Cognition 124 (2012) 209–215

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Scientific knowledge suppresses but does not supplant earlier intuitions

Andrew Shtulman *, Joshua Valcarcel

Department of Psychology, Occidental College, 1600 Campus Road, Los Angeles, CA 90041, United States

ARTICLE INFO

Article history:
Received 4 August 2011
Revised 16 April 2012
Accepted 19 April 2012
Available online 16 May 2012

Keywords:
Naive theories
Knowledge representation
Conceptual change
Science education

ABSTRACT

When students learn scientific theories that conflict with their earlier, naïve theories, what happens to the earlier theories? Are they overwritten or merely suppressed? We investigated this question by devising and implementing a novel speeded-reasoning task. Adults with many years of science education verified two types of statements as quickly as possible: statements whose truth value was the same across both naïve and scientific theories of a particular phenomenon (e.g., "The moon revolves around the Earth") and statements involving the same conceptual relations but whose truth value differed across those theories (e.g., "The Earth revolves around the sun"). Participants verified the latter significantly more slowly and less accurately than the former across 10 domains of knowledge (astronomy, evolution, fractions, genetics, germs, matter, mechanics, physiology, thermodynamics, and waves), suggesting that naïve theories survive the acquisition of a mutually incompatible scientific theory, coexisting with that theory for many years to follow.

1. Introduction

Decades of research in cognitive psychology, developmental psychology, and science education have dispelled the myth that students enter the science classroom as "empty vessels" ready to be filled with knowledge. Rather, they enter with rich, pre-instructional theories of the domain-relevant phenomena that often interfere with learning (Carey, 2000; Keil, 2011; Vosniadou, 1994). In the domain of mechanics, for instance, students hold theories of motion predicated on the belief that forces are transferred from one object to another upon contact and must dissipate before objects can come to a rest (Clement, 1982; McCloskey, 1983). In the domain of thermodynamics, students hold theories of heat predicated on the belief that heat is a kind of substance that flows in and out of objects and can ultimately be trapped or contained (Reiner, Slotta, Chi, & Resnick, 2000; Wiser & Amin, 2001). And in the domain of evolution, students hold theories of adaptation predicated on the belief that all members of a species evolve together, with each organism producing offspring better adapted to the environment than it was at birth (Shtulman, 2006; Shtulman & Schulz, 2008).

Science educators are thus charged with two tasks: not only must they help students learn the correct, scientific theory at hand, but they must also help students *unlearn* their earlier, less accurate theories. Psychologists who have studied this process – typically termed "conceptual change" – have characterized the transition from naïve theories to scientific theories in several ways. Some have emphasized the role of category knowledge, characterizing conceptual change as a series of conceptual differentiations, in which new category boundaries are established, and conceptual coalescences, in which old category boundaries are collapsed (Carey, 2009; Smith, 2007). Some have emphasized the role of ontological hierarchies, characterizing conceptual change as the reassignment of a key concept, or system of concepts, from one branch of an ontological hierarchy to another (Chi, Slotta, & de Leeuw, 1994; Thagard, 1992). And some have emphasized the role of causal expectations, characterizing conceptual change as a revision of the core presuppositions of a causal model or causal theory (Vosniadou, 1994; Wellman & Gelman, 1992). Common to all characterizations is a commitment

---

* Corresponding author. Address: Department of Psychology, Occidental College, Los Angeles, CA 90041, United States.
E-mail address: shtulman@oxy.edu (A. Shtulman).

0010-0277/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.04.005

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 the earlier 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. Alzheimer's patients, for instance, have been shown to endorse teleological explanations for natural phenomena that typically only children endorse (Lombrozo, Kelemen, & Zaitchik, 2007). While adults without Alzheimer's disease do not typically endorse these explanations, they can be induced to do so under speeded conditions (Kelemen & Rosset, 2009). Adults have also been shown to be slower and less accurate at categorizing plants as alive than at categorizing animals as alive (Goldberg & Thompson-Schill, 2009), reminiscent of young children's belief that animate entities are alive but inanimate entities are not (Piaget, 1929). Even biology professors were found to be slower and less accurate at classifying plants as alive than at classifying animals as alive, implying that years of professional experience had not erased an erroneous distinction first drawn in childhood (Goldberg & Thompson-Schill, 2009).

