Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
J. Experimental Child Psychology 87 (2004) 171–200
Journal of Experimental Child Psychology
www.elsevier.com/locate/jecp

What are theories for? Concept use throughout the continuum of dinosaur expertise

Kathy E. Johnson,[a,*] Paul Scott,[b] and Carolyn B. Mervis[c]

[a] *Department of Psychology, Indiana University–Purdue University Indianapolis Indianapolis, IN 46202, USA*
[b] *Department of Psychology, Rockhurst University, Kansas City, MO 64110, USA*
[c] *Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY 40292, USA*

Received 6 August 2002; revised 4 December 2003

Abstract

Although it is now well established that object concepts are situated within broader systems of theoretical knowledge, it is less clear how theories influence the use of object concepts at various points throughout the continuum of expertise. Two studies were conducted to investigate the impact of specific theories (concerning dinosaurs) and overarching framework theories (of biology) on children's and adults' performance on categorization tasks involving familiar and less familiar concepts. Although expertise increased the quantity of deep feature knowledge possessed by children and increased their understanding of biologically adaptive relations among features, few aspects of children's performance generalized beyond highly familiar dinosaurs. Children's specific theories related to dinosaurs were empirically constrained and relatively dissociated from other types of biological knowledge. The interaction of specific concept knowledge with broader framework theories of biology throughout the continuum of expertise is considered.

Keywords: Categorization; Expertise; Reasoning; Biological knowledge

---

Introduction

Some young children become fascinated with particular classes of objects and doggedly pursue their parents to read them books, play them videotapes, and acquire

---

[*] Corresponding author. Fax: 1-317-274-6756.
E-mail address: kjohnso@iupui.edu (K.E. Johnson).

doi:10.1016/j.jecp.2003.12.001

toys and models that help them to master the information contained within the domain. Consequently, these children may ultimately end up possessing concepts pertaining to kinds of things, such as dinosaurs and trains, that are far more differentiated than those possessed by most adults. Possession of specialized knowledge can have far-reaching effects on various aspects of information processing, including memory (Chi, 1978; Chi & Koeske, 1983), inference generation (Chi, Hutchinson, & Robin, 1989), reasoning (Gobbo & Chi, 1986), and categorization (Chi et al., 1989; Johnson & Eilers, 1998; Johnson & Mervis, 1994). In some cases, adults' typical performance advantage can even be eliminated when children's domain-specific knowledge is extensive (Chi, 1978; Johnson & Eilers, 1998). Findings such as these have led many researchers to conclude that conceptual development is domain specific and driven predominantly by knowledge acquisition (e.g., Carey, 1985, 1995; Keil & Lockhart, 1999; Springer, 1999).[1]

It is clear that experience with specific categories of objects is sufficient for developing elaborate networks of concepts and their concomitant features, even during the preschool period (Chi & Koeske, 1983; Johnson & Mervis, 1994). What is not yet understood is the specific interaction between children's broader theoretical knowledge related to these domains and conceptual structure. It is now well established that theoretical knowledge and conceptual structure are inextricably linked (e.g., Barrett, Abdi, Murphy, & Gallagher, 1993; Carey, 1985; Gelman, 1996; Keil, 1994b; Medin, 1989; Murphy & Medin, 1985). Concepts are individuated mental representations that form coherent units within semantic memory (Clark, 1983; Rosch, 1978). However, they are positioned within larger systems of knowledge that comprise theories. Theories influence which aspects of specific concepts individuals attend to and help to constrain similarity relations and patterns of inferential reasoning throughout the life span. The purpose of the current research was to explore variations in concept use as a function of various levels of specific theoretical knowledge pertaining to dinosaurs. A secondary question concerned the impact of specific theoretical knowledge about dinosaurs on the categorization of related, but less familiar, biological concepts (e.g., fish, birds, hypothetical dinosaurs). The article begins by introducing various types of theories and considering the effects of fact acquisition on theory construction and use. It then considers the means by which theories could mediate the effects of expert knowledge on related, but less familiar, biological concepts.

The notion that concepts are embedded within theories is now well established (Keil, 1995; Keil & Lockhart, 1999; Medin, 1989). When a child accepts that an eel is a fish but a dolphin is not, it is not because of a simple alteration in conceptual feature salience or prototypicality but rather because of a basic shift in the child's theoretical beliefs concerning the biological relations among fish and other animals. Theories are explanatory frameworks within which individual concepts are situated (Carey, 1985; Murphy, 1993). This implies that concepts possess a hybrid structure

---

[1] For a detailed review of issues pertaining to the definition of domains, see Hirschfeld and Gelman (1994). Johnson and Mervis (1998) also addressed the issue of defining domains associated with expertise on object concepts.

throughout development, consisting of an interdependent mixture of both associationist and theory-based components (Keil & Lockhart, 1999).

Just as concepts vary in terms of their levels of specificity, theories differ in terms of their levels of explanation. Wellman (1990) used the philosophy of science distinction between specific theories and more global theories (which Wellman refers to as "framework theories") in characterizing the child's developing theory of mind. Barrett and her collaborators (Barrett et al., 1993; Barrett, Abdi, & Sniffen, 1992) have usefully applied this distinction between framework and specific theories to the concepts-in-theories view. Framework theories are more abstract and help to define the ontology of the domain. Framework theories highlight causally relevant properties that structure domains and specify the kinds of explanations that are satisfactory within them. For example, attributing a change in size to growth processes would be acceptable within a biological domain but unacceptable within an artifact domain. Framework theories also shape how children learn and reason about domain-specific concepts.

Specific theories are constrained by the framework theory within which they are situated. Barrett and colleagues (1992, 1993) suggested that specific theories offer more detailed explanations for observations within each particular theoretical domain. Explanations of the bases for feature correlations indicative of specific concepts, whether learned directly from more knowledgeable sources or inferred based on perceived features, would constitute specific theories. It seems likely that experts on domains of taxonomic concepts acquire multiple layers of specific theories corresponding to concepts at increasingly more precise levels of categorization. For example, children who are experts on dinosaurs understand the features that determine membership in the dinosaur domain, but they also understand relations among various families of dinosaurs (e.g., hadrosaurs, ceratopsians) and the specific concepts included within those families (e.g., triceratops is a more recently evolved ceratopsian than is protoceratops), not to mention the bases for feature correlations indicative of specific concepts (e.g., triceratops' bony frill protected its neck and counterbalanced heavy facial horns). The current research is focused on two particular functions of specific theories: explaining bases for similarity decisions (Study 1) and generating inferences concerning feature relations (Study 2).

Framework and specific theories differ sharply in their potential for modification through fact acquisition. Specific theories are most likely to change as a function of experience with domain-specific exemplars; thus, they are most apt to differ considerably as a function of expertise. Although framework theories are subject to change, they are not apt to be revised based strictly on changes to specific theories. Change in framework theories is "typically a product of internal inconsistencies or changes in the scope of the framework theory" (Barrett et al., 1992, p. 279). As such, framework theories are good candidates for explaining developmental differences between adults and children with comparable levels of knowledge concerning specific concepts. Springer (1999) argued that the acquisition of naive biological theories depends critically on both fact acquisition and certain key inferences generated based on the facts that have been learned. Although specific theories can be altered when facts are learned, the basis for framework theory change has been subject to debate (cf. Carey, 1995; Inagaki, 1997; Springer, 1999). However, it appears to involve some form of qualitative reorganization of specific theoretical knowledge, based on an internal analysis of stored representations. For biological knowledge, this reorganization likely takes place during the middle childhood years. It also may proceed in an abstract-to-concrete direction, whereby children develop more concrete representations of biological mechanisms only after relatively sparse abstract representations of animals have been formed (Simons & Keil, 1995).

Within the expertise literature, it is widely accepted that experts' proficiency is constrained to the specific area in which knowledge has been acquired (Alexander, Johnson, & Schreiber, 2002; Chi, Glaser, & Farr, 1988; Ericsson & Lehmann, 1996; Glaser, 1987). Evidence for this stance comes from skill domains characterized by specialized procedural knowledge such as chess, medical diagnosis, and bridge. Johnson and Mervis (1998) argued that this assumption should be qualified for types of expertise characterized by conceptual knowledge associated with common taxonomic categories. They found that adult experts on birds based triad similarity solutions on deep features related to taxonomic membership for both familiar bird triads and triads involving biological concepts that were unfamiliar (tropical fish). In addition, bird experts rated surface features indicative of taxonomic categories in fish (e.g., fin shape) as significantly more perceptually salient than did novices. Johnson and Mervis concluded that this generalization was mediated by experts' biological theories, which helped to guide the extraction of subtle perceptual features of fish that were aligned with those recruited within the domain of birds. Because adults' framework theories of biology specify which perceptual features of biological kinds generally are predictive of conceptual relations, it is possible that such generalization may be found only among adult experts. Both of the current studies investigated the degree to which individuals with high specific knowledge on dinosaurs can extend this knowledge to parallel situations involving unfamiliar biological concepts. Study 1 specifically involved a replication with child dinosaur experts of the same kind of triad task that Johnson and Mervis used with adult bird experts.

