TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 8 (2016) 118–137

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12174

Tensions Between Science and Intuition Across the Lifespan

Andrew Shtulman, Kelsey Harrington

Department of Psychology, Occidental College

Received 30 June 2014; received in revised form 11 February 2015; accepted 20 February 2015

Abstract

The scientific knowledge needed to engage with policy issues like climate change, vaccination, and stem cell research often conflicts with our intuitive theories of the world. How resilient are our intuitive theories in the face of contradictory scientific knowledge? Here, we present evidence that intuitive theories in 10 domains of knowledge—astronomy, evolution, fractions, genetics, germs, matter, mechanics, physiology, thermodynamics, and waves—persist more than four decades beyond the acquisition of a mutually exclusive scientific theory. Participants (104 younger adults, M_{age} = 19.6, and 48 older adults, M_{age} = 65.1) were asked to verify two types of scientific statements as quickly as possible: those that are consistent with intuition (e.g., "the moon revolves around the Earth") and those that involve the same conceptual relations but are inconsistent with intuition (e.g., "the Earth revolves around the sun"). Older adults were as accurate as younger adults at verifying both types of statements, but the lag in response times between intuition-consistent and intuition-inconsistent statements was significantly larger for older adults than for younger adults. This lag persisted even among professional scientists. Overall, these results suggest that the scientific literacy needed to engage with topics of global importance may be constrained by patterns of reasoning that emerge in childhood but persist long thereafter.

Keywords: Conceptual development; Explanatory coexistence; Naïve theories; Scientific knowledge; Science education; Speeded reasoning

1. Introduction

Being an informed citizen in today's highly technological world requires a fair amount of scientific literacy. Engaging with debates on how to curtail climate change, for instance, requires an understanding of geological systems and how those systems are

Correspondence should be sent to Andrew Shtulman, Department of Psychology, Occidental College, 1600 Campus Road, Los Angeles, CA 90041. E-mail: shtulman@oxy.edu

affected by human carbon emissions. Engaging with debates on how to regulate antibiotics requires an understanding of immune systems and how those systems are affected by increased selection pressure on bacteria. And engaging with debates on how to manage invasive species requires an understanding of ecological systems and how those systems are affected by increased competition for resources. Many other issues of public import—cloning, vaccination, pesticides, organ donation, nuclear power, genetically modified foods—require high levels of scientific knowledge as well.

Does the average adult possess that knowledge? Three decades of research in cognitive development and science education suggest not. Researchers in these fields have discovered that learning science is a two-fold process: Students must not only learn unfamiliar concepts absent from everyday discourse, but they must also un-learn concepts acquired earlier in development for making sense of those same phenomena. In other words, students enter the science classroom with rich, pre-instructional theories—termed "folk theories," "naïve theories," or "intuitive theories"—that typically interfere with learning a more accurate, scientific theory of the same domain (Carey, 2009; Vosniadou, 1994). For instance, students charged with learning a kinetic theory of heat must un-learn a substance-based theory in which heat is construed as an immaterial substance that flows in and out of objects and can be trapped or contained (Reiner, Slotta, Chi, & Resnick, 2000). Students charged with learning a selection-based theory of evolution must un-learn a need-based theory in which evolution is construed as a process that guarantees organisms the traits they need in order to survive (Shtulman, 2006). And students charged with learning an inertial theory of mechanics must un-learn an "impetus"-based theory in which objects are assumed to move only if imparted an internal force, or impetus, and will remain in motion until that impetus dissipates (McCloskey, 1983).

Further complicating matters, recent research suggests that intuitive theories not only interfere with the *acquisition* of scientific theories but may also interfere with the *operation* of those theories many years past their acquisition. Recent research has shown that adults exhibit cognitive conflict when retrieving scientific information that contradicts the intuitive theories they had presumably abandoned as children. One of the best studied cases of this phenomenon is conceptual development in the domain of living things. Beginning with Piaget (1929), psychologists have long observed that young children conflate *life* with *animacy*. Not only do young children attribute life to animate, yet non-living, entities like the sun and the wind, but they also deny life to living, yet seemingly inanimate, entities like flowers and trees. By age 8, this pattern of attributions is typically replaced by a more biologically informed pattern in which life is now identified with metabolic processes (e.g., eating, breathing, growing) rather than mobility (Hatano & Inagaki, 1994). Nevertheless, the adult-like pattern gives way to the child-like pattern when adults are tested under speeded conditions (Babai, Sekal, & Stavy, 2010; Goldberg & Thompson-Schill, 2009). That is, when adults asked to make living/non-living judgments as quickly as possible, they make those judgments more slowly and less accurately for plants (e.g., orchids, elms) as compared to animals (e.g., pigs, sharks) and for animate non-living entities (e.g., comets, rivers) as compared to inanimate ones (e.g., brooms, towels).