These findings, among others (Dunbar, Fugelsang, & Stein, 2007; Legare & Gelman, 2008; Zaitchik & Solomon, 2008, 2009), suggest that scientific knowledge serves to mask, rather than replace, one's initial intuitions. They are consistent with the views of many science education researchers, who see the purpose of instruction as reanalyzing, rather than replacing, pre-instructional ideas (e.g., Caravita & Hallden, 1994; Clement, 1993). They are also consistent with models of theory change that emphasize the inferential competition *between* theories rather than the replacement of one theory by another (e.g., Ohlsson, 2009; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Nevertheless, current evidence for the persistence of early intuitions is limited to studies that tap a single conceptual distinction, like the distinction between living and nonliving things (Zaitchik & Solomon, 2008) or the distinction between teleological and nonteleological processes (Kelemen & Rosset, 2009). While these distinctions play a central role in folkbiology, it's unclear whether the effect extends to other concepts within the same domain or to concepts from other domains altogether. Documenting the exact scope of this phenomenon is critical to understanding its origin. While, on the one hand, the resilience of early intuitions may have implications for conceptual change as a whole, it may, on the other hand, have implications for only a particular kind of knowledge or a particular kind of learning.

Here, we provide evidence that the resilience of early intuitions is, in fact, a domain-general phenomenon by introducing a new method for probing those intuitions that can be applied to potentially *any* concept in *any* domain. Specifically, we compare the speed and accuracy with which adults verify two types of statements: statements whose truth-value is known to remain constant across conceptual change (e.g., "The moon revolves around the Earth," which is true on both naïve and scientific theories of astronomical phenomena) and syntactically analogous statements whose truth-value is known to reverse across that same change (e.g., "The Earth revolves around the sun," which is true on a scientific theory but not a naïve theory). We hypothesized that, if naïve theories survive the acquisition of a mutually incompatible scientific theory, then statements whose truth-value reverse across conceptual change should cause greater cognitive conflict than statements whose truth-value remain constant, resulting in slower and less accurate verifications for those statements. 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 more conflict than statements whose truth-value remain constant, since the naïve theories should no longer be present to cause such conflict.

2. Method

2.1. Participants

The participants were 150 college undergraduates recruited from introductory psychology courses. They reported having taken an average of 3.1 college-level math and science courses prior to the study (*SD* = 3.3), but this variable did not predict any of the effects reported below.

2.2. Materials

Participants verified, as quickly as possible, 200 statements about natural phenomena: 20 statements in each of 10 domains of knowledge, with each statement exemplifying one of five concepts within that domain (see Table 1). A quarter of the statements were true on both naïve and scientific theories of the domain ("steal is denser than foam"), a quarter were false on both naïve and scientific theories ("foam is denser than brick"), a quarter were true on naïve theories but false on scientific theories ("ice is denser than water"), and a quarter were true on scientific

Table 1
The five concepts covered in each domain.

Domain	Concept
Astronomy	Planet, star, solar system, lunar phase, season
Evolution	Common ancestry, phylogeny, variation, selection, adaptation
Fractions	Addition, division, conversion, ordering, infinite density
Genetics	Heritability, chromosome, dominance, expression, mutation
Germs	Contagion, contamination, infection, sterilization, microbe
Matter	Mass, weight, density, divisibility, atom
Mechanics	Force, velocity, acceleration, momentum, gravity
Physiology	Life, death, reproduction, metabolism, kinship
Thermodynamics	Heat, heat source, heat transfer, temperature, thermal expansion
Waves	Light, color, sound, propagation, reflection

Table 2
Sample items 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 (marked with an asterisk) were true on one theory but false on the other.