Theories may be recruited whenever individuals are pressed to justify their categorization decisions and whenever individuals move beyond information that is perceptually available when processing information related to concepts. Four specific functions of theories were investigated in the current studies. First, when multiple bases for similarity relations are available, theories direct which features are recruited when making and subsequently justifying similarity decisions. Second, theories may enable the generalization of solution strategies from familiar concepts to less familiar, but related, concepts. Third, theories provide an explanatory framework from which to infer the functional or behavioral correlates of perceptual features. Finally, theories enable individuals to determine whether particular combinations of features are plausible or not plausible. The first two functions were explored in Study 1, and the second, third, and fourth functions were investigated in Study 2.

Study 1: Effects of domain knowledge on making and justifying similarity decisions

When people make decisions of similarity, theories help to constrain the dimensions along which similarity is computed (Murphy, 1993). Theories are undoubtedly recruited whenever people are asked to justify their decisions of similarity. Many studies have shown that experts and novices select different bases for similarity decisions (Chi, Feltovich, & Glaser, 1981; Johnson & Mervis, 1994, 1998). Less is known about whether such bases may generalize to related, but unfamiliar, categories. Johnson and Mervis (1998) found that adults with expertise on birds spontaneously generalized their bases for similarity relations from bird triads to much less familiar fish triads. One goal of the current study was to determine whether a parallel result obtains among children. Triads were created such that deep feature (conceptual) bases for similarity were pitted against highly salient surface feature (morphological) bases for similarity. In each triad, a target exemplar shared a striking perceptual resemblance (in terms of body shape or color) with one exemplar (the surface feature match). At the same time, the target shared a nonvisible conceptual relation with a perceptually dissimilar exemplar (the deep feature match).

A second goal of the study was to examine the degree to which specific theories influenced justifications of similarity decisions by comparing the responses of children with high, moderate, and low levels of knowledge. Two additional types of data were collected relevant to each similarity decision: whether the children knew relevant deep features when explicitly asked about them and whether the children spontaneously produced deep (conceptual) features when asked to justify their similarity decisions. We anticipated that children with high knowledge would be more apt to both recruit and talk about deep features, whereas children with less knowledge would be in a transitional period where facts related to deep features were known, but not recruited, when either making or justifying similarity decisions.

Method

Participants

Participants were 32 children (27 boys and 5 girls), between 6 and $8\frac{1}{2}$ years of age, who were recruited based on their expressed interest in dinosaurs. This age range was targeted because children are likely to possess immature framework theories of biology, yet they still have high levels of domain knowledge that can be articulated. A total of 27 children were recruited in Atlanta, Georgia, from three sources: (a) inquiries regarding an article about the research published in a local newspaper, (b) registration lists of summer programs at a local zoo, and (c) a university child subject pool. In addition, 5 children were recruited and tested in Indianapolis, Indiana, at least 6 months after participating in an unrelated study of dinosaur expertise and problem solving. An additional 9 children were excluded (6 due to response bias, 2 due to experimenter error, and 1 due to loss of interest). Children received a small gift for participating.

Materials

Knowledge assessment measures. Three types of measures were used to assess children's relative level of knowledge of dinosaurs. First, a parent questionnaire was created to assess both parents' and children's interest in and knowledge of dinosaurs and fish. Three types of information were obtained through the questionnaire. First, parents rated their own and their children's level of knowledge of dinosaurs on a 9-point scale (1 = *very low level of knowledge*, 4 = *average level of knowledge*, 9 = *exceptional level of knowledge*). The same scale was used to make parallel ratings of fish knowledge. Second, parents rated both their own and their children's level of interest in dinosaurs and fish on a 9-point scale (1 = *no interest*, 4 = *average level of interest*, 9 = *extremely interested*). Finally, parents were given a checklist of 15 dinosaur names and were asked to indicate which of these names they had heard their children produce referentially.

The second knowledge measure involved a set of 36 realistic color pictures of individual dinosaurs, presented on laminated index cards in the same canonical orientation, for use in a name production task. A wide range of dinosaurs was presented, including all of the dinosaurs that Gobbo and Chi (1986) used to assess children's expertise as well as other species from representative dinosaur groups (Norman, 1985).

The third converging knowledge measure was the brief test of dinosaur knowledge used by Gobbo and Chi (1986) and Chi and colleagues (1989) to assign children to expert and novice conditions. The test consists of 20 fill-in-the-blank items that assess multiple aspects of children's understanding of the dinosaur domain (e.g., fossils, functions of body parts, name derivations, habitats, characteristic behaviors).

Intelligence measure. The revised Peabody Picture Vocabulary Test (PPVT-R, Form L) was used to provide an estimate of general intelligence. Scores on this test correlate positively with several intelligence measures, including the Stanford–Binet and the verbal component of the Wechsler Intelligence Scale for Children (Dunn & Dunn, 1981). The PPVT-R involves trials composed of sets of four pictures. The children are asked to point to the picture in each set associated with the word provided by the tester.

Triad task stimuli. Triad task stimuli were digitized photographs and realistic color drawings of objects, presented to scale and in the same canonical orientation on a 13-in. Apple RGB color high-resolution monitor. No labels were provided. Pictures of 36 dinosaurs and 36 fish were obtained from a variety of published references and field guides (e.g., Axelrod et al., 1987; Wilson, 1986). Pictures of dinosaurs were presented against a white background, and pictures of fish were presented against a rectangular section of the aquarium within which they were photographed. (It was not possible to digitally remove this background without distorting transparent segments of the fins.) A list of the pictures included in all test triads is included in the appendix. Pictures of 18 additional artifact and animal exemplars were used for practice triads. None of the pictures was labeled. Triads were constructed such that a "deep feature" solution was always pitted against a "surface feature" solution. For example, the allosaurus–compsognathus–iguanodon triad included two carnivores (allosaurus–compsognathus) and two exemplars that were highly similar in terms of their overall morphology (allosaurus–iguanodon) but that did not share the deep feature "diet" relation. Deep feature bases for solutions were related either to taxonomic membership within the domain of dinosaurs (e.g., Dinosaur A, Dinosaur B, nondinosaur that resembles Dinosaur A or B), to taxonomic membership within a subgroup of dinosaurs (e.g., Hadrosaur A, Hadrosaur B, nonhadrosaur that resembles Hadrosaur A or B), or to similarity in diet (e.g., Carnivore A, Carnivore B, herbivore that resembles Carnivore A or B). Surface feature solutions were based on salient visible similarities in body shape, color, posture, and size.

Procedure

Children were tested in university-based laboratories during a single session that lasted approximately 75 min, including at least two short breaks. The procedure for children consisted of five tasks administered in the following order: triad task with solution justifications, deep feature knowledge assessment, name production, the brief test of dinosaur knowledge, and the PPVT-R. While each child was tested, a parent (typically the mother) completed the questionnaire developed to gauge children's level of interest in and knowledge of dinosaurs.

Triad task. The triad task was used to assess the impact of children's theories on the features recruited during categorization decisions. To ensure that the task was understood, six practice triads were administered first. Practice triads featured pictures of artifacts and animals that were not included in the test triads and were used to demonstrate that perceptual similarity was not the only basis for grouping objects. On four practice triads, surface feature similarity was correlated with deep feature similarity (e.g., school bus–city bus–airplane), and on others, they were not (e.g., knife–scissors–umbrella). On two triads, surface feature similarity and deep feature similarity were orthogonal (e.g., orange–banana–ball). In reference to each triad, children were instructed to "touch the two that are most like the same kind of thing." Triads were presented on a computer monitor, and responses were recorded through a touch screen. Triad presentation and online data collection were accomplished through specially created Authorware (Version 1.6) programs. If children did not provide the expected answer to practice triads, they were given feedback concerning the deep feature solution and the triad was repeated until it was solved correctly.

Following the practice triads, two sets of 12 test triads were administered. One set involved dinosaur exemplars and one set involved fish exemplars, and the order in which the two sets were presented was counterbalanced across participants. Trials within a set of triads were presented in random order. In reference to each of the 24 triads, children were asked, "Which two of these are most like the same kind of thing?" Children responded by touching two pictures on each triad, and responses were recorded through Authorware. Solutions in which children paired the target exemplar with the perceptually similar exemplar were coded as surface feature solutions. Solutions in which children paired the target with the conceptually related exemplar were coded as deep feature solutions. Solutions in which children paired the two nontarget exemplars were considered uninterpretable. All deep feature bases for solution are included in the appendix. Following children's responses, the experimenter asked, "Can you tell me why?" Children's justifications for their solutions were audiotaped and later transcribed.

Assessment of deep feature knowledge. After children completed all 24 triads, they were shown the triads a second time in the same order as before. To assess whether the children knew the features on which deep feature solutions were based, they were asked specific deep feature questions concerning each of the three animals in the triad. For example, for each of the three animals in a dinosaur triad involving a potential deep feature solution based on diet, children were asked, "Does this one eat meat?" Children were instructed to give what they thought was the best answer, even if they were unsure. No feedback concerning the correctness of answers was provided.

Knowledge assessment. In the name production task, children were shown a series of 36 pictures of individual dinosaurs and were asked to name each one. Feedback was not provided. The brief test of dinosaur knowledge was administered aurally. Both tasks were audiotaped.

