to 7 years appear to show a considerably better understanding of the shape of the earth than was reported by Vosniadou and Brewer (1992). The implications of these findings deserve close attention because our methods of assessing naïve astronomical beliefs were based on those of Vosniadou and Brewer (1992). In response we would make two points. First, it should be noted that the results suggesting that young children have a good grasp of the earth's shape have been largely based on assessments of children's responses at the item level rather than on an analysis of their underlying mental models. Second, as argued earlier, even if our methods did underestimate the initial state of children's earth concepts, this does not invalidate our findings concerning the conditions that promote conceptual change. As long as children's understanding of the earth is below ceiling we would expect our pattern of results regarding the relative efficacy of dual- and single-belief training to be replicated across alternative methods of assessing earth concepts.

Finally, there is a need both to clarify the kind of conceptual change that was found in this study and to examine whether the training methods used are likely to be effective in promoting change in domains outside naïve astronomy. Thagard (1992) identifies at least seven different types of conceptual change. Our data on correlations between item responses and shifts from a nonsphere to sphere model suggest that this involved a number of the different kinds of epistemic change identified by Thagard. Conceptual change was associated with the addition (e.g., you look down to see the earth) and deletion (you can't fall off the edge of the earth) of specific beliefs. It also involved some degree of reorganization such that change was associated with an ability to resolve the contradiction between the apparent flatness shown in children's drawings and knowledge about sphericity.

In considering the generality of our findings to other conceptual domains, we need to be mindful that conceptual change may take very different forms across domains (Siegal et al., 2003; Thagard, 1992). It would therefore be premature to expect that the kinds of interventions used in the current study will always produce the same kinds of changes in domains as different as children's naïve physics, biology, or understanding of emotions and beliefs. Nevertheless, we see the principle that challenging diverse aspects of beliefs rather than repeatedly challenging the same belief as one that is a general feature of induction, and therefore one that seems likely to promote relatively greater levels of change across a range of belief systems or domains.

In summary, we have shown that the diversity principle derived from general models of scientific reasoning and induction is likely to have an important role in promoting change in children's naïve mental models. Our results show that children are more likely to revise their beliefs in the face of challenges to multiple assumptions that underlie these beliefs. It is important to remember, however, that our brief intervention achieved only a modest level of conceptual change overall. This is, perhaps, not surprising given the relatively low proportion of children showing a spherical model at pre-test and the brevity of the intervention. To produce more substantial shifts in children's understanding of basic science phenomena, it may be necessary to consider other factors such as children's beliefs about the relations between different kinds of evidence or data (Kim & Keil, 2003; Medin, Coley, Storms, & Hayes, in press).

Acknowledgments

This research was supported by an Australian Research Council Discovery Grant to the first and third authors. We thank the staff and students of Maryland Public School, Maryland, N.S.W, the Ravenswood School for Girls, Gordon, N.S.W, Queenwood School for Girls, and Mosman Preparatory School, Mosman, N.S.W., for their participation in this study, and Natalie Shoolman for her assistance in data collection. We thank William Brewer and the reviewers of the earlier versions of the manuscript for their insightful comments and suggestions. Parts of this work were presented at the Biennial Meeting of the Society for Research in Child Development, Minneapolis, MN, April 2001.

Appendix A. Adaptation of the Vosniadou and Brewer (1992) interview protocol

The protocol used to assess children's mental models of the earth followed closely the protocol described by Vosniadou and Brewer (1992, p. 551). Items 1 to 7 and 10 to 14 were identical to the corresponding items used by Vosniadou and Brewer (1992), with the same items presented in both pre- and posttests. The alternate forms of Items 8 and 9 are given in italics.

