Ghosts of Theories Past: The Ever-Present Influence of Long-Discarded Theories

Andrew Shtulman (shtulman@oxy.edu)
Joshua Valcarcel (jvalcarcel@oxy.edu)
Department of Psychology, Occidental College
1600 Campus Road, Los Angeles, CA 90041

Abstract

When individuals replace their naïve theories of natural phenomena with more accurate, scientific ones, what happens to the original theories? Are they overwritten or merely suppressed? We investigated this issue by asking college undergraduates to verify two types of statements as quickly as possible: statements whose truth value was the same across naïve and scientific theories of the phenomena at hand (e.g., "The moon revolves around the Earth") and statements whose truth value differed across those theories (e.g., "The sun revolves around the Earth"). Participants verified the former more quickly and more accurately than the latter, though this difference was smaller for early-developing domains (fractions, germs, inheritance, matter, physiology) than late-developing domains (astronomy, evolution, mechanics, thermodynamics, waves). These findings suggest that intuitive theories survive the acquisition of a mutually incompatible scientific theory, coexisting with, or even competing with, that theory for many years to come.

Keywords: intuitive theories, conceptual change, knowledge acquisition, science education

Introduction

Knowledge acquisition can take two very different forms. One form, known as "knowledge enrichment," involves the accretion of beliefs expressible in terms of preexisting concepts, like learning the traits of an unfamiliar animal or learning the history of an unfamiliar country. Another form, known as "conceptual change," involves revising the very concepts that articulate those beliefs, like learning that the earth is a sphere, not a plane, or learning that weight is a relational property of objects, not an intrinsic property. The result is a capacity to represent information incommensurate with our prior beliefs and experiences.

Numerous instances of conceptual change have been documented in the cognitive development and science education literatures. They include the transition from an impetus-based theory of motion to an inertial theory (Clement, 1982; McCloskey, 1983); the transition from a substance-based theory of energy to a process-based theory (Reiner, Slotta, Chi, & Resnick, 2000; Wiser & Amin, 2001); the transition from a geocentric model of the solar system to a heliocentric model (Siegal, Butterworth, & Newcombe, 2004; Vosniadou & Brewer, 1994); the transition from a tactile theory of matter to a particulate theory (Nakhleh, Samarapungavan, & Saglam, 2005; Smith, 2007); the transition from an integer-based understanding of fractions to a division-based understanding (Hartnett & Gelman, 1998; Moss & Case, 1999); the transition from a psychological theory of bodily functions to a vitalist theory (Johnson & Carey, 1998; Slaughter & Lyons, 2003); the transition from a behavioral theory of illness to a germ-based theory (Au et al., 2008; Solomon & Cassimatis, 1999); and the transition from an essentialist theory of evolution to a selection-based theory (Shtulman, 2006; Shtulman & Schulz, 2008).

Different scholars have characterized these transitions in different ways. Carey (2009) and Smith (2007) have characterized them as a series of conceptual differentiations, in which new conceptual boundaries are established, and conceptual coalescences, in which old conceptual boundaries are collapsed. Thagard (1992) and Chi (2008) have characterized them as the reassignment of a key concept or system of concepts from one branch of an ontological hierarchy to another. And Vosniadou (1994) and Wellman and Gelman (1992) have characterized them as a revision of the core presuppositions of a framework theory, or a theory that defines a domain's ontological categories and causal mechanisms. Common to all characterizations is a commitment to knowledge restructuring, or the conversion of one conceptual system into another by radically altering the structure (and not just the content) of that system.

Implicit in the idea of knowledge restructuring is the idea that early modes of thought, once restructured, should no longer be accessible, for the basic constituents of that system are no longer represented. A number of recent findings have challenged this idea, however, by showing that early modes of thought do sometimes reemerge later in life. Lombrozo, Kelemen, and Zaitchik (2007), for example, found that adults with Alzheimer's disease endorse teleological explanations for natural phenomena that young children also endorse but that age-matched adults without Alzheimer's disease do not – explanations like "trees exist so that plants and animals have shade" and "rain exists so that plants and animals have water to drink." Kelemen and Rosset (2009) found that, under speeded conditions, college undergraduates are also inclined to endorse the same kinds of explanations. The fact that individuals who explicitly disavow such explanations under normal circumstances will still endorse them under abnormal circumstances (i.e., cognitive load or biological impairment) implies that teleological thought is suppressed with additional education but never completely overwritten.