Results and discussion

Participants were assigned to knowledge groups based on composite knowledge scores that reflected each of four different knowledge indexes: (a) parents' ratings of children's relative level of dinosaur knowledge, (b) number of correct names produced at either the species or family level for the set of 36 realistic color pictures of dinosaurs presented in the laboratory, (c) number of dinosaur names that parents reported their children had produced (out of 15 names listed on the questionnaire), and (d) children's scores on the dinosaur knowledge test (out of 20). Each of the four scores was standardized, and then *z* scores were summed to create an overall knowledge index. Interrelations among the individual knowledge measures, PPVT-R scores, and the overall knowledge index are presented at the top of Table 1. All four knowledge measures were highly intercorrelated, although parents' ratings of children's knowledge levels were less strongly, but still significantly, positively related to the other three measures. PPVT-R scores were significantly related only to performance on the brief test of dinosaur knowledge. Standard scores were added, rather than averaged, because knowledge is a cumulative construct and there was concern that extreme scores on one measure (e.g., parental rating) might distort performance on other measures if means were computed. Because knowledge scores were based largely on explicit (rather than implicit) tests of dinosaur knowledge, the measure could have underestimated children's domain competence.

Natural breaks in the distribution of the cumulative knowledge index were used to divide participants into low, moderate, and high knowledge groups.[2] Descriptions of the children assigned to each of the three knowledge conditions are presented in the top portion of Table 2. One-way analyses of variance (ANOVAs) were conducted to determine whether PPVT-R scores, ages, or children's levels of knowledge of fish

---

[2] We recognize that expertise is a relative phenomenon and that the continuum of knowledge has many intermediate stages of competency. The current article concentrates on three points along this continuum: high, moderate, and low levels of knowledge.

Table 1
Relations among knowledge measures and PPVT-R scores

Study 1 ($N = 32$)	Name production	Name checklist	Parent rating	Knowledge test	Composite index
Name checklist	.84**				
Parent rating	.68**	.59**			
Knowledge test	.81**	.71**	.62**		
Composite index	.94**	.89**	.82**	.89**	
PPVT-R score	.26	.24	.11	.52**	.32
Study 2: Children ($N = 46$)	**Name production**	**Name checklist**	**Parent rating**	**Composite index**	
Name checklist	.70**				
Parent rating	.55**	.60**			
Composite index	.87**	.89**	.83**		
PPVT-R score	.41**	.45**	.36*	.47**	
Study 2: Adults ($N = 27$)	**Name production**	**Name checklist**	**Self-rating**	**Composite index**	
Name checklist	.78**				
Self-rating	.71**	.82**			
Composite index	.90**	.94**	.92**		
PPVT-R score	.71**	.66**	.54**	.69**	

* $p = .05$ (two-tailed).
** $p = .01$ (two-tailed).

Table 2
Participant characteristics (mean scores)

	Measure	High knowledge	Moderate knowledge	Low knowledge
Study 1		$N = 9$	$N = 10$	$N = 13$
	Age (months)	86.4 (11.3)	86.7 (7.5)	84.31 (8.9)
	PPVT-R score	121 (12.22)	116 (14.08)	112 (16.02)
	Fish knowledge (rating from 1 to 9)	4.32 (1.79)	3.43 (2.03)	3.31 (1.93)
	Dinosaur knowledge index	4.66 (1.30)	0.16 (.44)	–3.35 (1.57)
Study 2	Children	$N = 17$	$N = 14$	$N = 15$
	Age (months)	81.65 (8.7)	87.64 (11.9)	92.4 (14.1)
	PPVT-R score	122 (8.60)	118 (11.36)	108 (9.57)
	Bird knowledge (parent rating from 1 to 8)	3.62 (0.78)	3.36 (1.15)	2.93 (0.81)
	Dinosaur knowledge index	2.75 (0.34)	–0.30 (0.70)	–2.84 (1.02)
	Adults	$N = 8$	$N = 10$	$N = 9$
	Age (years)	33 (13)	29 (9)	30 (17)
	PPVT-R score	127 (9.45)	109 (11.91)	105 (8.10)
	Bird knowledge (self-rating from 1 to 8)	3.50 (0.76)	2.70 (1.06)	2.78 (1.78)
	Dinosaur knowledge index	3.19 (1.62)	0.25 (0.56)	–3.12 (0.95)

Note. Standard deviations are in parentheses.

varied significantly across the three knowledge groups. None of these differences was significant (all *p*s $> .10$).

Recruitment of deep features during triad solutions

For each triad, we first determined whether or not children made a deep feature solution by selecting the target animal and the deep feature match as the two that "were most like the same kind of thing." Indeterminate solutions were generated less than 5% of the time across the three groups of children. Due to experimenter error on one fish triad, responses for that triad were dropped, necessitating the use of proportions to compare deep feature solutions across the two types of triads. Mean proportions of deep feature solutions were compared in a 3 (Knowledge: high, moderate, or low) × 2 (Order of Domain Presentation) × 2 (Triad Type: dinosaurs or fish) mixed ANOVA, with triad type as the within-subject factor. The only significant effects were of triad type, $F(1, 26) = 15.50$, $p = .001$, and order of domain presentation, $F(1, 26) = 4.29$, $p < .05$. The main effect of triad type was attributable to children generating significantly more deep feature solutions for dinosaur triads ($M = 24\%$) than for fish triads ($M = 4\%$). The main effect of order of domain presentation was attributable to higher proportions of deep feature solutions being generated by children who were presented with dinosaur triads first, suggesting that once children's representations of dinosaurs were activated, children were more apt to generalize solutions based on deep features to the less familiar domain. Although interactions with knowledge group only approached significance (presumably due to the small samples of children available), we tested next whether the test order effect held up equally well for all three knowledge groups when considering only deep feature solutions for dinosaur triads. Although these results must be interpreted with caution, the order effect appeared to be driven by the performance of children in the low-knowledge group, $t(11) = 3.79$, $p < .01$. As depicted in Table 3, children in the moderate-knowledge and high-knowledge groups generated comparable proportions of deep feature solutions for dinosaur triads, regardless of the order in which the dinosaur and fish triads were presented. That is, by the time moderate levels of knowledge had been acquired, children's deep feature solutions (although still relatively rare) appeared to be more stable and less susceptible to order effects. The preponderance of surface feature solutions generated by all groups presumably was due to the striking perceptual similarities shared by targets and their respective surface

Table 3
Mean proportions of deep feature solutions generated across triad types

First domain presented	Knowledge group	Dinosaur triads	Fish triads
Dinosaurs	High	.22 (.20)	.08 (.11)
	Moderate	.18 (.13)	.09 (.15)
	Low	.21 (.09)	.08 (.06)
Fish	High	.25 (.08)	.00 (.00)
	Moderate	.10 (.10)	.04 (.04)
	Low	.05 (.06)	.05 (.07)

Note. Standard deviations are in parentheses.

feature matches. In contrast to the solutions of fish triads by adult bird experts (Johnson & Mervis, 1998), children rarely generated deep feature solutions for triads involving exemplars from the unfamiliar domain.

Knowledge of deep features

We next wanted to verify whether there were differences among children in the three knowledge groups in terms of their knowledge of deep features associated with dinosaur triads. For each child, the proportion of triads for which he or she correctly indicated that the target and deep feature match possessed the relevant deep feature, but the surface feature match did not, was determined. A mixed analysis of variance with two between-groups factors (Knowledge: low, moderate, or high; Order of Domain Presentation: dinosaur or fish first) and one within-group factor (Triad Type: dinosaurs or fish) was conducted, with the number of triads for which children reported correct knowledge of deep features as the dependent variable. There were main effects of knowledge, $F(2, 26) = 5.11$, $p = .01$, and triad type, $F(1, 26) = 55.09$, $p < .001$. Furthermore, there was a significant interaction between triad type and knowledge, $F(2, 26) = 8.42$, $p < .01$. No effects involving order of domain presentation were significant.

To explicate the Triad Type × Knowledge interaction, separate one-way ANOVAs were conducted for each domain. The means and standard deviations from this analysis are depicted in Table 4. A significant main effect of knowledge emerged only for the dinosaur domain, $F(2, 29) = 8.69$, $p < .01$. The main effect of knowledge was followed up by a linear trend analysis, which revealed a strong linear effect of deep feature knowledge as knowledge level increased, $F(1, 29) = 16.81$, $p < .001$. Thus, as knowledge increased, children were more apt to comprehend the deep feature bases for more triads. Children's performance on the deep feature knowledge questions was also compared with chance (defined as 1.5 of the 12 triads) using one-sample t tests. Only children in the moderate-knowledge group, $t(9) = 3.23$, $p = .01$, and the high-knowledge group, $t(8) = 5.92$, $p < .001$, were consistently above chance levels in their responses to deep feature questions for dinosaurs. All groups performed at chance levels when responding to deep feature questions for fish. Clearly, the moderate- and high-knowledge groups knew substantial amounts of information regarding deep features for dinosaurs but knew virtually no deep feature information for the fish. Low-knowledge children had little or no knowledge of deep features for either domain, despite their professed interest in dinosaurs.[3]

---

[3] A follow-up linear trend analysis was conducted on the proportion of deep feature triad solutions generated out of the subset of triads identified in this analysis (i.e., triads for which children correctly indicated that the target and deep feature match possessed the relevant deep feature but that the surface feature match did not). There was a strong linear effect of deep feature solutions as knowledge level increased, $F(1, 29) = 8.87$, $p < .01$. Thus, when facts related to deep features had been acquired, children with more knowledge were more likely to make deep feature solutions than were children with less knowledge.