1. What is the shape of the earth?
2. Which way do we look to see the earth?
3. What is above the earth?
4. What is below the earth?
5. What is to the sides of the earth?
6. Can you draw a picture of the earth?
7. Now on this drawing, show me where the moon and stars go. Now draw the sky.
8. Show me where the people live.
8. Suppose you and I were in this picture. Draw two people in the place where we would be.
9. Here is a picture of a house. This house is on the earth, isn't it? How come the earth here is flat, but before you made it round?
9. Here is a picture of a playground. Does the earth seem flat or round? Why does the earth seem flat here but you drew the earth round there?
10. If you walked for many days in a straight line, where would you end up?
11. Would you ever reach the end or edge of the earth?
11a. Is there an end or an edge to the earth?
12. Can you fall off that end or edge?
13. Where would you fall?
14. Now tell me what is down here below the earth (with reference to the child's drawing)?

Appendix B. Summary of the content of the training videos

Size of the earth video (4 training episodes)

SPOKEN CONTENT	VISUAL CONTENT
Training episode 1	
It is hard to imagine that the earth is shaped like a ball when it seems flat if we look around, but it is a matter of the size of the earth and where we are looking. The problem is that the earth is so big that we have trouble picturing it. Let's first have a look at what earth looks like from space. Now we can see that the earth looks like a round ball that shines bright and blue. So why does it look flat from where we are?	Shots of the earth from space.
Training episode 2	
The earth is very large. We can only see a very small part of the earth at a time. The earth is so big that it looks flat but really it isn't. The earth seems flat to us because we are so tiny in comparison to the size of the earth. Let me show you what I mean.	Tiny human figures placed next to globe.

Appendix B. (*continued*)

SPOKEN CONTENT	VISUAL CONTENT
Training episode 3	
Imagine that this rockmelon (cantaloupe) is the earth. It is shaped like a ball. But if we just look at one small piece of the skin, it looks flat. This little piece of skin is like what we can see of the earth. It is like where we live and it looks flat, just like where we live looks flat. But it is part of a round ball. If we put this part back into the rockmelon we can see how something that looks flat can be part of something round. What we see of the earth is flat but it is part of something round like a ball.	Demonstration with a rockmelon (cantaloupe).
Training episode 4	
The earth is so big that we can't see it all at once. That makes it hard to imagine. From space, it is easy to see the earth as a round ball.	Animated characters and rocket taking off.
Narrator: Why don't we go for a trip into space to see for ourselves? Let's climb into this rocket. Get ready for take-off! 5-4-3-2-1!	Characters in a space ship, traveling through space, looking back at the earth to see its shape.
Narrator: Now we are in space, let's have a look at what the earth really looks like. Animated child: Wow! It really is round! Can we travel through space and see what the earth looks like from different places? Narrator: Definitely, let's go! Animated child: Look! It does look like a ball! So no matter where we look at the earth from, it's round. Narrator: Yes, it is! Animated child: Because the earth is so big when we look at it, it looks flat. Narrator: That's right, but now you know that it is really like a big round ball. So what we have learnt is that the earth looks flat because it is so big. But it is actually shaped like a huge round ball. Gravity video (4 training episodes)	
Training episode 1	
The earth is shaped like a ball. Let's pretend that this toy person is a real person living on the earth. How can this person live here on the bottom of the earth without falling off? The answer is gravity.	Shot of a yellow ball. Shot of a toy figure. Shot of a toy falling

Appendix B. (*continued*)