Animistic intuitions reemerge not only when adults are placed under speeded conditions, but also when they sustain permanent cognitive impairments, like those produced by Alzheimer's disease. Zaitchik and Solomon (2008) asked Alzheimer's patients what it means for something to be alive and found that these patients were more likely to cite motion as a prerequisite for life than to cite any truly biological properties (e.g., eating, breathing, growing). Healthy age-matched controls, on the other hand, tended to cite only the latter. When Alzheimer's patients were asked to provide examples of living things, they almost always mentioned animals but rarely mentioned plants; age-matched controls tended to mention both animals and plants. And when Alzheimer's patients were asked to judge the life status of entities presented to them, they tended to err in the same ways as children, judging the sun and the wind as alive, but judging flowers and trees as not alive; age-matched controls continued to provide a biologically informed pattern of judgments. Errors by Alzheimer's patients do not appear to derive from a more general deficit in cognitive functioning, as these patients show no impairments on tasks that impose the same information-processing demands but assess physical reasoning rather than biological reasoning (Zaitchik & Solomon, 2009).

Similar findings have been documented in the domain of teleology, or the perception of design in nature. Previous research has shown that children are more "promiscuous" with their teleological explanations than adults are (Kelemen, 1999). Although both children and adults will provide teleological explanations for human artifacts (e.g., "pencils are for writing") and for biological parts (e.g., "ears are for hearing"), only children will provide teleological explanations for whole organisms (e.g., "birds are for flying") and for naturally occurring events (e.g., "it rains so that flowers can drink"). Children become more selective in their teleological explanations by early adolescence, but that selectivity is tenuous. When college-educated adults are asked to judge the acceptability of teleological explanations under speeded conditions, they tend to accept unwarranted explanations, like "ferns grow in forests because they provide ground shade" and "the sun radiates heat because warmth nurtures life," which they tend not to accept under un-speeded conditions and presumably have not accepted under such conditions for many years (Kelemen, Rottman, & Seston, 2013).

Furthermore, just as Alzheimer's patients endorse animistic conceptions of life under normal (non-speeded) conditions, they endorse teleological conceptions of nature under normal conditions as well. Lombrozo, Kelemen, and Zaitchik (2007) provided Alzheimer's patients with both mechanistic and teleological explanations for a variety of natural phenomena, some of which warranted a teleological explanation (e.g., eyes exist "so that people and animals can see") and some of which did not (e.g., rain exists "so that plants and animals have water for drinking and growing"). Compared to healthy elderly adults, Alzheimer's patients were more likely to judge unwarranted teleological explanations as acceptable. They were also more likely to judge those explanations as *preferable* to mechanistic ones. These findings suggest that teleology, like animism, is a deep-seated form of intuition that can be suppressed by a more scientific worldview but cannot be eradicated altogether.

Tensions between science and intuition have been documented not only at the level of behavior, but also at the level of the brain. Dunbar, Fugelsang, and Stein (2007) used fMRI to determine whether college-educated adults exhibit different patterns of brain activity when watching motion displays that were either consistent or inconsistent with the laws of physics. The physics-consistent displays depicted two balls of unequal size falling to the ground at the same rate; the physics-inconsistent displays depicted the larger ball falling to the ground more quickly than the smaller ball. Dunbar et al. found that, among participants who judged the physics-consistent displays as natural and the physics-inconsistent displays as unnatural, watching those displays increased activation in the anterior cingulate cortex, an area of the brain associated with error detection and conflict monitoring. That is, participants who exhibited no behavioral evidence of holding the misconception “heavier objects fall faster than lighter ones” still exhibited neural evidence of holding that misconception insofar that their brains appeared to be detecting and inhibiting contradictory beliefs. Similar results have been documented in the domain of electricity: Physics experts show increased activation in the anterior cingulate cortex, among other areas associated with conflict monitoring, when evaluating electric circuits that are intuitively correct but physically impossible (Masson, Potvin, Riopel, & Foisy, 2014; Potvin, Turmel, & Masson, 2014).