In a different line of experiments, Goldberg and Thompson-Schill (2009) found that adults are slower and less accurate at categorizing plants as living things than at categorizing animals as living things, despite knowing full well that both plants and animals are alive. While young children explicitly deny that plants are alive, they typically revise this belief by the age of eight or nine (Carey, 1985). Yet this belief apparently persists at an implicit level across decades of additional education and experience, for even Yale biology professors were found to be slower and less accurate at classifying plants as alive than at classifying animals as alive. Years of professional experience had not apparently erased an erroneous distinction these biologists had once drawn as children.

These findings, among others (e.g., Legare & Gelman, 2008; Shafto, Coley, & Baldwin, 2007), suggest that scientific knowledge serves to mask, rather than replace, one's initial, intuitive conceptions of the relevant domain. A more conservative interpretation, however, is that teleology and animism are particularly resilient modes of thought but that nothing comparable occurs in other domains of knowledge. We decided to explore this issue by adopting Goldberg and Thompson-Schill's (2009) methods but by expanding the scope of inquiry beyond the living-nonliving distinction to a variety of other conceptual distinctions in a variety of other domains: astronomy, evolution, fractions, germs, inheritance, matter, mechanics, physiology, thermodynamics, and waves. For each domain, we measured the speed and accuracy with which well educated adults verified two types of domain-specific statements: statements whose truth-value is known to remain constant across a conceptual change within that domain (e.g., "The moon revolves around the Earth," which is true on both naïve and scientific theories of astronomical phenomena) and statements whose truth-value is known to reverse across conceptual change (e.g., "The sun revolves around the Earth," which is true on naïve theories of astronomical phenomena but is not true on a scientific theory).

We hypothesized that, if naïve theories survive the acquisition of a mutually incompatible scientific theory, then statements whose truth-value is known to reverse across conceptual change should cause greater cognitive conflict than statements whose truth-value is known to remain constant across that change, resulting in slower and less accurate verifications of the former as compared to the latter. If, on the other hand, naïve theories are generally overwritten by scientific theories, then statements whose truth-value reverse across conceptual change should cause no greater cognitive conflict than statements whose truth-value remain constant, since the naïve theories would no longer be present to cause such conflict.

We also hypothesized that, if long-discarded theories do indeed cause cognitive conflict when reasoning about phenomena once covered by those theories, then participants should experience different levels of conflict for theories discarded at different points in time, with theories discarded early in life yielding the least conflict and theories discarded late in life yielding the most. To test this hypothesis, we included both domains that tend to undergo conceptual reorganization during the first decade of life, like fractions and physiology, and domains that tend to undergo conceptual reorganization during the second decade of life (or later), like evolution and mechanics.

Method

Participants

The participants were 65 college undergraduates (19 men, 46 women) recruited from introductory psychology and cognitive science courses. Nine were majoring (or intending to major) in the natural sciences, 25 were majoring in the arts and humanities majors, and 31 were majoring in the social sciences. The number of college-level science and math courses that participants had taken prior to the study ranged from 0 to 9 and averaged 2.2.

Materials

Participants were presented 200 statements about natural phenomena and asked to decide whether each was true or false. The statements were designed to exemplify one of five conceptual relations specific to one of ten domains (four statements per relation, five relations per domain). Five of the domains entailed relations that typically undergo restructuring in the first decade of life (fractions, germs, inheritance, matter, physiology), and five entailed relations that typically undergo restructuring in the second decade of life, if at all (astronomy, evolution, mechanics, thermodynamics, waves). We refer to the former as "early developing domains" and the latter as "late developing domains." This classification, though inexact, was based on prior findings within the cognitive development and science education literatures (noted above). Our selection of domain-specific conceptual relations was based on these findings as well.