SPOKEN CONTENT	VISUAL CONTENT
Gravity is the force that pulls everything down onto the earth. It is an important force that affects everything on the earth. We can't see gravity, but we can feel it. Let's have a look at some examples of what gravity does.	from the base of a ball. Shot of toy stuck to a ball.
Training episode 2	
The mysterious force of gravity holds us down on the ground. No matter how hard we try to push away from the earth, gravity pulls us back down. When we jump up into the air on a trampoline, gravity is the force that pulls us back down. When we drop something, it falls down toward the ground because of gravity. When a diver leaps off the diving board, gravity pulls him faster and faster towards the pool below. Gravity makes rain fall, and basketballs drop through hoops. Gravity pulling us down gives us our weight. It makes us feel heavy.	Slow motion shot of children jumping. Shot of a child jumping on a trampoline. Shots of a child dropping a book. Shot of a diver. Shot of rain falling. Shot of a basketball. Shot of toddler on a set of scales.
Training episode 3	
Gravity works the same way as a magnet does on your fridge at home. It holds things in place. Gravity keeps objects from floating into space.	Shot of a magnet holding some paper in place.
Training episode 4	
Gravity pulls everything toward the center of the earth. No matter where you are on the planet, gravity pulls objects toward the center. This brings us back to our first question. If the earth is round like a ball, how can people live all over it? Won't people fall off if they live on the sides or the bottom of the ball? We now know that people won't fall off the earth, because of a force called gravity. Gravity holds everything onto the earth no matter what part of the earth we are on. This special force called gravity is how people can live on all different parts of the earth without falling off.	A picture of the earth with arrows pointing and moving towards the center of the earth. Cartoon of earth with figures all over the edge again. Shot of a slowly spinning globe with tiny figures of people attached all around the earth.

Appendix C. Pretest and posttest correlations between response accuracy for individual interview items

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		.33*	.08	.12	.11	.15	−.08	.29*	.04	.15	.17	.12	.16	.06
2	*.32**		.08	.12	.11	−.14	−.17	.00	.04	.15	.26*	−.19	.16	.14
3	*.14*	*.08*		.15	.25*	.28*	−.07	.28*	−.01	.12	.03	−.19	−.17	−.13
4	*−.07*	*.16*	*.18*		.25*	−.05	−.10	.06	.19	.14	.07	.07	.05	.07
5	*.18*	*.07*	*.17*	*.28**		.17	.17	.28*	.00	−.02	.16	.20	.09	.46*
6	*.25**	*−.12*	*.44**	*−.02*	*.12*		.26*	.57*	.17	.26*	.00	.12	.17	.07
7	*−.12*	*−.16*	*.13*	*.11*	*.09*	*−.10*		−.01	−.08	.00	−.05	−.01	.06	.28*
8	*.33**	*.17*	*.25**	*.03*	*.25**	*.63**	*−.16*		.17	.26*	.19	.18	.23	.18
9	*.17*	*−.10*	*−.13*	*.01*	*.16*	*.15*	*.06*	*.25**		.32*	.06	.10	.04	.04
10	*−.02*	*−.15*	*−.13*	*−.06*	*−.09*	*.05*	*.2*	*.02*	*.21*		.44*	.09	.11	−.13
11	*−.05*	*.16*	*−.09*	*.04*	*.04*	*−.12*	*−.09*	*−.08*	*.16*	*.47**		.37*	.30	−.04
12	*−.06*	*−.15*	*−.15*	*−.15*	*.13*	*−.11*	*.07*	*−.13*	*.12*	*.21*	*.34**		.88*	.04
13	*−.06*	*.19*	*−.13*	*−.11*	*.04*	*−.10*	*−.05*	*−.12*	*.08*	*.16*	*.31**	*.90**		−.07
14	*.22*	*.06*	*−.07*	*.33**	*.37**	*.15*	*.25**	*.16*	*.25**	*.17*	*.09*	*−.19*	*−.17*	

Note. Pretest correlations are given in normal typeface above the diagonal; posttest correlations are given in italics below the diagonal. An asterisk denotes significant at $p < .01$, two-tailed. This conservative significance criterion was adopted because of the large numbers of correlations computed.

References

Bacon, F. (1620/1898). *Novum organum*. London: George Bell and Sons.

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (2000). Explanation in scientists and children. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 279–298). Cambridge, MA: The MIT Press.

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

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology, 21*, 13–19.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1–49.

Choi, I., Nisbett, R. E., & Smith, E. E. (1998). Culture, category salience, and inductive reasoning. *Cognition, 65*, 15–32.