Table 4
Mean proportions of triads for which children evidenced knowledge of deep features

Knowledge group	Dinosaur triads	Fish triads
High	.54 (.21)	.06 (.05)
Moderate	.29 (.16)	.12 (.10)
Low	.21 (.18)	.06 (.06)

Note. Standard deviations are in parentheses.

Conceptually based justifications of similarity decisions

Children's verbal justifications for their triad solutions provided a window through which we could potentially glimpse the nature of the specific theories that drove those solutions. Verbal justifications were coded as "conceptually based" if they referred to bases for similarity that were not visible and that pertained to concepts (or groups of concepts) from within the relevant domain (e.g., they both have bird hips and they both eat plants, so they are ceratopsians). All justifications were coded by one of the authors, and then a second individual recoded 20% of the transcripts. Reliability between the two coders (agreements/agreements + disagreements) was 93%. Coding was conducted prior to determining children's relative knowledge levels to ensure that coders were blind to knowledge group membership.

The mean proportion of conceptually based propositions out of all propositions produced was calculated for each child across both the dinosaur and fish triads. Preliminary analyses indicated no effects involving the order in which domains were presented, so we collapsed across this factor. Computed proportions were compared in a mixed ANOVA with one between-groups factor (Knowledge: low, moderate, or high) and one within-group factor (Triad Type: dinosaurs or fish). There were significant main effects of both knowledge, $F(2, 29) = 8.39$, $p < .001$, and triad type, $F(1, 29) = 10.96$, $p < .01$. More specific conceptually based propositions were generated for solutions involving dinosaurs than for those involving fish, and production increased as a function of knowledge across both types of triads. Separate linear trend analyses conducted for each domain indicated a strong linear effect of conceptually based justifications for both domains as knowledge level increased (dinosaurs: $F(1, 29) = 12.28$, $p < .01$; fish: $F(1, 29) = 11.85$, $p < .01$). Although children's actual solutions did not generalize to the less familiar domain, the types of explanations offered for their solutions were fairly consistent across the two domains. This effect is reminiscent of Inagaki's (1990) finding that 5-year-olds who raised goldfish were able to use their knowledge as the source in generating inferences about unfamiliar, but related, animals (e.g., frogs).

In sum, children with high levels of specific knowledge were surprisingly unwilling to base their similarity decisions on deep features related to conceptually salient relations such as diet and taxonomic relations, at least when these deep feature similarities were pitted against highly salient surface feature similarities. Deep feature solutions were particularly rare when triads from the less familiar domain of fish (for which surface feature solutions were nearly exclusively generated) were solved first. Order effects were weaker when knowledge was relatively high. In contrast, adult bird experts invariably solved similar types of triads based on specific attributes correlated with deep features and even transferred this strategy to fish categories they knew significantly less about (Johnson & Mervis, 1998). We were unable to recruit adults with comparable levels of dinosaur knowledge for the current study. However, we have found a similar pattern of results for adults and children on a subordinate category extension task (Johnson & Eilers, 1998), where adults generalized their extension strategies to novel shorebirds but children did not. Additional direct comparisons between children and adults with comparable levels of knowledge about dinosaurs are reported in Study 2.

A novel contribution of the current study was the finding that children's failure to solve triads based on deep feature relations was not due to a simple lack of knowledge. In many cases, children possessed the deep feature knowledge on which a potential solution could be based but still chose to base their decisions on surface feature similarities. Similar dissociations between knowledge and performance on a card-sorting task have been attributed to the gradedness of knowledge representations (Munakata & Yerys, 2001). It might be that relatively weak representations of deep features can readily support children's answers to questions regarding whether or not a particular deep feature is associated with a specific dinosaur. However, a stronger representation is required for the generation of similarity decisions based on deep features, particularly when competing perceptual bases for similarity judgments are available. Children with high levels of knowledge were significantly more likely to mention conceptual features when justifying both dinosaur and fish triad solutions than were children with low knowledge. In the General Discussion section later, we consider the question of how children's less developed framework theories related to biology might have contributed to this pattern of results.

Study 1 was conducted to examine the role of specific theories in making and justifying decisions of similarity. In this context, a broad array of both surface (perceptual) and deep (conceptual) features are available, and we were interested in the degree to which specific theoretical knowledge would affect the salience of particular features associated with familiar and less familiar categories. Study 2 addresses the question of how specific theories guide the generation of inferences concerning causal relations between surface features and deep features.

Study 2: Effects of knowledge and development on inferences concerning feature relations

An important contribution of the concepts-in-theories view has been the shift in focus from individuals' knowledge of features associated with concepts to individuals' understanding of the relations among those features. Understanding of feature relations develops as individuals acquire specific theories related to concepts throughout the life span. However, some young children develop elaborate networks of feature knowledge as a function of intense interest in a specific domain. In Study 2, children and adults with low, moderate, and high levels of knowledge about dinosaurs made inferences concerning the behaviors associated with physical features of three types of stimuli: actual dinosaurs, hypothetical dinosaurs created by combining features from pictures of actual dinosaurs, and unfamiliar shorebirds. Comparisons across children and adults with comparable levels of specific dinosaur knowledge were made to determine the relative influences of specific theoretical knowledge and broader framework theories pertaining to biology (e.g., Barrett et al., 1993; Wellman, 1990) on inference generation. If children and adults exhibit similar patterns of performance when they possess similar levels of specific knowledge, this provides support for the view that conceptual development is driven primarily by knowledge acquisition (Gelman, 1996; Keil & Batterman, 1984). If children and adults with comparable levels of specific knowledge perform differently on inference generation tasks, we could potentially infer that such differences are attributable to developmental differences in the broader theories of biology within which dinosaur concepts are situated.

Method

Participants

Participants were 46 children (mean age = 7 years 3 months, range: 5 years 9 months to 9 years 11 months) and 27 adults (mean age = 31 years, range: 17–73 years) who expressed high levels of interest in dinosaurs. Participants were recruited in Indianapolis, Indiana, through notices in the local newspaper and through a local museum exhibit on dinosaurs. Additional adult participants were recruited through the geology department at Indiana University–Purdue University at Indianapolis. Of the 27 adult participants, 4 were parents of child participants. Adult novices were undergraduates who had registered for, but not yet taken, a university course on dinosaurs. Children received a small gift and adults were paid for their participation.

Materials

Materials consisted of realistic color illustrations of nine dinosaurs, nine shorebirds, and four hypothetical dinosaurs that were digitally created by blending together features from two individual species. Two additional pictures (a kangaroo and a mug) were used during practice trials in the functional correlate task. All pictures were digitized and edited so that all exemplars were portrayed to scale and in the same canonical orientation against a white background. All exemplars were presented individually on a computer monitor. The experimenter controlled the rate of picture presentation through mouse clicks.

Procedure

Participants completed three tasks separated by an unrelated intervening activity. First, to assess participants' relative level of dinosaur knowledge, participants were also asked to name a set of 15 dinosaur pictures in the laboratory, with pictures presented individually on a computer monitor.[4] While each child was tested, a parent (typically the mother) was asked to complete a questionnaire that assessed his or

---

[4] The brief test of dinosaur knowledge (Gobbo & Chi, 1986) was not administered to participants in Study 2.

her own and the child's interest in and knowledge of dinosaurs and birds. Adult participants responded to the same questionnaire items prior to their completion of the picture-naming task. The content of this questionnaire was the same as that described for Study 1. However, ratings of knowledge and interest were made along an 8-point scale (1 = *knows nothing*, 3 = *average level of knowledge*, 5 = *a good deal more knowledge than average*, 8 = *knows just about all there is to know*) rather than the 9-point scale reported for Study 1. Adult participants were instructed to rate their own levels of knowledge and interest related to dinosaurs and birds and to indicate which of the 15 dinosaurs on the checklist they thought they could identify correctly. Following the knowledge assessment, half of the participants completed the functional correlate task first and half completed the inferential reasoning task first. All responses were audiotaped. At the end of the session, the PPVT-R was administered to assess verbal intelligence.

Inferences concerning functional correlates. Participants were presented with exemplars possessing a salient physical attribute and were asked to infer the function or behavior with which the feature was associated. Two practice trials were first presented using pictures of a kangaroo and a coffee mug. The experimenter indicated that she would be pointing out "something special that each thing has" and that participants would be asked to "tell me what that part is for." Participants were then asked to specify what the kangaroo's pouch and the mug's handle were for. All participants specified reasonable functional correlates for the two attributes that were provided. During the test trials, the realistic color pictures of nine dinosaurs and nine shorebirds were presented. Two orders of pictures within each domain were created, and half of the participants in each age group received each of the two orders. Presentation of pictures was blocked by domain, and the order in which the domains were presented also was counterbalanced across participants. In reference to each picture, participants were told, "This dinosaur [or shorebird] has X [where X was the target feature]. Can you tell me why?" Participants were encouraged to try to infer functions even if the animal possessing the feature was unfamiliar. After participants had responded, they were asked to name the pictured animal if they were certain they knew what it was. The list of dinosaurs, shorebirds, and their respective surface features and deep feature (i.e., functional) correlates are listed in the appendix.