Domain	T1	T2	Statement
Matter	True	True	Rocks are composed of matter.
	False	False	Numbers are composed of matter
	True	False	Fire is composed of matter*
	False	True	Air is composed of matter*
Physiology	True	True	People turn food into energy
	False	False	Rocks turn food into energy
	True	False	Plants turn food into energy*
	False	True	Bacteria turn food into energy*
Evolution	True	True	Humans are descended from tree-dwelling creatures
	False	False	Humans are descended from plants
	True	False	Humans are descended from chimpanzees*
	False	True	Humans are descended from sea-dwelling creatures*
Mechanics	True	True	A moving bullet loses speed
	False	False	A moving bullet loses weight
	True	False	A moving bullet loses force*
	False	True	A moving bullet loses height*
Thermodynamics	True	True	Ovens produce heat
	False	False	Rain produces heat
	True	False	Coats produce heat*
	False	True	Pressure produces heat*

theories but false on naïve theories ("cold pennies are denser than hot pennies").

Sample statements are shown in Table 2. All materials were derived from prior research in cognitive development and science education on the particular form of restructuring that occurs in each domain. For the domain of astronomy, the statements were designed to assess the transition from a geocentric model of the solar system to a heliocentric model (Siegal, Butterworth, & Newcombe, 2004; Vosniadou & Brewer, 1994); for evolution, the transition from an essentialist theory of species adaptation to a selection-based theory (Shtulman, 2006; Shtulman & Schulz, 2008); for fractions, the transition from an integer-based model of rational number to a division-based model (Hartnett & Gelman, 1998; Moss & Case, 1999); for genetics, the transition from a trait-based theory of inheritance to a chromosomal theory (Duncan, Rogat, & Yarden, 2009; Springer & Keil, 1989); for germs, the transition from a behavioral theory of illness to a microbial theory (Au et al., 2008; Solomon & Cassimatis, 1999); for matter, the transition from a tactile theory of material substances to a particulate theory (Nakhleh, Samarapungavan, & Saglam, 2005; Smith, 2007); for mechanics, the transition from an impetus theory of motion to an inertial theory (Clement, 1982; McCloskey, 1983); for physiology, the transition from a psychological theory of bodily functions to a vitalist theory (Johnson & Carey, 1998; Slaughter & Lyons, 2003); for thermodynamics, the transition from a substance-based theory of heat to a kinetic theory (Slotta & Chi, 2006; Wiser & Amin, 2001); and for waves, the transition from a substance-based theory of light and sound to an frequency-based theory (Linder & Erickson, 1989; Mazens & Lautrey, 2003). The full list of 200 statements can be found online in the Supplemental materials.

The materials were counterbalanced in three respects: (1) there were an equal number of objectively true and objectively false statements per domain, discouraging the adoption of a response bias; (2) 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; and (3) the average number of words per statement was held constant across statement type, response type, and domain.

2.3. Procedure

Stimuli were presented to participants with MediaLab v1.21 software, which recorded the speed and accuracy of their judgments. The mean response time across items and across subjects was 3971 ms, and all response times that fell more than two standard deviations beyond this mean were eliminated from the dataset. Statements from the same domain ($n = 20$) were presented as a block (prefaced by the domain name), in order to minimize abrupt changes in content, but their ordering was randomized within that block, and the ordering of the domains was randomized as well.

3. Results

Participants' accuracy at verifying statements whose truth value was consistent across both naïve and scientific theories ("consistent statements") and statements whose truth value differed across those theories ("inconsistent statements") is displayed in Fig. 1A. A repeated-measures analysis of variance (ANOVA) confirmed that participants were significantly more accurate at verifying the former than the latter ($F(1,149) = 1582.49$, $p < .001$), and simple effects tests confirmed that this difference was robust