The studies reviewed thus far tracked the persistence of a single misconception—that is, the misconception that life is synonymous with animacy, the misconception that everything in nature exists for a purpose, and the misconception that heavier objects fall faster than lighter objects. The focus on one, and only one, misconception was necessitated by the type of judgment participants were asked to make, such as a living/non-living judgment (Goldberg & Thompson-Schill, 2009) or a warranted/unwarranted judgment (Kelemen et al., 2013). Shtulman and Valcarcel (2012) extended this paradigm to many more misconceptions (50 in total) by asking participants to make true/false judgments instead. More specifically, they asked participants to verify, as quickly as possible, two types of statements: statements whose truth value is the same on both intuitive and scientific theories of a domain (e.g., “the moon revolves around the Earth,” “genes that code for eye color can be found in the eye”) and statements involving the same predicates but whose truth value differs across intuitive and scientific theories (e.g., “the Earth revolves around the sun,” “genes that code for eye color can be found in the liver”). The logic behind this design is that if intuitive theories survive the acquisition of a mutually incompatible scientific theory, then the latter type of statement should cause greater cognitive conflict than the former, resulting in (a) slower verifications and (b) less accurate verifications.

Using this method, Shtulman and Valcarcel (2012) documented evidence of long-term conflict between science and intuition in 10 domains of knowledge: astronomy, evolution, fractions, genetics, germs, matter, mechanics, physiology, thermodynamics, and waves. What is most notable about these findings is their robustness. Shtulman and Valcarcel probed for conflict between science and intuition with respect to five concepts in each of 10 domains and observed such conflict for the vast majority of them (86%). They also observed conflict both for statements that are scientifically true but intuitively false (e.g., “air is composed of matter,” “humans are descended from sea-dwelling creatures”) and for statements that are scientifically false but intuitively true (e.g., “fire is composed of matter," "humans are descended from chimpanzees"), indicating that conflict arose both for statements that underextended scientific principles (e.g., failing to classify air as matter) and for statements that overextended scientific principles (e.g., classifying fire as matter). Furthermore, the participants in Shtulman and Valcarcel's study had taken more science courses than the average American—around three college-level courses, plus 4 to 6 years of middle and high school courses—and virtually all showed the effect. The robustness of this phenomenon across domains, concepts, statements, and participants suggests that it reflects more than just a handful of stubborn misconceptions. Rather, it appears to reflect a fundamental property of science learning, namely, that intuition can be *overridden* but not *overwritten*.

Here, we adopt Shtulman and Valcarcel's (2012) method to assess the resilience of intuitive theories across two dimensions that were held constant in the original study: age and expertise. Whereas participants in the original study were college undergraduates of approximately the same age (18–22) and with approximately the same amount of science expertise, participants in the present study were adults with at least 40 additional years of life experience (M_{age} = 65.1), a quarter of whom were professional scientists. The speed and accuracy with which older adults verified scientific statements were compared to the speed and accuracy with which younger adults verified the same statements to determine whether conflict between science and intuition diminishes with age and/or education.

There are at least two reasons to expect that older adults should *not* show the effect of interest (i.e., a lag in response times between intuition-consistent and intuition-inconsistent statements). First, older adults would have learned the relevant scientific theories much earlier in life, affording more time for their intuitive theories to fade in strength or relevance. Second, older adults would have had considerably more opportunity to use their scientific knowledge outside the classroom, allowing for greater integration and consolidation of that knowledge with preexisting beliefs. The latter consideration applies even more forcefully to professional scientists, who would not only have consolidated their knowledge, but would also have consulted that knowledge on a near-daily basis. To be certain, scientists are more proficient than non-scientists at using scientific knowledge to encode domain-relevant information (Feil & Mestre, 2010), solve domain-relevant problems (Chi, Feltovich, & Glaser, 1981), and make domain-relevant decisions (Shanteau, 1992). The question under investigation, however, is whether scientists are more efficient at retrieving that knowledge when it conflicts with an earlier-acquired intuitive theory. The results presented below suggest they are not.

2. Method

2.1. Participants

Two groups of participants were recruited for this study: 104 younger adults, recruited from introductory psychology courses at Occidental College and compensated with extra credit, and 48 older adults, recruited from either the local community ($n = 27$) or the faculty at Occidental College ($n = 21$) and compensated monetarily. The younger adults reported having taken an average of 3.6 college-level math and science courses prior to the study ($SD = 3.3$, range = 0–15) and averaged 19.6 years in age ($SD = 1.2$, range = 18–22). The older adults reported having taken an average of 6.5 college-level math and science courses ($SD = 5.9$, range = 0–21) and averaged 65.1 years in age ($SD = 11.1$, range = 50–87). Approximately half of the older adults were women, and approximately a quarter of the younger adults were women. Preliminary analyses revealed that gender was not associated with either response accuracy or response latency and was not therefore included as a variable in subsequent analyses.