Diakidoy, I. A., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: A comparison of the effectiveness of two instructional approaches. *Learning and Instruction, 11*, 1–20.

Diakidoy, I. A., Vosniadou, S., & Hawks, J. D. (1997). Conceptual change in astronomy: Models of the earth and of the day/night cycle in American-Indian children. *European Journal of Education, 12*, 159–184.

di Sessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction, 10*, 105–225.

Gopnik, A., & Meltzoff, A. N. (1994). *Words, thoughts and theories*. Cambridge, MA: MIT Press.

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

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences, 5*, 97–127.

Harris, P. L. (1996). Thinking by children and scientists: False analogies and neglected similarities. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 294–315). Cambridge, England: Cambridge University Press.

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.

Heit, E., & Hahn, U. (2002). *Knowledge effects in children's diversity-based reasoning*. Unpublished manuscript.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.

Horwich, P. (1982). *Probability and evidence*. New York: Cambridge University Press.

Kikas, E. (1998). The impact of teaching on students' definitions and explanations of astronomical phenomena. *Learning & Instruction, 8*, 439–454.

Keil, F. C. (1989). *Concepts, kinds and conceptual development*. Cambridge, MA: MIT Press.

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

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674–689.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). London: Cambridge University Press.

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 & 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.

Mali, G. B., & Howe, A. (1979). Development of earth and gravity concepts among Nepali children. *Science Education, 63*, 685–691.

Medin, D. L., Coley, J. D., Storms, G. & Hayes, B. K. (in press). A relevance theory of induction. *Psychonomic Bulletin and Review*.

Nagel, E. (1939). *Principles of the theory of probability*. Chicago: University of Chicago Press.
Nobes, G., Moore, D., Martin, A., Clifford, B., Butterworth, G., Panagiotaki, G., et al. (2003). Children's understanding of the earth in a multicultural community: Mental models or fragments of knowledge? *Developmental Science, 6*, 74–87.
Nussbaum, J. (1979). Children's conceptions of the earth as a cosmic body: A cross age study. *Science Education, 63*, 83–93.
Nussbaum, J., & Novak, J. D. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education, 60*, 535–550.
Nussbaum, J., & Sharoni-Dagan, N. (1983). Changes in second grade children's preconceptions about the earth as a cosmic body resulting from a short series of audio tutorial lessons. *Science Education, 67*, 99–114.
Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.
Samarapungavan, A., Vosniadou, S., & Brewer, W. F. (1996). Mental models of the earth, sun and moon: Indian children's cosmologies. *Cognitive Development, 11*, 491–521.
Schoultz, J., Säljö, R., & Wyndhamn, J. (2001). Heavenly talk: Discourse, artifacts, and children's understanding of elementary astronomy. *Human Development, 44*, 103–118.
Siegal, M., Butterworth, G., & Newcombe, P. A. (2003). Culture and children's cosmology. *Developmental Science, 6*, 72–85.
Sneider, C. I., & Ohadi, M. M. (1998). Unraveling students' misconceptions about the earth's shape and gravity. *Science Education, 82*, 265–284.
Solomon, G. E. A., & Cassimatis, N. L. (1999). On facts and conceptual systems: Young children's integration of their understandings of germs and contagion. *Developmental Psychology, 35*, 113–126.
Spellman, B. A., Lopez, A., & Smith, E. E. (1999). Hypothesis testing: Strategy selection for generalizing versus limiting hypotheses. *Thinking and Reasoning, 5*, 67–91.
Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.
Von Eye, A., & Niedermeier, K. E. (1999). *Statistical analysis of longitudinal categorical data in the social and behavioural sciences*. Mahwah, NJ: Erlbaum.
Vosniadou, S. (1991). Conceptual development in astronomy. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 147–177). Hillsdale, NJ: Erlbaum.
Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 49–69.
Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.
Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science, 62*, 111–121.
Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon (Ed.), *Handbook of child psychology* (Vol. 2, (5th ed., pp. 523–573). New York: Wiley.