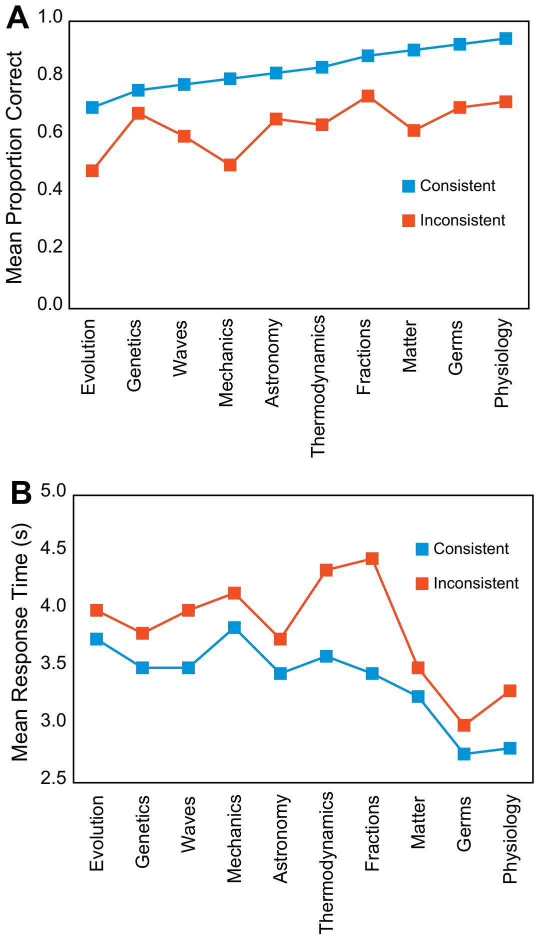

Fig. 1. (A) Mean proportion of consistent and inconsistent statements correctly verified in each domain (all SE < .02). (B) Mean response times (in seconds) for correct responses (all SE < .1).

across domains ($t(149) > 5.0$, $p < .001$ for all within-domain comparisons).

The speed at which participants made their verifications is displayed in Fig. 1B. Only correct verifications are included in Fig. 1B, though the results do not change if incorrect verifications are included as well. As predicted, participants were significantly slower at verifying inconsistent statements than at verifying consistent ones, both across domains ($F(1,149) = 349.98$, $p < .001$) and within-domains ($t(149) > 3.0$, $p < .01$ for all within-domain comparisons). This pattern was robust across individual concepts as well; the average response latency for inconsistent statements was greater than that for consistent statements for 43 of the 50 concepts (see Table 3).

To determine whether the effect varied by response type ("true" vs. "false"), we collapsed the latency data across domains and analyzed them for effects of response type and statement type with a repeated-measures ANOVA. Not surprisingly, participants verified true statements significantly faster than they verified false statements ($M = 3698$ ms vs. $M = 3466$ ms, $F(1,149) = 71.04$, $p < .001$; see Fig. 2), but there was no interaction between statement type and response type, indicating that participants correctly verified consistent statements faster than inconsistent ones regardless of whether those statements were true on scientific theories but false on naïve theories or true on naïve theories but false on scientific theories. Indeed, the mean difference in response latency between consistent and inconsistent statements for the false items (415 ms) was virtually the same as that for the true items (428 ms).

Table 3
Distribution of conceptual relations in each domain ($n = 5$) as a function of the difference in response times between consistent and inconsistent statements; positive differences are indicative of a conceptual relation in which participants were faster to respond to consistent versions than to inconsistent versions.

Domain	Mean difference in response time			
	<0 ms	0–500 ms	500–1000 ms	>1000 ms
Astronomy	2	1	1	1
Evolution	0	4	1	0
Fractions	1	0	1	3
Genetics	2	1	2	0
Germs	1	3	1	0
Matter	0	3	1	1
Mechanics	0	3	1	1
Physiology	0	3	0	2
Thermodynamics	0	1	3	1
Waves	1	1	3	0
Total	7	20	14	9

Lastly, we compared response latency to response accuracy. Overall, participants were more accurate for domains in which conceptual change tends to occur within the first decade of life (e.g., fractions, physiology) than for domains in which conceptual change tends to occur in the second decade of life (e.g., evolution, mechanics), if at all. Accompanying this difference in accuracy was a difference in speed: participants verified consistent statements an average of 562 ms faster than inconsistent statements in domains for which overall accuracy was 80% or greater (fractions, germs, physiology), 388 ms faster in domains for which overall accuracy was between 70% and 80% (astronomy, genetics, thermodynamics, matter), and 334 ms faster in domains for which overall accuracy was 70% or below (mechanics, evolution, waves). A contrast analysis confirmed that speed varied linearly with accuracy ($F(1,149) = 23.40$, $p < .001$), indicating that participants were slower to verify inconsistent statements in domains where they were more likely to verify those statements correctly.