Inferences concerning hypothetical dinosaurs. Four hypothetical dinosaurs were digitally created to depict either plausible or implausible feature combinations. Two adult experts who did not participate in this study assisted with determining which combinations of features were plausible or implausible. Implausible combinations involved surface features that were extremely unlikely to have coevolved given paleontological evidence related to their respective functions. For example, one of the implausible combinations involved a sauropod with a fin on its back. Fins typically are associated with thermoregulatory functions and generally are found on smaller dinosaurs that needed to move about quickly so as to capture prey. Sauropods were extremely large herbivores that were able to maintain their body temperatures largely through their enormous size; thus, they would not need to have evolved an additional thermoregulatory mechanism. A list of the created dinosaurs is presented in the appendix. Participants were instructed to imagine that each picture depicted a recently discovered dinosaur and were told, "Sometimes when scientists decide what new dinosaurs must have looked like, they do a good job, and sometimes they make mistakes. I'm going to show you some pictures of some brand new dinosaurs and ask you questions about them. One thing I'd like you to think about is whether the scientists did a good job figuring out what each dinosaur looked like or whether they probably made a mistake." Two random orders of the dinosaurs were created, and half of the participants received each of the two orders. The six questions asked about each of the four dinosaurs are presented in Table 5 in the order in which they were asked.

The audiotaped recording of each participant's responses across the two tasks was transcribed. One of the authors then checked each transcript against the original audiotape to ensure accuracy and to insert annotations for pauses in participants' responses and other nonverbal information (e.g., laughter).

Results

Assignments to knowledge groups

Knowledge group assignments were made based on a composite knowledge index score similar to that described for Study 1. Standard scores were derived within each age group for each of three variables: (a) parents' (or self-) ratings of relative level of dinosaur knowledge, (b) number of correct names produced at either the species or family level for the set of 15 dinosaurs presented in the laboratory, and (c) number of dinosaur names that parents (or adult participants) reported were known (out of 15 names listed on the questionnaire). The three standard scores were added together, and natural breaks in the distribution of the summed knowledge index scores were again used to divide same-age participants into low-, moderate-, and high-knowledge groups. Interrelations among the individual knowledge measures, PPVT-R scores, and overall knowledge index are presented for the child and adult groups near the bottom of Table 1. Again, the three knowledge measures were highly intercorrelated within each age group. In contrast to Study 1, PPVT-R scores were significantly related to performance across all knowledge measures. This may have been attributable to the

Table 5
Questions asked during inferential reasoning task

Dimension	Question
Diet	What do you think this dinosaur would have eaten?
Habitat	Where do you think this dinosaur would have lived?
Defense	How do you think this dinosaur would have defended itself?
Taxonomic relations	Which dinosaurs do you think would have been related to this dinosaur?
Functional correlate	What was the X on this dinosaur for? (X = salient feature presented in digitized drawing)
Plausibility	Did the scientists do a good job, or did they make a mistake, when figuring out what this dinosaur looked like?

greater number of participants recruited for Study 2 as well as the greater range of PPVT-R scores generated by Study 2 participants.[5]

Within each age group, simple one-way ANOVAs were conducted to test whether the three knowledge groups differed in terms of age, reported level of bird knowledge, and/or PPVT-R standard scores. Descriptions of the participants assigned to each of the three knowledge conditions within each age group are presented near the bottom of Table 2. Among children, there was a main effect of age, $F(2, 43) = 3.43$, $p < .05$. Bonferroni post hoc comparisons indicated that high-knowledge children were significantly younger than low-knowledge children. There were no differences across same-age groups in reported levels of bird knowledge. However, there were significant differences in PPVT-R scores across the three knowledge groups for both children, $F(2, 43) = 7.69$, $p < .01$, and adults, $F(2, 24) = 10.96$, $p < .001$. Post hoc tests indicated that low-knowledge children scored significantly lower on the PPVT-R than did either moderate- or high-knowledge children. Differences in verbal intelligence among the moderate- and high-knowledge groups were not significant. In the adult cohort, high-knowledge adults scored significantly higher on the PPVT-R than did both moderate- and low-knowledge adults, and the latter two groups did not differ from each other. Due to the confounding of level of knowledge with level of verbal intelligence and the presence of significant correlations between PPVT-R scores and task performance reviewed later, we used analyses of covariance in all subsequent comparisons across knowledge groups so as to control statistically for initial differences between knowledge groups in intelligence. Finally, independent groups t tests compared the child and adult groups in terms of the individual knowledge measures and the overall composite index. The only difference between children and adults pertained to parental and self-ratings of dinosaur knowledge, $t(71) = 2.43$, $p < .05$, attributable to parents' ratings of children's knowledge tending to be higher than adults' self-ratings of their own knowledge. All other knowledge measures were comparable across the two groups.

Responses to the naming task at the end of the functional correlate task verified that the shorebird exemplars were unfamiliar to all participants and that individuals across the three knowledge groups differed in their degrees of familiarity with the dinosaur exemplars. No participants identified any of the shorebird exemplars with the correct species names. Adults in the high-, moderate-, and low-knowledge groups correctly identified 60, 41, and 10% of the dinosaurs, respectively, whereas children in the high-, moderate-, and low-knowledge groups correctly identified 60, 38, and 28% of the dinosaurs, respectively, either by their species names or by family nicknames (e.g., "duckbill" for the anatosaurus).

Task 1: Inferences concerning deep feature correlates

Participants' responses concerning the deep feature correlates of specified surface features were transcribed and coded in terms of both overall quality (correct vs incorrect) and whether responses were qualified or not. Correctness of deep feature

---

[5] For a more detailed consideration of the relation between expertise on domains of object concepts and intelligence, see Johnson and Eilers (1998).

correlates was determined through consultation with two adult experts on dinosaurs and two adult experts on birds. (None of these adults participated in the current study.) Qualified responses were defined as those in which participants' answers were accompanied by phrases such as "it might be because ___," "maybe so that it ___," and "I don't know, but it could be ___" or instances in which at least 10 s elapsed between the experimenter's question and the participant's response (suggesting that the participant did not immediately generate an inference or was waffling between at least two alternative responses). All responses were coded by one individual, and then one of the authors recoded 30% of the transcripts from each age group. Coding for this analysis and subsequent analyses was conducted prior to determinations of participants' relative knowledge levels to ensure that coders were blind to knowledge group membership. Reliability (agreements/agreements + disagreements) was .93, and disagreements were resolved through discussions between the two coders.

We first were interested in the degree to which generation of correct deep feature correlates varied within each age group as a function of knowledge level and across children and adults with comparable levels of dinosaur knowledge. We also were interested in the frequency with which children and adults with similar levels of knowledge would make correct inferences concerning the deep feature correlates of less familiar shorebird features. Preliminary analyses indicated that there were no effects involving the order in which the two tasks were presented or the order in which domains were presented, so these factors were collapsed across in subsequent analyses. Participants' correct responses generated for both dinosaur and shorebird exemplars were first compared in a 2 (Age) × 3 (Knowledge: high, moderate, or low) × 2 (Exemplar Type: dinosaurs or shorebirds) mixed analysis of covariance (ANCOVA), with age and knowledge as between-groups factors and exemplar type as the within-group factor. The means for this analysis, adjusted for the effects of PPVT-R scores, are presented in Fig. 1. There were significant main effects for both age, $F(1, 66) = 32.79$, $p < .001$, and knowledge, $F(2, 66) = 19.91$, $p < .001$, as well as a significant interaction between age and exemplar type, $F(1, 66) = 16.17$, $p < .001$. Collapsing across both domains, adults generated more correct responses than did children, and across both age groups, participants generated more correct responses as their level of knowledge increased. However, the Age × Exemplar Type interaction suggests that the two age cohorts diverged in terms of their generation of correct inferences for shorebird exemplars.