All older adults were, to our knowledge, neurologically healthy. Those recruited from the local community occupied a variety of professions—accountant, author, writer, therapist—whereas those recruited from the Occidental College faculty were either humanities professors ($n = 11$) or science professors ($n = 10$). The humanities professors came from the departments of Asian Studies, Economics, English, History, Philosophy, and Spanish; the science professors came from the departments of Biology, Chemistry, Geology, Kinesiology, Psychology, and Physics. Not surprisingly, science professors reported having taken significantly more college-level science and math courses than did humanities professors ($M = 15.9$ vs. $M = 4.2$, $t(19) = 8.47$, $p < .001$). Humanities professors, on the other hand, reported having taken an equivalent number to the non-professors ($M = 4.2$ vs. $M = 3.9$, $t(36) < 1$). All three groups were of approximately the same age: science professors, $M = 64.7$, $SD = 4.5$; humanities professors, $M = 58.7$, $SD = 11.9$; non-professors, $M = 67.9$, $SD = 11.7$; $F(2, 45) = 2.85$, *ns*.

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. All domains and concepts are displayed in

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, gene 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, thermal expansion, temperature
Waves	Light, color, sound, wave propagation, reflection

Table 1. A quarter of the statements were true on both intuitive and scientific theories of the domain ("rocks are composed of matter"), a quarter were false on both intuitive and scientific theories ("numbers are composed of matter"), a quarter were true on intuitive theories but false on scientific theories ("fire is composed of matter"), and a quarter were false on intuitive theories but true on scientific theories ("air is composed of matter"). These types of statements will henceforth be referred to as "TT," "FF," "TF," and "FT," respectively. TT and FF statements are both intuition-consistent, whereas TF and FT statements are both intuition-inconsistent.

Sample statements are shown in Table 2. All materials were taken from Shtulman and Valcarcel (2012), who derived them from prior research on the nature of the conceptual change that occurs in each domain. For the domain of astronomy, the statements were designed to assess the change from a geocentric model of the solar system to a heliocentric model (Vosniadou & Brewer, 1994); for evolution, the change from a need-based theory of species adaptation to a selection-based theory (Shtulman & Calabi, 2012); for fractions, the change from an integer-based model of rational number to a division-based model (Siegler, Fazio, Bailey, & Zhou, 2013); for genetics, the change from a trait-based theory of inheritance to a chromosomal theory (Duncan, Rogat, & Yarden, 2009); for germs, the change from a behavioral theory of illness to a microbial theory (Au et al.,

Table 2
Sample intuition-consistent (TT, FF) and intuition-inconsistent (TF, FT) statements

Domain	Type	Statements
Germs	TT	Being sneezed on can make a person sick
	TF	Being cold can make a person sick
	FT	Being depressed can make a person sick
	FF	Being happy can make a person sick
Matter	TT	A log can be cut in half
	TF	A shadow can be cut in half
	FT	A grain of sand can be cut in half
	FF	An idea can be cut in half
Evolution	TT	Humans are more closely related to apes than monkeys
	TF	Apes are more closely related to monkeys than humans
	FT	Whales are more closely related to humans than fish
	FF	Whales are more closely related to plants than fish
Thermodynamics	TT	Heat increases an object's temperature
	TF	Heat increases an object's weight
	FT	Heat increases an object's size
	FF	Heat increases an object's color
Waves	TT	Sound travels through air
	TF	Sound travels through a vacuum
	FT	Sound travels through metal
	FF	Sound travels through foam

Note. TT = scientifically and intuitively true, TF = scientifically true but intuitively false, FT = intuitively false but scientifically true, FF = scientifically and intuitively false.

2008); for matter, the change from a tactile theory of material substances to a particulate theory (Smith, 2007); for mechanics, the change from an impetus theory of motion to an inertial theory (McCloskey, 1983); for physiology, the change from a psychological theory of bodily functions to a vitalist theory (Johnson & Carey, 1998); for thermodynamics, the change from a substance-based theory of heat to a kinetic theory (Reiner et al., 2000); and for waves, the change from a substance-based theory of light and sound to a frequency-based theory (Mazens & Lautrey, 2003).