This effect was observed at the level of the participant as well. Although all participants exhibited above-chance levels of accuracy ($t(199) > 2.5$, $p < .05$ for all comparisons to chance), individual participants ranged from 59% correct to 89% correct. Fig. 3 displays the mean response times for participants in each quartile of the accuracy distribution. Response times for the consistent statements did not vary by quartile, but response times for the inconsistent statements did ($F(3,146) = 2.69$, $p < .05$), with participants in the fourth quartile responding an average of 465 ms slower than participants in the first quartile. Indeed, the difference in response times between inconsistent and consistent statements was significantly correlated with accuracy scores in six of the 10 domains (fractions: $r = .27$; genetics:

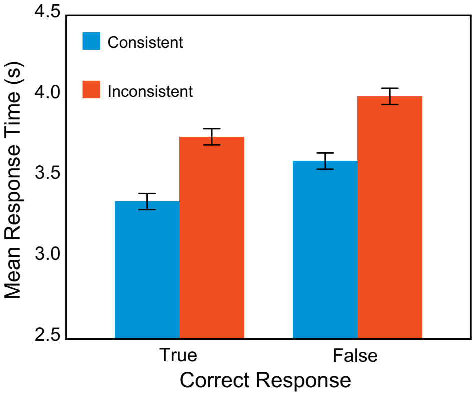

Fig. 2. Mean response times (in seconds) as a function of statement type (consistent vs. inconsistent) and truth-value (true vs. false).

$r = .19$; germs: $r = .27$; matter: $r = .21$; physiology: $r = .22$; thermodynamics: $r = .18$; all p's < .05), suggesting that participants with greater domain expertise exhibited more cognitive conflict on the inconsistent items than did participants with less expertise. Although this finding needs to be explored further, it indicates, at the very least, that the effect was not an artifact of confusion or carelessness.

4. Discussion

When students learn scientific theories that conflicts with earlier, naïve theories, what happens to the earlier theories? Our findings suggest that naïve theories are *suppressed* by scientific theories but not *supplanted* by them. Across 10 domains, participants were significantly slower and less accurate at verifying statements whose truth-value reversed across a conceptual change (e.g., "1/13 is greater than 1/30") than at verifying structurally analogous statements whose truth-value remained constant across that change (e.g., "12/13 is greater than 1/13"). This effect was observed not only in domains where participants were introduced to the correct, scientific concepts in late adolescence but also in domains where they were introduced to those concepts in early childhood. Indeed, the latency data suggest that participants exhibited more cognitive conflict in the latter than in the former, possibly because naïve theories in the latter domains emerge earlier and are thus more deeply entrenched.

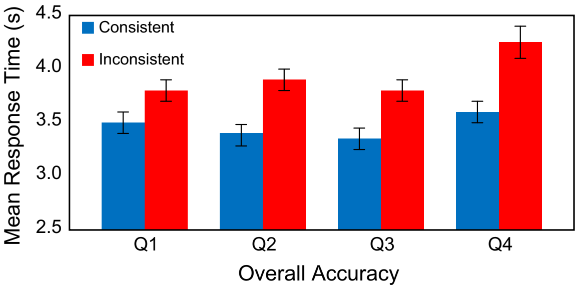

Fig. 3. Mean response times (in seconds) as a function of statement type (consistent vs. inconsistent) and overall accuracy (by quartile).

Might the observed differences between consistent and inconsistent statements reflect something other than the "tug" of a previously discarded theory? One possibility is that the inconsistent statements were syntactically more difficult to process than the consistent ones, yet this possibility is unlikely given how the two types of statements were constructed. Consistent and inconsistent statements did not differ, after all, in the complexity of their linguistic form (e.g., "[entity] has heat") but in the objects or properties to which a particular linguistic expression was applied (e.g., "the sun has heat" vs. "ice has heat"). Clearly, the inconsistent statements were more *counterintuitive* than the consistent statements but that difference constituted the manipulation of interest. Moreover, the sense in which the inconsistent statements were counterintuitive was not random or capricious but was rather constrained by decades of research in the relevant domains. For instance, in the domain of fractions, the relation "is greater than" was verified more quickly and more accurately when applied to the numbers 12/13 and 1/13 than when applied to the numbers 1/13 and 1/30. The reason for this difference presumably stems from the fact that our initial understanding of number, acquired in the context of counting, lacks the properties of divisibility and infinite density needed to represent non-integral quantities and must therefore be restructured before the concept of a denominator can be grasped (Moss & Case, 1999; Smith, Solomon, & Carey, 2005; Vamvakoussi & Vosniadou, 2004). While there may be something intrinsically more difficult about comparing numerators than comparing denominators, a more feasible explanation, in light of the math education literature, is that this difference is a historical byproduct of the conceptual trajectory from an understanding of *whole number* to that of *rational number*.