The Age × Exemplar Type interaction was examined through separate 2 (Age) × 3 (Knowledge: high, moderate, or low) ANCOVAs conducted on participants' responses for each of the two types of exemplars. For responses generated in reference to dinosaurs, only the main effect of knowledge was significant, $F(2, 66) = 27.81$, $p < .001$. Adults and children with comparable levels of knowledge performed very similarly when making inferences concerning causal relations between surface features and deep features. For responses generated in reference to shorebird exemplars, both the main effects of age, $F(1, 66) = 41.03$, $p < .0001$, and those of knowledge, $F(2, 66) = 3.36$, $p < .05$, were significant. Although low-knowledge children and adults produced comparable numbers of correct responses for shorebirds, both moderate- and high-knowledge adults tended to produce more correct responses for shorebirds than did children with comparable levels of knowledge. These results

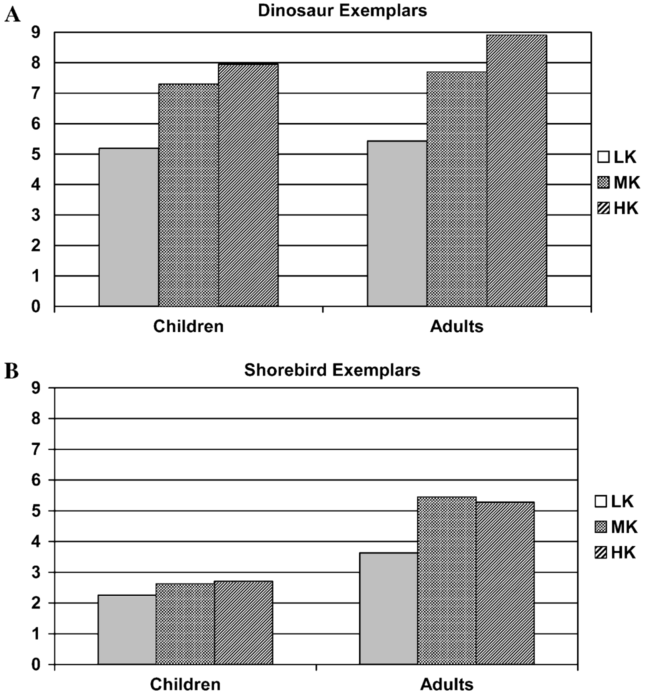

Fig. 1. Deep feature correlate task: mean numbers of correct inferences (out of 9) produced. HK, high knowledge; MK, moderate knowledge; LK, low knowledge.

suggest that although specific theories related to attribute correlations within the familiar domain become more accurate for both children and adults as knowledge accrues, only adults with at least a moderate level of knowledge about dinosaurs are able to generate biologically accurate inferences in reference to exemplars from the unfamiliar domain. This effect may have arisen because adults were more apt to analogically transfer their teleological inferences pertaining to physical features of dinosaurs to a less familiar (but still biological) domain. It also may have been due to adults' increased exposure to factual information about birds (and other animals) more generally. However, the fact that low-knowledge adults and low-knowledge children did not significantly differ in terms of their generation of inferences pertaining to shorebirds suggests that the result was not simply due to adults possessing greater knowledge about causal relations among general biological features.

We also were interested in the extent to which responses were qualified to indicate uncertainty across children and adults with comparable levels of dinosaur knowledge. Indications of uncertainty could support the premise that responses are generated based on inductive inference rather than on simple fact retrieval. The numbers of qualified responses were compared across exemplars in a parallel 2 (Age) × 3 (Knowledge: high, medium, or low) × 2 (Exemplar Type: dinosaurs or shorebirds) mixed ANCOVA, with age and knowledge as between-groups factors and exemplar type as the within-group factor. Only main effects of age, $F(1, 66) = 50.39$, $p < .001$, and knowledge, $F(2, 66) = 6.21$, $p < .01$, emerged. Adults were significantly more likely to express uncertainty concerning their responses than were children. Planned pairwise comparisons indicated significant differences among all combinations of knowledge groups; low-knowledge individuals ($M = 5.21$) qualified their responses significantly more often than did medium-knowledge individuals ($M = 3.99$), who in turn qualified their responses significantly more often than did high-knowledge individuals ($M = 2.81$) (all *p*s $< .05$).

Task 2: Inferences concerning hypothetical dinosaurs

Participants were asked during the inferential reasoning task to make decisions concerning the plausibility of hypothetical ("newly discovered") dinosaurs and to infer behavioral characteristics (e.g., diet, habitat, taxonomic relations, mode of defense) based on the dinosaurs' appearance. The appropriateness of participants' responses to these questions was determined based on published dinosaur encyclopedias and through consultation with a geology instructor who regularly taught college-level courses related to dinosaurs. Responses to each question were coded separately by a single coder, and again one of the authors coded 30% of the transcripts. Agreement between the two coders (agreements/agreements + disagreements) was .91, and disagreements were resolved through discussion.

We first were interested in the degree to which participants would consider hypothetical dinosaurs to be plausible or implausible as well as the justifications that they provided for their decisions. We began by considering the number of instances for which participants correctly identified a plausible dinosaur to be hypothetically possible (i.e., cases where participants decided that the paleontologist did a "good job" in deciding what the dinosaur might have looked like). The numbers of such cases were compared across participants in a 2 (Age) × 3 (Knowledge: high, moderate, or low) ANCOVA, which revealed only a significant main effect of age, $F(1, 66) = 17.42$, $p < .001$. The main effect of knowledge and the Age × Knowledge interaction only approached significance ($p < .11$), presumably due to sample size limitations. The means for this analysis, adjusted for the effects of the PPVT-R covariate, are depicted at the top of Fig. 2, with asterisks denoting responses that differed from chance (as assessed by separate one-sample *t* tests). Children were more apt to reject such dinosaurs as plausible than were adults. When queried further, many children indicated that the paleontologist had to be mistaken because they had never encountered any dinosaur that looked like that before. The trend toward the Age × Knowledge interaction was driven primarily by individuals in the high-knowledge group. Children in the high-knowledge group seemed to have particular difficulty in entertaining the notion that such dinosaurs may have existed, even though they had never seen them before. Although it is possible that children simply understood the task less well than did adults or that children had more of a tendency to respond negatively, it also could suggest that high levels of specific theoretical knowledge do not alter children's tendency to engage in concrete operational reasoning.

The parallel analysis performed on the number of implausible dinosaurs that were correctly rejected by participants as "mistakes" yielded a different pattern of results; only a main effect of knowledge group emerged, $F(2, 66) = 3.04$, $p = .05$. Pairwise

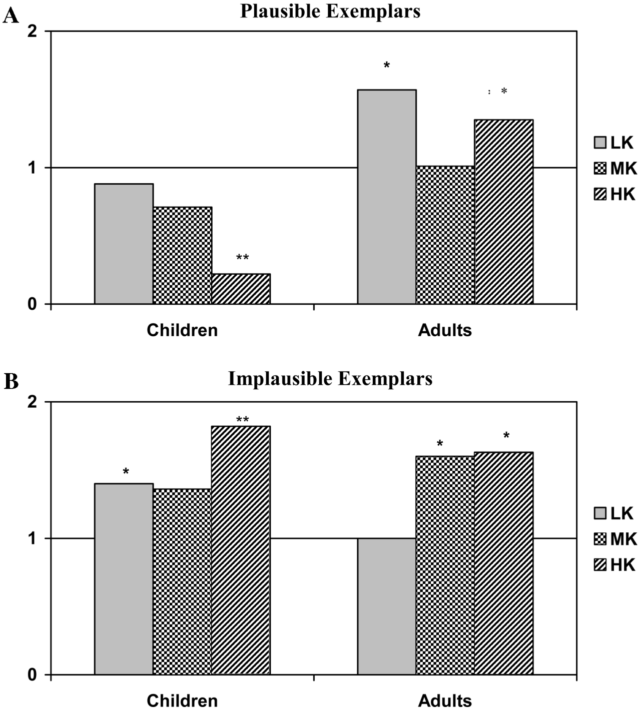

Fig. 2. Inferential reasoning task: mean numbers of correct "accepts" of plausible exemplars and correct "rejects" of implausible exemplars. HK, high knowledge; MK, moderate knowledge; LK, low knowledge. *Different from chance (1) at $p < .05$; **different from chance (1) at $p < .001$ (as assessed through one-sample *t* tests).

comparisons indicated that high-knowledge individuals were significantly more apt to reject implausible dinosaurs than were low-knowledge individuals; no other differences among knowledge groups were significant. Parallel one-sample *t* tests were run to compare each group's responses with chance levels, as depicted at the bottom of Fig. 2. However, participants' justifications of their responses differed markedly as a function of age, with most children again indicating that the scientists must have made a mistake because "I've never seen a dinosaur like that before" (83%), most adults indicating either that the animal would not have evolved to look like that (e.g., "If it had all that armor, it couldn't run quickly") (55%), or that the animal just did not look like it could exist (e.g., "It just looks kind of busy with all that stuff on it") (30%). Thus, children were not generating the same types of inferences as were adults. Regardless of their level of specific theoretical knowledge, children indicated that exemplars that had never been encountered before were likely to be "mistakes."

Responses to the individual questions pertaining to the hypothetical dinosaurs generally revealed the same pattern of findings obtained for the deep feature correlate inferencing task. Numbers of appropriate responses made to the questions concerning diet, taxonomic relations, and functional attribute correlates yielded only significant main effects of knowledge (diet: $F(2, 66) = 7.26$, $p < .001$; taxonomic relations: $F(2, 66) = 8.20$, $p < .001$; functional attribute correlates: $F(2, 66) = 6.01$, $p < .01$). Linear trend analyses confirmed that within each age group, more appropriate inferences concerning these properties were generated as knowledge increased ($p < .01$). The question concerning habitat yielded only a main effect of age, $F(1, 66) = 15.06$, $p < .001$, primarily because children tended to misinterpret the question as referring to the time period in which the animal lived.