The materials were counterbalanced in three respects. First, there were an equal number of objectively true and objectively false statements per domain, discouraging participants from adopting a response bias. Second, the average number of words per statement was held constant across statements and domains, give or take a few words. Third, the linguistic complexity of the statements was held constant across the four statements designed to probe any given concept so that simpler statements (e.g., "[entity] is composed of matter") were represented as often as more complex statements (e.g., "[material$_1$] is denser than [material$_2$]") among each stimulus category (TT statements, FF statements, TF statements, and FT statements). Note that, by equating the logical complexity of each statement in a four-statement grouping, those statements differed only in content (i.e., the content of the objects to which the target conceptual relation was applied). Our task was thus qualitatively different from those used to assess belief bias and other types of deductive errors (e.g., Evans, Barston, & Pollard, 1983) as those tasks purposely pit content against logic. The full list of 200 statements is available upon request.

2.3. Procedure

Stimuli were presented to participants with MediaLab v1.21 software (Empirisoft, New York, NY, USA), which recorded the speed and accuracy of their statement-verification judgments. The mean response time across items and across subjects was 3.67 s for the younger adults and 5.56 s for the older adults, and all response times that fell more than 2 *SD* beyond the means for each group were eliminated from the data set. Statements from the same domain 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, as was the ordering of the blocks themselves.

3. Results

Effects of age (younger vs. older adults) on the speed and accuracy of participants' statement verifications are presented first, followed by effects of expertise (science professors vs. other older adults). Only correct responses were retained for the analysis of response latency, as correct responses provide the cleanest test of the hypotheses of interest. Nevertheless, the same effects were observed when incorrect responses were included as well.

3.1. Effects of age

3.1.1. Response accuracy

Fig. 1A shows the mean proportion of intuition-consistent statements and intuition-inconsistent statements correctly verified by younger and older participants, averaged across domain and across the statements' truth value. (Effects of truth value are analyzed

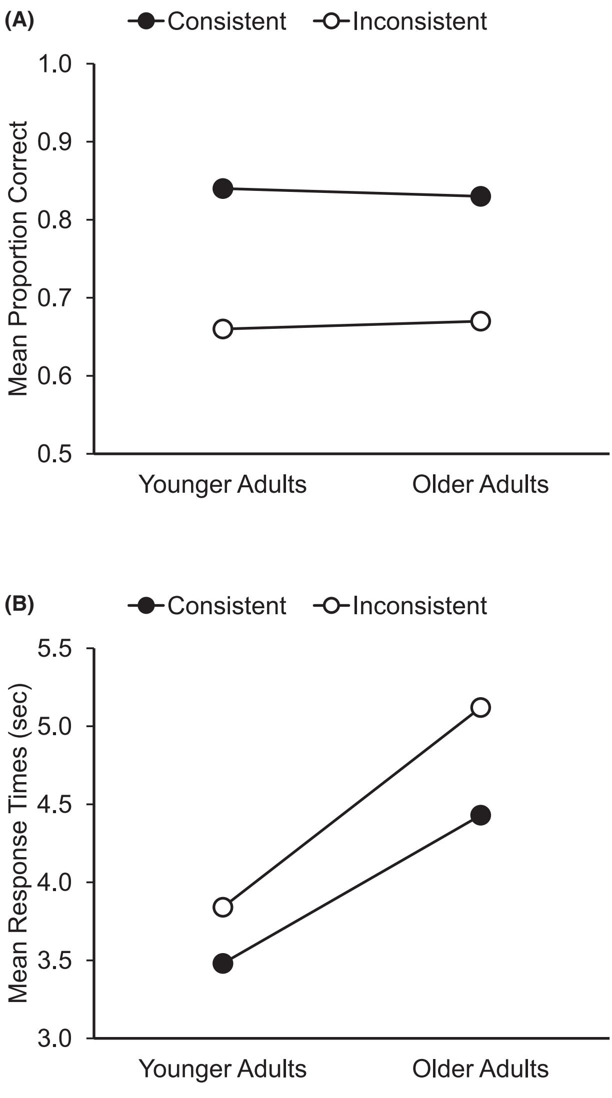

Fig. 1. Mean proportion of correct verifications (A) and mean response times (B) by statement type (intuition-consistent vs. intuition-inconsistent) and age group (younger vs. older adults); all $SE < 0.02$ for (A) and all $SE < 0.025$ for (B).

below.) We submitted those data to a repeated measures analysis of variance (ANOVA) in which statement type (intuition-consistent vs. intuition-inconsistent) was treated as a within-participants variable and age group (younger adults vs. older adults) was treated as a between-participants variable. This analysis revealed a significant effect of statement type ($F(1, 150) = 786.58$, $p < .001$), but no effect of age ($F(1, 150) < 1$). That is, participants in both age groups verified intuition-consistent statements more accurately than they verified intuition-inconsistent statements, but neither group was more accurate, on the whole, than the other.