As an illustration, consider the five relations used to represent the domain of matter: "[material_1] is denser than [material_2]," "[material_1] weighs more than [material_2]," "[entity] is composed of matter," "[entity] can be cut in half," and "atoms have [property]." Each of these relations was represented by four particular types of statements: statements true on both naïve and scientific conceptions of the domain (e.g., "steal is denser than foam"); statements false on both naïve and scientific conceptions of the domain (e.g., "foam is denser than brick"); statements that true on naïve conceptions of the domain but false on scientific conceptions (e.g., "ice is denser than water"), and statements that true on scientific conceptions of the domain but false on naïve conceptions (e.g., "cold pennies are denser than hot pennies"). The first two statements were classified as "consistent," because their truth-value was consistent across both naïve and scientific theories, and the last two were classified as "inconsistent," because their truth-value was *not* consistent across the two theories. Sample statements of each type and from each domain are displayed in Table 1.

This design proved useful for a number of reasons. First, it ensured there would be an equal number of objectively true and objectively false statements (as determined by their correspondence with the scientific theory), which would discourage participants from adopting a response bias toward one truth-value or the other. Second, it ensured that

Table 1: Sample items from the early-developing domains (top) and late-developing domains (bottom) involving the same conceptual relation. Consistent items were true on both the naïve theory (T1) and the scientific theory (T2) or false on both theories; inconsistent items were true on one theory but false on the other.

Domain	Type	T1	T2	Statement
Fractions	Consistent	True	True	12/13 is greater than 1/13.
		False	False	1/17 is greater than 16/17.
	Inconsistent	True	False	1/17 is greater than 1/9.
		False	True	1/13 is greater than 1/30.
Germs	Consistent	True	True	Being sneezed on can make a person sick.
		False	False	Being happy can make a person sick.
	Inconsistent	True	False	Being cold can make a person sick.
		False	True	Being depressed can make a person sick.
Inheritance	Consistent	True	True	Hair color is heritable.
		False	False	Pierced ears are heritable.
	Inconsistent	True	False	Immunity to chickenpox is heritable.
		False	True	Intelligence is heritable.
Matter	Consistent	True	True	Rocks are composed of matter.
		False	False	Numbers are composed of matter.
	Inconsistent	True	False	Fire is composed of matter.
		False	True	Air is composed of matter.
Physiology	Consistent	True	True	Fish are alive.
		False	False	Rocks are alive.
	Inconsistent	True	False	The sun is alive.
		False	True	Coral is alive.
Astronomy	Consistent	True	True	The moon revolves around the Earth.
		False	False	The sun revolves around the moon.
	Inconsistent	True	False	The sun revolves around the Earth.
		False	True	The Earth revolves around the sun.
Evolution	Consistent	True	True	Humans are descended from tree-dwelling creatures.
		False	False	Humans are descended from plants.
	Inconsistent	True	False	Humans are descended from chimpanzees.
		False	True	Humans are descended from sea-dwelling creatures.
Mechanics	Consistent	True	True	A moving bullet loses speed.
		False	False	A moving bullet loses weight.
	Inconsistent	True	False	A moving bullet loses force.
		False	True	A moving bullet loses height.
Thermodynamics	Consistent	True	True	Ovens produce heat.
		False	False	Rain produces heat.
	Inconsistent	True	False	Coats produce heat.
		False	True	Pressure produces heat.
Waves	Consistent	True	True	Red objects reflect red light.
		False	False	Red objects reflect blue light.
	Inconsistent	True	False	Red objects absorb red light.
		False	True	Red objects absorb blue light.

the complexity of each conceptual relation was balanced across statement type (consistent vs. inconsistent) and response type (true vs. false) so that simple relations (e.g., “[entity] evolves over time”) were represented as often as complex relations (e.g., “[species$_1$] is descended from [species$_2$]”) within each stimulus category. Third, it ensured that the average number of words per statement remained relatively constant across both statement type (consistent vs. inconsistent) and response type (true vs. false), which was a necessary precondition for comparing response times across categories. Additional care was taken to balance the average number of words per statement across domains, which ranged from 5.1 to 6.9 and averaged 5.9. No domain was an outlier on this measure.

Procedure

Stimuli were presented to participants with MediaLab v1.21 software, which also recorded the speed and accuracy of their truth judgments. The mean response time across items and across subjects was 3928 ms, and all response times that fell more than two standard deviations beyond this mean (200 of 13200, or 1.5%) were eliminated from the dataset.