Another possibility is that the inconsistent statements represented information that was simply less familiar than that represented by the consistent statements, with the less familiar information proving more difficult to retrieve than the more familiar information. The problem with this explanation is that what constitutes "familiar" and "unfamiliar" should vary by when learning has occurred, yet the effect of interest was observed for all 10 domains, as well as the vast majority of concepts within those domains. While it is true that the effect was stronger in some domains (e.g., fractions, germs) than in others (e.g., mechanics, waves), this difference was actually in the opposite direction of that predicted by a familiarity account in that participants were slowest to verify inconsistent statements in the familiar domains, not the unfamiliar ones. Moreover, there were no consistent differences in how participants treated statements that were true on naive theories but false on scientific theories (i.e., positive misconceptions) and statements that were true on scientific theories but false of naive theories (i.e., negative misconceptions). These two types of misconceptions ostensibly vary in their familiarity by domain; positive misconceptions are ostensibly less familiar in early-developing domains (e.g., "the sun is alive") than in late-developing domains (e.g., "the earth's distance from the sun causes the season") owing to the amount of time that has passed since the relevant intuitions were first challenged, and negative misconceptions are ostensibly less familiar in late-developing domains (e.g., "the earth's tilt causes the seasons") than in early-developing domains (e.g., "coral is alive") owing to the amount of time that has passed since the relevant scientific information was first introduced. However, a statistical comparison of the response lag between consistent and inconsistent statements associated with positive misconceptions and that associated with negative misconceptions revealed no interaction between type of misconception (positive vs. negative) and domain ($F(9, 1341) < 1$, *ns*).

Our findings are thus inconsistent with either alternative. They are, however, consistent with previous findings regarding the re-emergence of teleological thought and animistic thought under cognitive impairment (Lombrozo et al., 2007; Zaitchik & Solomon, 2008) or cognitive load (Goldberg & Thompson-Schill, 2009; Kelemen & Rosset, 2009). And they do not merely replicate those findings; they extend them across multiple domains of knowledge – from the life sciences to the physical sciences to mathematics – and across multiple concepts within those domains. Indeed, the consistency of the effect within and across domains suggests that it is not merely the byproduct of a few particularly resilient intuitions but is rather a domain-general consequence of conceptual restructuring. Consequently, these findings present a strong challenge to standard models of conceptual change, like coalescence-and-differentiation models (Carey, 2009) or ontological-reassignment models (Chi et al., 1994), as they demonstrate the continued influence of intuitions that should have been rendered obsolete by the kinds of radical restructuring these models entail. While such models could 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 same individual across different contexts or different tasks.

One possible modification would be to stipulate that naïve theories and scientific theories take different *forms* of representation. Naïve theories might take a more contextual or more phenomenological representation, whereas scientific theories might take a more abstract or more propositional representation (see DiSessa, 1993). While this view seems plausible for domains in which naïve theories are constructed from firsthand sensory data (e.g., mechanics, thermodynamics), it seems less plausible for domains in which the relevant phenomena must be learned through testimony and inference (e.g., astronomy, evolution). Another possibility is that naïve theories and scientific theories occupy different *levels* of representation. On this view, beliefs about an entire ontology (e.g., matter) may be represented at a different level of abstraction than beliefs about a particular instance of that ontology (e.g., ice), giving rise to potentially many undetected inconsistencies (e.g., "all matter has heat" vs. "ice has no heat"). Preliminary support for this view comes from studies showing that conceptual change is perhaps best facilitated by instructional approaches that emphasize the integration of knowledge across multiple levels of representation (Slotta & Chi, 2006; Wiser & Amin, 2001). Further research, however, is needed to determine whether successful instruction ever truly eliminates early, erroneous intuitions or whether such intuitions persist regardless of the comprehensiveness or systematicity of the newly acquired theory.