There was, however, a significant interaction between statement type and age ($F(1, 150) = 3.94$, $p < .05$) such that the difference in correct verifications for intuition-consistent and intuition-inconsistent statements was slightly more pronounced for younger participants (0.18) than for older participants (0.16). That difference withstanding, younger and older participants performed comparably with respect to the statements' consistency with intuition. They also performed comparably with respect to the statements' truth value, judging objectively true statements as "true" (younger participants: $M = 0.81$, $SD = 0.10$; older participants: $M = 0.80$, $SD = 0.11$) and objectively false statements as "false" (younger participants: $M = 0.69$, $SD = 0.13$; older participants: $M = 0.70$, $SD = 0.14$) at statistically equivalent rates.

3.1.2. Response latency

Mean response times for intuition-consistent and intuition-inconsistent statements are displayed as a function of age group in Fig. 1B. These data were analyzed with repeated measures ANOVAS of the same type as that described earlier. Participants in both age groups verified inconsistent statements significantly slower than they verified consistent statements ($F(1, 150) = 287.48$, $p < .001$), and older adults verified both types of statements significantly slower than younger adults did ($F(1, 150) = 48.33$, $p < .001$). A significant interaction between statement type and age group ($F(1, 150) = 28.20$, $p < .001$) indicated that the lag in response times between consistent and inconsistent statements was not equivalent for younger and older adults. For younger adults that lag averaged 0.36 s, whereas for older adults that lag averaged 0.69 s.

To assess the consistency of these effects, we repeated our analyses separately for each domain. Response times are displayed as a function of domain, age group, and statement type in Table 3. In all domains, participants in both age groups verified inconsistent statements more slowly than they verified consistent statements (all Fs > 4.8, all ps < .05), and older participants verified both types of statements significantly slower than younger participants did (all Fs > 23.5, all ps < .001). The interaction between statement type and age was significant in only four domains (evolution: $F(1, 150) = 10.01$, $p < .01$; fractions: $F(1, 150) > 75.04$, $p < .001$; matter: $F(1, 150) > 4.13$, $p < .05$; physiology: $F(1, 150) > 6.32$, $p < .05$). Nevertheless, older participants exhibited a larger lag in response times between intuition-consistent and intuition-inconsistent statements than that exhibited by younger participants in all domains but one (genetics).

Finally, we explored whether the effect of statement type on response latency held both for statements that were objectively true ("fish are alive" vs. "coral is alive") and

Table 3
Mean response times (seconds) for intuition-consistent and intuition-inconsistent statements in each domain and age group, as well as response lags between the two types of statements (intuition-inconsistent minus intuition-consistent) and differences in response lags between the two age groups (older minus younger)

Domain	Age group	Inconsistent	Consistent	Response Lag	Difference
Astronomy	Younger	3.86	3.68	0.18	0.11
	Older	5.08	4.79	0.29	
Evolution	Younger	4.32	3.96	0.36	0.60
	Older	5.68	4.72	0.96	
Fractions	Younger	4.54	3.70	0.84	1.27
	Older	6.77	4.66	2.11	
Genetics	Younger	3.92	3.56	0.36	−0.24
	Older	4.75	4.63	0.12	
Germs	Younger	3.07	3.00	0.07	0.16
	Older	3.91	3.68	0.23	
Matter	Younger	3.70	3.43	0.27	0.36
	Older	4.96	4.33	0.63	
Mechanics	Younger	4.04	3.94	0.10	0.25
	Older	5.26	4.91	0.35	
Physiology	Younger	3.11	2.62	0.49	0.35
	Older	4.27	3.43	0.84	
Thermodynamics	Younger	4.30	3.78	0.52	0.22
	Older	5.64	4.90	0.74	
Waves	Younger	3.90	3.44	0.46	0.21
	Older	5.15	4.48	0.67	

for statements that were objectively false ("rocks are alive" vs. "the sun is alive"). To do so, we computed separate means for the two types of intuition-consistent statements (TT and FF statements) and the two types of intuition-inconsistent statements (FT and TF statements). Paired-samples t tests revealed that participants in both age groups verified intuition-inconsistent statements significantly slower than intuition-consistent statements regardless of whether those statements were objectively true (younger participants: $t(103)= 12.90$, $p < .001$; older participants: $t(47)= 8.87$, $p < .001$) or objectively false (younger participants: $t(103)= 6.70$, $p < .001$; older participants: $t(47) = 6.39$, $p < .001$).