Participants were instructed to verify each statement as quickly as possible without sacrificing accuracy. Responses were recorded by pressing “1” or “2” on the keypad, corresponding to “True” and “False” respectively. Once a response had been selected participants were then asked to rate the certainty of that response on a 5-point scale (1 = not certain, 2 = 25% certain, 3 = 50% certain, 4 = 75% certain, 5 = 100% certain). Confidence ratings were included mainly to provide a brief respite between statements and are not analyzed here. It should be noted, however, that the mean confidence rating was 4.1 and the modal confidence rating was 5 (or “100% certain”), indicating that participants were highly confident in their responses.

Statements from the same domain ($n = 20$) were presented as a block, in order to minimize abrupt changes in content, but their presentation was randomized within that block, and the presentation of each domain was randomized as well. No two participants therefore encountered the same domains or the same statements in the same order.

Results

Response Accuracy

Participants' accuracy at verifying statements whose truth value was consistent across naïve and scientific theories (“consistent statements”) and statements whose truth value differed across those theories (“inconsistent statements”) is displayed in Table 2. Participants correctly verified consistent statements significantly more often than they correctly verified inconsistent statements in all ten domains (paired samples $t(64) > 3.9$, $p < .01$ for all comparisons).

Table 2: Mean proportion of consistent and inconsistent statements correctly verified in the early-developing domains (top) and late-developing domains (bottom).

Domain	Consist.	Inconsist.	Difference
Fractions	.90	.74	.16**
Germs	.90	.69	.21**
Inheritance	.78	.70	.08**
Matter	.90	.61	.29**
Physiology	.94	.71	.23**
Astronomy	.83	.66	.17**
Evolution	.72	.47	.25**
Mechanics	.78	.50	.28**
Thermodynamics	.85	.61	.24**
Waves	.81	.59	.22**

A repeated-measures analysis of variance (ANOVA) was used to determine whether, and how, the effects of statement type (consistent vs. inconsistent) varied by domain type (early-developing vs. late-developing). This analysis revealed a significant main effect of statement type ($F(1,64) = 705.73$, $p < .001$), as expected, but it also revealed a significant main effect of domain type ($F(1,64) = 196.00$, $p < .001$), with participants demonstrating greater accuracy for statements about early-developing domains ($M = 0.79$) than for statements about late-developing domains ($M = 0.68$), and a significant interaction between statement type and domain type ($F(1,64) = 9.36$, $p < .01$). The source of this interaction was a 21% greater difference between consistent and inconsistent statements for the late-developing domains ($M_{\text{diff.}} = 0.23$) than for the early-developing domains ($M_{\text{diff.}} = 0.19$). Participants apparently erred on more inconsistent statements in the late-developing domains than in the early-developing domains, implying that naïve theories held greater sway in the late-developing domains. This finding is consistent with the fact that many high school students never actually make conceptual change in domains like mechanics (McCloskey, 1983) or evolution (Shtulman, 2006), though participants had most likely been introduced to the relevant scientific theories previously.

Response Latency

Participants' speed at verifying consistent and inconsistent statements is displayed in Table 3. As predicted, participants verified consistent statements more quickly than they verified inconsistent statements in all ten domains, though the effect was statistically significant in only eight (fractions, $t(64) = 7.94$, $p < .01$; germs, $t(64) = 4.31$, $p < .01$; matter, $t(64) = 5.33$, $p < .01$; physiology, $t(64) = 6.25$, $p < .01$; astronomy, $t(64) = 2.97$, $p < .05$; evolution, $t(64) = 2.06$, $p < .05$; thermodynamics, $t(64) = 5.85$, $p < .01$; waves, $t(64) = 4.52$, $p < .01$). These effects remained significant even when incorrect responses were removed from the dataset (fractions, $t(64) = 9.67$, $p < .01$; germs, $t(64) = 5.58$, $p < .01$; matter, $t(64) = 3.12$, $p < .01$; physiology, $t(64) = 5.44$, $p < .01$; astronomy, $t(64) = 2.80$, $p < .01$; evolution, $t(64) = 2.21$, $p < .05$; thermodynamics, $t(64) = 6.85$, $p < .001$; waves, $t(64) = 3.13$, $p < .01$). In fact, one non-significant effect became significant on this analysis (mechanics, $t(64) = 2.33$, $p < .05$).

Table 3: Mean response times (ms) for consistent and inconsistent statements in the early-developing domains (top) and late-developing domains (bottom).