In sum, both children and adults in the moderate- and high-knowledge groups were significantly more accurate than their low-knowledge peers in inferring the deep features correlated with dinosaurs' surface features (for both real and hypothetical dinosaurs) and in deducing the likely diets and defensive behaviors of hypothetical dinosaurs. Developmental differences emerged in two principal areas. First, adults in the medium- and high-knowledge groups were significantly more likely than adults in the low-knowledge group to infer correct functional correlates for shorebirds. There was no parallel effect of knowledge on children's responses. High-knowledge children were significantly more likely than high-knowledge adults to indicate that a mistake had been made during the reconstruction of "plausible" hypothetical dinosaurs. High-knowledge children often reported that the scientists must have made a mistake because "I've never seen a dinosaur like that before." High-knowledge adults frequently reported that such dinosaurs were plausible because "they might have evolved to fill a particular ecological niche." We speculate that developmental differences in framework theories of biology could have accounted for differences between children and adults with comparable levels of specific theoretical dinosaur knowledge. An alternative possibility is that children actually possessed some semblance of a framework theory of biology (Simons & Keil, 1995) yet lacked the skills necessary to draw connections between dinosaurs and shorebirds. The possibility that child experts might be differentially susceptible to competence–performance discrepancies could be investigated further by using less demanding, comprehension-based inferencing measures.

General discussion

The original impetus for this research was to begin to determine what role theories play in influencing concept use by individuals with various levels of expertise on biological domains. Two facets of concept use were explored: making and justifying similarity decisions and generating inferences concerning the relations between deep features and surface features. Our findings replicated past work indicating that child experts attend to relatively subtle surface features that are causally related to biological adaptations (Chi et al., 1989; Gobbo & Chi, 1986; Johnson & Mervis, 1994). However, we also discovered the basis for an important qualification to this established pattern: Heightened knowledge is necessary, but not sufficient, for sophisticated types of reasoning about concepts. The ability to access explicit knowledge of deep features did not automatically lead to the use of that knowledge as the basis for similarity decisions, suggesting that such knowledge is graded. Munakata and Yerys (2001) suggested that when dissociations between knowledge and action arise, they reflect relatively weak representations that may be sufficient for some tasks (e.g., responses to direct questions about deep features) but not for others (e.g., triad solutions). This finding also is reminiscent of production deficiencies reported in young children's strategic memory performance (Schneider & Bjorklund, 1998). Even when children were highly knowledgeable about the bases for surface feature–deep feature correlations, they refused to acknowledge that unfamiliar instances were hypothetically plausible. Finally, children demonstrated relatively little generalization of their specific conceptual knowledge to related, but less familiar, biological domains. Children's expertise on object concepts appears to be more empirically constrained and more isolated from other systems of related biological knowledge than does adults' expertise. We begin by considering the effects of expertise on conceptual structure. We then address the transition from children's "folkbiological" knowledge to adults' folkbiological knowledge and consider how high levels of specific conceptual knowledge might affect the course of this evolution.

Expertise and conceptual structure

Becoming an expert on conceptual object domains such as birds, trees, and airplanes entails learning new concepts and associated features. Studies of adults have reliably yielded differences in the likelihood of use of surface and deep features during categorization as a function of level of expertise (e.g., Anzai, 1991; Chi et al., 1981; Hardiman, Dufresne, & Mestre, 1989; Johnson & Mervis, 1998; Medin, Lynch, Coley, & Atran, 1997; Patel & Groen, 1991). Gobbo and Chi (1986) found similar results with children, reporting that children who were experts on dinosaurs were more likely to mention deep features when discussing dinosaurs. Furthermore, Chi and colleagues (1989) found that child experts were more apt to justify dinosaur groupings based on diet than were child novices. Johnson and Mervis (1994) found that for one 4-year-old expert on birds, deep features consistently were used when making and justifying similarity decisions, even when deep features were pitted directly against surface features in a triad judgment task.

Expertise acquisition also entails a shift in the kinds of surface features that are noticed. Subtle features that differentiate among coordinate categories become more salient through experience (Biederman & Shiffrar, 1987; Gibson & Gibson, 1955; Johnson, 2002; Johnson & Mervis, 1998; Proctor & Dutta, 1995). Gibson's (1969, 1991; see also Gibson & Levin, 1975) research on perceptual learning illustrates that sheer exposure to stimuli is sufficient for expert detection of both differentiating and higher order features of which novices are unaware. Keil (1994a) pointed out that expertise in perceptually based skill domains such as chick sexing is "blind" in that it depends on mechanisms of association and the automatization of highly repetitive routines rather than on theoretically based explanations. Dinosaur knowledge resides at the interface of perceptual learning and folkbiological theories in that its acquisition entails both perceptual differentiation and the elaboration of intuitive theories. As in previous research involving bird experts, we argue that perceptual learning is necessary, but not sufficient, for attaining high levels of competence. Also necessary is experts' understanding of why features tend to co-occur (i.e., the notion of teleology) (Keil, 1994a) and of the conceptual bases for abstract concepts that share invariant features (Johnson & Mervis, 1998).

Even very young children go beyond perceptual information when making decisions about category membership and when making inductive inferences (Barrett et al., 1993; Gelman & Markman, 1986; Keil, 1989). Perceptual information is interpreted in light of theoretical knowledge that children possess about biology, kinship, and the essence of things. Theories likely constrain the kinds of inferences that children draw about surface and the deep features with which they are associated. Theories also may allow analogical predictions to be drawn concerning unfamiliar, but related, exemplars (Inagaki, 1990). Given that individuals with high levels of domain-specific knowledge possess theories that are maximally rich and cohesive, experts should infer different kinds of feature relations than should novices. Adults also possess different intuitive theories than do children (Carey, 1985), but it is unclear to what extent theories change as a function of development versus fact acquisition. The relative lack of child experts makes this a particularly difficult question to address empirically. Future efforts to disentangle these relations by studying children from rural or indigenous cultures (Coley, 2000) or by seeking out child experts on specific types of biological concepts will provide valuable contributions to the literature.

Children's and adults' theories of biology

The question of how individuals' understanding of biological concepts changes throughout the life span has generated a wealth of research since the publication of Carey's (1985) seminal book. Coley (2000) characterized the field as divided into two major camps on the issue of folkbiological knowledge acquisition. First, the radical theory change view stipulates that the adult-state version of biological knowledge is radically different from that of children. Developmental differences are on the order of framework theory differences, and it is unlikely that even passionately interested child experts could achieve precocious knowledge restructuring. Alternatively, the knowledge enrichment view maintains that children's conceptions of biological kinds are in many ways similar to those of adults. Development entails a much more continuous process of fleshing out this initially skeletal structure with more concrete conceptual information (Simons & Keil, 1995). This view is far more compatible with the idea that child experts can outperform adult novices at some level. Our findings suggest that a more temperate position, situated between these two extremes, is warranted. Although many aspects of children's and adults' performance on categorization tasks are similar when children's knowledge is dense (supporting the knowledge enrichment view), systematic developmental differences continued to obtain. In particular, children are more susceptible to competence–performance discrepancies than are adults, and children's reasoning is more empirically constrained than that of adults.

Fact acquisition enriches specific theories by creating rich networks of associated features and heightened understanding of the bases for their relations. When children and adults are moderately familiar with concept exemplars, categorical inferences are highly similar. Specific theories related to dinosaurs enabled more knowledgeable children to attend to specific surface features and to infer biological functions that were correlated with those features. Nevertheless, specific theories did not influence all aspects of children's categorization performance. Striking perceptual similarities continued to influence children's similarity decisions throughout the continuum of expertise, regardless of how much knowledge about the domain children possessed. Children's reasoning about feature relations was more empirically constrained than that of adults. High-knowledge children tended to reject hypothetical dinosaurs that had not been encountered before, regardless of whether or not these dinosaurs' features might plausibly co-occur. This suggests that specific theories cannot override children's tendency to reason concretely about novel category exemplars. The empirical constraints on high-knowledge children's inference generation is at odds with evidence that children as young as 6 years can use covariation evidence to formulate simple hypotheses (Ruffmann, Perner, Olson, & Doherty, 1993; Sodian, Zaitchik, & Carey, 1991). One possibility is that young children's early competence in scientific reasoning may obtain only when domain knowledge is relatively low. An important question for future research concerns whether young children might be less willing to formulate hypotheses based on feature covariations when those covariations conflict with the children's expert knowledge of feature relations aligned with the familiar domain.

One of the most flexible aspects of the human cognitive system is the fact that humans can, under some conditions, generalize their knowledge and skills to new domains. Across a variety of categorization tasks, high knowledge yields a performance benefit only to adults (Johnson & Eilers, 1998; Johnson & Mervis, 1998). For children, specific theories do not always support generalization of solution strategies to similarly structured, but less familiar, domains (for an exception to this pattern, see Inagaki, 1990). We suggest that framework theories of biology, within which specific theories pertaining to dinosaurs are embedded, may help to support such generalizations. We suspect that the application of expert knowledge occurs through analogical mapping of specific feature relations from familiar to less familiar concepts included within the same framework theory. It is likely that such generalization effects obtain for expertise on biological object concepts but not on skill domains such as chess and backgammon (Chi et al., 1988) because the latter types of domains do not share overarching theoretical commonalities.