That said, the difference in response times between the two types of statements was larger for true statements ($M = 0.55$ s) than for false statements ($M = 0.40$ s), and a repeated measures ANOVA confirmed that the interaction between statement type (intuition-consistent vs. intuition-inconsistent) and statement truth value (true vs. false) was significant ($F(1, 151) = 9.29$, $p < .01$). False statements were apparently less reliable than true statements at eliciting conflict between science and intuition, possibly because false statements were verified more slowly overall (M seconds for false statements = 4.17, M seconds for true statements = 3.88, $F(1, 151) = 58.65$, $p < .001$). Nevertheless, the size of the interaction between statement type and truth value (partial $\eta^2 = 0.06$) was substantially smaller than the size of the main effect of statement type (partial $\eta^2 = 0.58$), indicating that a statement's truth value played a minor role in how quickly it was verified compared to whether or not that statement was consistent with intuition.

3.2. *Effects of expertise*

3.2.1. *Response accuracy*

The mean proportion of intuition-consistent and intuition-inconsistent statements correctly verified by the three groups of older adults is displayed in Fig. 2A. We submitted those data to a repeated measures ANOVA in which statement type (intuition-consistent vs. intuition-inconsistent) was treated as a within-participants factor and occupation

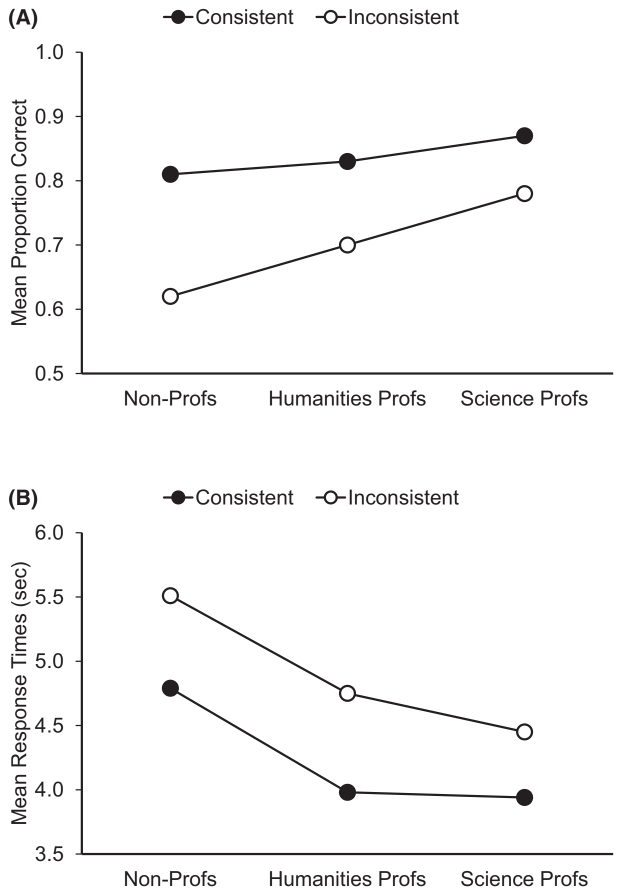

Fig. 2. Mean proportion of correct verifications (A) and mean response times (B) as a function of statement type (intuition-consistent vs. intuition-inconsistent) and occupation (non-professors, humanities professors, science professors) for the older adults; all $SE < 0.03$ for (A) and all $SE < 0.035$ for (B).

(science professors vs. humanities professors vs. non-professors) was treated as a between-participants factor. Correct verifications varied not only by statement type ($F(1, 45) = 149.38$, $p < .001$), but also by occupation ($F(2, 45) = 13.76$, $p < .001$) and by the interaction between statement type and occupation ($F(2, 45) = 9.28$, $p < .001$).

We explored the effect of occupation with linear contrast analyses in which scores for non-professors were weighted "−1," scores for humanities professors "0," and scores for science professors "1." These analyses revealed that correct verifications increased monotonically from non-professors to humanities professors to science professors, both for consistent statements ($F(1, 47) = 8.92$, $p < .01$) and for inconsistent statements ($F(1, 47) = 30.58$, $p < .001$). And Bonferroni comparisons revealed that science professors performed significantly better than the other two groups for both types of statements ($p < .05$). The superior performance of science professors helps validate the task itself, particularly the difference in performance between science professors and humanities professors. Both types of participants had attained similar levels of education and had engaged in similar kinds of professional activities throughout their careers, yet only science professors possessed the requisite content knowledge for verifying the vast majority of statements correctly. Science professors also exhibited a smaller discrepancy in accuracy between intuition-consistent and intuition-inconsistent statements than did the other two groups (science professors, 9%; humanities professors, 13%; non-professors, 20%), as revealed by a significant interaction between statement type and occupation noted above. This finding indicates that a statement's consistency with intuition matters less the more one knows about science, at least with respect to accuracy.