Domain	Consist.	Inconsist.	Difference
Fractions	3534	4134	-600**
Germs	2720	2976	-256**
Inheritance	3667	3769	-102
Matter	3248	3652	-404**
Physiology	2832	3350	-518**
Astronomy	3457	3700	-243*
Evolution	3800	3971	-172*
Mechanics	3997	4149	-151
Thermodynamics	3591	4151	-560**
Waves	3480	3853	-374**

As with the accuracy data, the latency data were analyzed with a repeated-measures ANOVA for effects of statement type (consistent vs. inconsistent) and domain type (early-developing vs. late-developing). This analysis revealed significant main effects of both (statement type: $F(1,64) = 188.94$, $p < .001$; domain type: $F(1,64) = 52.40$, $p < .001$), indicating that participants were not only faster at verifying consistent statements (*M* = 3324) than at verifying inconsistent statements (*M* = 3765) but were also faster at verifying statements about early-developing domains (*M* = 3387) than at verifying statements about late-developing domains (*M* = 3702). This analysis also revealed a significant interaction between statement type and domain type ($F(1,64) = 10.43$, $p < .01$), owing to a greater difference between consistent and inconsistent statements in the late-developing domains ($M_{\text{diff.}} = 507$) than in the early-developing ones ($M_{\text{diff.}} = 375$).

Thus, just as participants were *most accurate* at verifying consistent statements in early-developing domains (*M* = 88% correct) and *least accurate* at verifying inconsistent statements in late-developing domains (*M* = 57% correct), they were *fastest* at verifying consistent statements in early-developing domains (*M* = 3200 ms) and *slowest* at verifying inconsistent statements in late-developing domains (*M* = 3956 ms). While the accuracy data seem to suggest that many participants simply defaulted to naïve theories in the late-developing domains, the latency data do not, for if they had, the difference in response latency between consistent and inconsistent statements in the late-developing domains should have been smaller, not larger, than that in the early-developing domains. Thus, participants appeared to experience more cognitive conflict between naïve and scientific theories in the late-developing domains than in the early-domains, perhaps because the relevant scientific theory was less well understood or less well consolidated.

Discussion

A significant component of science education is replacing inaccurate, pre-instructional theories of natural phenomena with more accurate, scientific ones, but what happens to the original theories? Our findings strongly suggest that those theories are *suppressed*, not *supplanted*. Across many different domains, participants were significantly faster and more accurate at verifying statements whose truth-value remained constant across a conceptual change (e.g., "people turn food into energy," "whales are more closely related to fish than to plants") than at verifying structurally analogous statements whose truth-value reversed (e.g., "bacteria turn food into energy," "whales are more closely related to humans than to fish"). This effect was observable not only in domains where participants were likely to have undergone conceptual change within the past five years but also in domains where they were likely to have undergone conceptual change as children, implying that even children's naïve theories remain intact, at some level of representation, following conceptual change.

These findings are consistent with many other recent findings, including Lombrozo et al.'s (2007) demonstration that Alzheimer's patients endorse teleological explanations of natural phenomena typically endorsed only by children and Goldberg and Thompson-Schill's (2009) demonstration that adults are slower and less accurate at classifying plants as alive than at classifying animals as alive. They extend these findings, however, by showing that this effect is pervasive across many domains of knowledge – from the life sciences (physiology, evolution) to the physical sciences (astronomy, thermodynamics) to mathematics (fractions) – and across many conceptual relations within those domains. They also extend this literature by showing that the influence of naïve theories appears to diminish over time, though the magnitude of difference between early-developing and late-developing domains was much smaller than the magnitude of difference between consistent and inconsistent statements.

While the present findings are consistent with other findings in the literature, they are not particularly consistent with many prominent models of conceptual change, like Carey's (2009) differentiation-coalescence model or Chi's (2008) ontology-reassignment model, because these models assume a kind of conceptual restructuring that should render early modes of thought inaccessible. While these models could certainly be amended to account for the data at hand, doing so would require a specification of how a single concept, like *heat* or *force*, could hold different meanings for the very same individual across different contexts or different tasks.