Acknowledgments

This research was supported by National Institute of Child Health and Development (NICHD) Grant HD27042 and by National Institute of Mental Health (NIMH) Grant MH53134. We thank Dick Neisser, Howard Rollins, and Byron Robinson for their helpful discussions of this work. Cindy Mervis assisted with software development, and Jodie Edwards, Amy Eilers, and Joann Helferich assisted with the collection, transcription, and coding of data. We are especially grateful to the dinosaur enthusiasts who participated in this research and to Suzanne Vizzie and Jeff Swanagan for helping us to recruit them.

Appendix

Triad stimuli: Dinosaurs

Target	Deep feature basis	Deep feature match	Surface feature match
Deinonychus	True dinosaur	Styracosaurus	Dimorphodon
Coelophysis	True dinosaur	Triceratops	Nothosaur
Parasaurolophus	True dinosaur	Compsognathus	Pteranodon
Spinosaurus	True dinosaur	Diplodocus	Dimetrodon
Dromaeosaurus	Carnivorous	Daspletosaurus	Hypsilophodon
Anatosaurus	Herbivorous	Stegosaurus	Tyrannosaurus
Allosaurus	Carnivorous	Compsognathus	Iguanodon
Ouranosaurus	Herbivorous	Opisthocoelicaudia	Acrocanthosaurus
Tenontosaurus	Ornithiscian	Torosaurus	Camarasaurus
Edmontosaurus	Hadrosaur	Corythosaurus	Plateosaurus
Struthiomimus	Armored dinosaur	Scediolosaurus	Kentrosaurus
Polacanthus	Armored dinosaur	Nodosaurus	Lexovisaurus

Triad stimuli: Fish

Target	Deep feature basis	Deep feature match	Surface feature match
Mudskipper	True fish	Lemon cichlid	Mole salamander
Fatheaded catfish	True fish	Betta	Three-toed amphiuma
Mottled spiney eel	True fish	Bigeye cachorro	False pit viper
Flagfin mudskipper	True fish	Earth eater	Salamander
Goldtone barb	Eat dry food	Rainbow shark	False alestes
Silvery falying barb	Eat worms	Rosy barb	Blacklined leporinus
Dwarf distichodus	Eat plants	Redhook mytinnis	False gold tetra
Archer fish	Eat worms	Roundspotted puffer	Mimic perissodus
Blacklined tetra	Lay eggs	Mpozo cichlid	Darter goodied
Bengal loach	Feeding depth	Spotted elephant	Eight-banded loach
Merry widow	Live birth	Caudo	Curimata
Bartailed rivulus	Feeding depth	Bitterling	Marbled hatchetfish

Functional correlate task stimuli: Dinosaurs

Exemplar	Physical feature	Functional correlate
Mamenchisaurus	Long neck	Eat leaves from tree tops
Spinosaurus	Fin on its back	Thermoregulation, display
Struthiomimus	Stiff tail	Balance while running
Anatosaurus	Teeth like grinding files	Grind plants, bark, pine needles
Deinonychus	Sickle-shaped claw	Disembowel prey
Pachycephalosaurus	Rounded skull	Head-ramming
Ankylosaurus	Clubbed tail	Defense
Tyrannosaurus Rex	Very sharp teeth	Kill prey
Triceratops	Bony frill	Protection, counterweight

Functional correlate task stimuli: Shorebirds

Exemplar	Physical feature	Functional correlate
Avocet	Long upcurved bill	Sweep through mud while hunting
Coot	Lobed toes	Swimming
Gallinule	Long toes	Support weight on water plants
Merganser	Toothed bill	Grip fish
Oystercatcher	Thick chiseled bill	Crack shells open
Skimmer	Bottom part of bill longer than top part	Scoop fish on the wing
Spoonbill	Spoon-shaped bill	Scoop and filter prey
Stilt	Very long legs	Wade in deep water while hunting
Woodcock	Very large eyes	Hunt at night

Inferential reasoning task stimuli

Stimulus type	Exemplar	Basis for correct inference
Implausible	Armored carnosaur	Predator would not be hunted; would not need armor
	Finned sauropod	Large sauropod would not need thermoregulatory mechanism (large size would maintain temperature)
Plausible	Crested sauropod	Crest may have evolved for display, communication
	Plated ceratopsian	Plates may have evolved for display, protection

References

Alexander, J. M., Johnson, K. E., & Schreiber, J. B. (2002). Knowledge is not everything: Analysis of children's performance on a haptic comparison task. *Journal of Experimental Child Psychology, 82*, 341–366.

Anzai, Y. (1991). Learning and use of representations for physics expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 64–92). New York: Cambridge University Press.

Axelrod, H. R., Burgess, W. E., Emmens, C. W., Pronek, N., Walls, J. G., & Hunziker, R. (1987). *Dr. Axelrod's mini-atlas of freshwater aquarium fishes: Mini-edition*. Neptune City, NJ: THF Publications.

Barrett, S. E., Abdi, H., Murphy, G. L., & Gallagher, J. M. (1993). Theory-based correlations and their role in children's concepts. *Child Development, 64*, 1595–1616.

Barrett, S. E., Abdi, H., & Sniffen, J. (1992). Reflecting on representation and process: Children's understanding of cognition. In B. Burns (Ed.), *Percepts, concepts, and categories: The representation and processing of information* (pp. 275–322). Amsterdam: Elsevier.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 640–645.

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

Carey, S. (1995). On the origin of causal understanding. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 268–302). Oxford, UK: Clarendon.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Lawrence Erlbaum.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Chi, M. T. H., Hutchinson, J. E., & Robin, A. F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill–Palmer Quarterly, 35*, 27–62.

Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19*, 29–39.

Clark, E. (1983). Meanings and concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology, Vol. 3. Cognitive development* (4th ed., pp. 787–840). New York: John Wiley.

Coley, J. D. (2000). On the importance of comparative research: The case of folkbiology. *Child Development, 71*, 82–90.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test—revised*. Circle Pines, MN: American Guidance Service.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273–305.

Gelman, S. A. (1996). Concepts and theories. In R. Gelman & T. K. Au (Eds.), *Perceptual and cognitive development* (pp. 117–150). San Diego: Academic Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–208.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton–Century–Crofts.

Gibson, E. J. (1991). *An odyssey in learning and perception*. Cambridge, MA: MIT Press.

Gibson, E. J., & Gibson, J. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*, 32–41.

Gibson, E. J., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.

Glaser, R. (1987). Thoughts on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 81–94). Norwood, NJ: Ablex.

Gobbo, C., & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development, 1*, 221–237.

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition, 17*, 627–638.

Hirschfeld, L. A., & Gelman, S. A. (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Inagaki, K. (1990). The effects of raising animals on children's biological knowledge. *British Journal of Developmental Psychology, 8*, 119–129.

Inagaki, K. (1997). Emerging distinctions between naive biology and naive psychology. In H. M. Wellman, K. Inagaki (Eds.), *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena* (New Directions for Child Development, No. 75, pp. 27–44) San Francisco: Jossey–Bass.

Johnson, K. E. (2002). Determinants of typicality throughout the continuum of expertise. *Memory & Cognition, 29*, 1036–1050.

Johnson, K. E., & Eilers, A. T. (1998). Effects of knowledge and development on subordinate level categorization. *Cognitive Development, 13*, 515–545.

Johnson, K. E., & Mervis, C. B. (1994). Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. *Developmental Psychology, 30*, 418–435.

Johnson, K. E., & Mervis, C. B. (1998). Impact of intuitive theories on feature recruitment throughout the continuum of expertise. *Memory & Cognition, 26*, 382–401.

Keil, F. C. (1994a). The birth and nurturance of concepts. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). New York: Cambridge University Press.

Keil, F. C. (1994b). Explanation, association, and the acquisition of word meaning. In L. Gleitman & B. Landau (Eds.), *The acquisition of the lexicon* (pp. 169–196). Cambridge, MA: MIT Press.

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

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 234–262). Oxford, UK: Oxford University Press.

Keil, F. C., & Batterman, N. (1984). A characteristic to defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23*, 221–236.

Keil, F. C., & Lockhart, K. L. (1999). Explanatory understanding in conceptual development. In E. K. Scholnick, K. Nelson, S. A. Gelman, & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 103–130). Mahwah, NJ: Lawrence Erlbaum.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*, 1469–1481.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology, 32*, 49–96.

Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*, 335–337.

Murphy, G. L. (1993). Theories and concept formation. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 173–200). San Diego: Academic Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Norman, D. (1985). *The illustrated encyclopedia of dinosaurs*. New York: Crescent Books.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93–125). New York: Cambridge University Press.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Ruffmann, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development, 64*, 1617–1636.

Schneider, W., & Bjorklund, D. F. (1998). Memory. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology, Vol 2. Cognition, perception, and language* (pp. 467–521). New York: John Wiley.

Simons, D. J., & Keil, F. C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition, 56*, 129–163.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753–766.

Springer, K. (1999). How a naive theory of biology is acquired. In M. Siegal & C. C. Peterson (Eds.), *Children's understanding of biology and health* (pp. 45–70). New York: Cambridge University Press.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wilson, R. (1986). *100 dinosaurs from A to Z*. New York: Grosset & Dunlap.