Further research is also needed to determine why early intuitions conflict with scientific knowledge to different degrees in different domains. One possibility is that intuitive theories in early developing domains hold more sway over reasoning than those in later developing domains because they are more deeply entrenched. In other words, intuitive theories in domains like physiology and germs are constructed earlier than those in domains like evolution and waves and may thus have influenced the encoding and organization of domain-specific information to a greater extent. Another possibility is that the intuitive theories in early developing domains hold no more sway than those in later developing domains but the *scientific* theories do. On this account, participants experienced greater cognitive conflict in early developing domains not because their pre-scientific intuitions in those domains were any more potent or resilient but because their newfound scientific beliefs posed a more formidable challenge to those intuitions. This explanation would account not only for differences in average response latencies across domains but also for differences between individual participants within the same domain. After all, participants who exhibited greater expertise in a particular domain (as indexed by response accuracy) also exhibited greater cognitive conflict in that domain (as indexed by response latency).

One way to disentangle these two possibilities would be to compare performance in the present task with performance on independent measures of conceptual understanding – measures that could probe understanding in more depth and more detail (e.g., structured interviews, inferential reasoning tasks, think-aloud protocols). While the "differential entrenchment" account predicts that differences in performance in the present task should be unrelated to differences in the quality or consistency of one's scientific knowledge (beyond some particular threshold), the "differential learning" account predicts that the two measures should be strongly related. Indeed, the differential learning account makes the counterintuitive prediction that the response lag between consistent and inconsistent statements should actually *increase* with expertise. While the present study provides some preliminary support for this hypothesis, stronger support would come from studies that directly compared the response profiles of novices and experts or the response profiles of novices before and after instruction. Critically, the effect should be domain-specific, such that instruction in physics should affect performance on the mechanics items but not the evolution items and instruction in biology should affect performance on the evolution items but not the mechanics items. Such findings, in combination with the present findings, promise to shed additional light not only on the relation between naïve theories and scientific theories but also on the nature of conceptual representation and conceptual change more generally.

Acknowledgment

This research was supported by National Science Foundation Grant DRL-0953384 awarded to Andrew Shtulman.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2012.04.005.

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.

Caravita, S., & Hallden, O. (1994). Reframing the problem of conceptual change. *Learning and Instruction, 4*, 89–111.

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

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

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27–43.

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

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30*, 1241–1257.

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*, 105–225.

Dunbar, K., Fugelsang, J., & Stein, C. (2007). Do naive theories ever go away? Using brain and behavior to understand changes in concepts. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 193–206). New York: Lawrence Erlbaum Associates.

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. *Journal of Research in Science Teaching, 46*, 655–674.

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.

Keil, F. C. (2011). Science starts early. *Science, 331*, 1022–1023.

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.

Linder, C. J., & Erickson, G. L. (1989). A study of tertiary physics students' conceptualizations of sound. *International Journal of Science Education, 11*, 491–501.

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.

Mazens, K., & Lautrey, J. (2003). Conceptual change in physics: Children's naïve representations of sound. *Cognitive Development, 18*, 159–176.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & 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.

Piaget, J. (1929). *The child's conception of physical causality*. London: Routledge.

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

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

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and 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.

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24*, 261–289.

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

Smith, C. L., Solomon, G. E. A., & Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology, 51*, 101–140.

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.

Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development, 60*, 637–648.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Structure, statistics, and abstraction. *Science, 331*, 1279–1285.

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

Vamvakoussi, X., & Vosniadou, S. (2004). Understanding the structure of the set of rational numbers: A conceptual change approach. *Learning and Instruction, 14*, 453–467.

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

Vosniadou, S., & Brewer, W. F. (1994). 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.

Zaitchik, D., & Solomon, G. E. A. (2008). Animist thinking in the elderly and in patients with Alzheimer's disease. *Cognitive Neuropsychology, 25*, 27–37.

Zaitchik, D., & Solomon, G. E. A. (2009). Conservation of species, volume, and belief in patients with Alzheimer's disease: The issue of domain specificity and conceptual impairment. *Cognitive Neuropsychology, 26*, 511–526.