One model of conceptual change that may hold more promise in this regard is Ohlsson's (2009) "resubsumption" model, in which a class of phenomena (e.g., the motion of objects) initially subsumed under one theory (e.g., impetus theory) is *re*-subsumed under another (e.g., Newtonian mechanics). The process is initiated by the learner noticing a correspondence between one class of phenomena and two different theories – say, a naïve theory and a scientific theory – and is completed by a competitive evaluation of the cognitive utility of each correspondence. While this model does not currently explain where scientific theories come from or how learners notice correspondences between scientific theories and phenomena currently subsumed by other theories (Shtulman, 2009), it does help to explain the findings documented here in that it explicitly stipulates that the outcome of conceptual change is not a single theory, derived step-by-step from some parent theory, but *two* theories in active competition.

An important prediction of the resubsumption model is that the longer an individual holds a scientific theory, the less he or she should rely on the naïve theory with which it competes (assuming, of course, that the scientific theory will ultimately prove more useful in its explanatory power and explanatory scope than the naïve one). Consistent with this prediction, we found that naïve theories held less sway for early-developing domains than for late-developing domains, presumably because the competition between naïve theories and scientific theories had been resolved (in favor of the scientific theories) to a much larger extent in the early-developing domains than in the late-developing ones. Similar results were obtained by Goldberg and Thompson-Schill (2009) in that biologists exhibited less cognitive conflict classifying plants as alive than undergraduates did, even though both groups exhibited more conflict classifying plants as alive than classifying animals as alive. Expertise in biology apparently decreased the influence of a long-discarded ontological distinction between plants and animals but did not eliminate that distinction altogether.

Future research could explore this issue more directly by explicitly manipulating the cognitive utility of a newly acquired scientific theory – by, for example, involving students in inferential-reasoning tasks easily accomplished in light of the scientific theory but not possible to accomplish in light of the naïve theory – and measuring the effects of that manipulation on the speed and accuracy of their responses to the kinds of stimuli presented here. If cognitive utility is indeed a determinant of the continued survival of a naïve theory, then decreasing the cognitive utility of that theory should decrease the likelihood that it will be activated when reasoning about the phenomena it once subsumed. No manipulation, however, would likely the eliminate the effect altogether, as the extant data suggest that naïve theories continue to persist even after decades of disuse and disrepair.

Acknowledgments

This work was supported by NSF grant DRL-0953384 to Andrew Shtulman.

References

Au, T. K. F., Chan, C. K. K., Chan, T. K., Cheung, M. W. L., Ho, J. Y. S., & Ip, G. W. M. (2008). Folkbiology meets microbiology: A study of conceptual and behavioral change. *Cognitive Psychology, 57*, 1-19.

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

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change.* Hillsdale, NJ: Erlbaum.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*, 66-70.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science, 20*, 480-487.

Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning and Instruction, 8*, 341-374.

Johnson, S. C., & Carey, S. (1998). Knowledge enrichment and conceptual change in folkbiology: Evidence from Williams Syndrome. *Cognitive Psychology, 37*, 156-200.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 11*, 138-143.

Legare, C., & Gelman, C. H. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science, 32*, 607-642.

Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations for patients with Alzheimer's disease. *Psychological Science, 18*, 999-1006.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner and A. L. Stevens (Eds.), *Mental Models* (pp. 299-324). Hillsdale, NJ: Erlbaum.

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Science Education, 30*, 122-147.

Nakhleh, M. B., Samarapungavan, A., & Saglam, Y. (2005). Middle school students' beliefs about matter. *Journal of Research in Science Teaching, 42*, 581–612.

Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*, 20-40.

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naïve physics reasoning: A commitment to substance-based conceptions. *Cognition and Instruction, 18*, 1-34.

Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review, 14*, 890-894.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*, 170-194.

Shtulman, A. (2009). Rethinking the role of resubsumption in conceptual change. *Educational Psychologist, 44*, 41-47.

Shtulman, A. & Schulz, L. (2008). Essentialist beliefs about species and their relationship to evolutionary reasoning. *Cognitive Science, 32*, 1049-1062.

Siegal, M., Butterworth, G., & Newcombe, P. A. (2004). Culture and children's cosmology. *Developmental Science, 7*, 308-324.

Slaughter, V., & Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology, 43*, 1-30.

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories. *Cognition and Instruction, 25*, 337-398.

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.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45-69.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the day/night cycle. *Cognitive Science, 18*, 123-183.

Wellman, H. M. & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43*, 337-375.

Wiser, M., & Amin, T. (2001). Is heat hot? Inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction, 11*, 331-355.