3.2.2. Response latency

Mean response times for intuition-consistent and intuition-inconsistent statements are displayed as a function of occupation in Fig. 2B. Similar to the accuracy data, the latency data varied significantly by both statement type ($F(1 ,45) = 76.96$, $p < .001$) and occupation ($F(2, 45) = 3.72$, $p < .05$). That is, older adults, on the whole, verified intuition-inconsistent statements significantly slower than they verified intuition-consistent statements, and older adults in some occupations verified both types of statements significantly slower than did those in other occupations.

To explore the latter effect, we submitted response times to linear contrasts of the same form described earlier. These analyses revealed that response times decreased monotonically from non-professors to humanities professors to science professors, both for intuition-consistent statements ($F(1, 47) = 4.66$, $p < .05$) and for intuition-inconsistent statements ($F(1, 47) = 5.47$, $p < .05$). Nevertheless, there was no interaction between statement type and occupation ($F(2, 45) = 1.57$, *ns*), as the lag in response times between intuition-consistent and intuition-inconsistent statements was comparable across groups: non-professors (0.72 s), humanities professors (0.82 s), science professors (0.46 s). Thus, although science professors outperformed their age-matched peers in terms of accuracy, they performed similarly to their peers in terms of speed.

4. Discussion

Scientific knowledge is critical for making informed decisions about many globally important issues, but scientific knowledge does not come easily. Learning science requires revising and restructuring earlier-acquired, intuitive theories that typically contradict scientific theories of the same phenomena. Moreover, recent research suggests that intuitive theories are not replaced by scientific theories but coexist with them, obstructing the retrieval of scientific theories by providing an alternative explanation for the phenomena at hand that must first be inhibited. In the present study, we explored the robustness of this finding across age and expertise. We found that, regardless of how old our participants were or how much science expertise they had acquired, they verified scientific statements that were inconsistent with intuition significantly slower than they verified scientific statements that were consistent with intuition. This pattern is remarkable given that, in many domains (e.g., germs, matter, physiology), participants acquired their relevant scientific knowledge as children. Yet the lag in response times between intuition-consistent and intuition-inconsistent statements did not diminish with age; if anything, it increased. And this lag was evident not only among adults in non-scientific occupations, but also among professional scientists with three to four decades of career experience.

Our finding that older adults are no more immune to the conflict between science and intuition than are younger adults accords well with prior research on scientific reasoning in Alzheimer's patients (Lombrozo et al., 2007; Zaitchik & Solomon, 2008). Although the cognitive impairments wrought by Alzheimer's disease appear to liberate pre-scientific intuitions at an explicit level, our findings indicate that the same intuitions persist at an implicit level among neurologically healthy adults of a similar age. Moreover, our finding that scientists are no more immune to the conflict between science and intuition than are non-scientists accords well with prior research involving other expert populations. Under speeded conditions, professional biologists reveal animistic intuitions of the same sort revealed by non-biologists, for example, that comets are alive and that orchids are not alive (Goldberg & Thompson-Schill, 2009), and professional physicists endorse unwarranted teleological explanations of the same sort endorsed by non-physicists, for example, that "moss forms around rocks in order to stop soil erosion" and that "the sun makes light so that plants can photosynthesize" (Kelemen et al., 2013). Our results suggest that these concept-specific instances of cognitive conflict are symptomatic of a more general pattern, one that encompasses multiple concepts in multiple domains.

Thus far, we have interpreted the observed effects as evidence of conflict between intuitive theories and scientific theories, but it is also possible that these effects reflect a more localized conflict—that is, a conflict between isolated beliefs. In other words, the conflict observed for statements like "ice has heat" (which is scientifically true but intuitively false) or "coats produce heat" (which is scientifically false but intuitively true) may reflect discrepancies not in our *theories* of heat but in our *beliefs* about heat, some of which are consistent with science and some of which are not. This interpretation is consistent with a "knowledge-in-pieces" view of folk beliefs that treats such beliefs as fragmented and incoherent (DiSessa, 1993; Hammer, 1996). Whether our folk beliefs are best characterized as self-consistent theories or as piecemeal conglomerations is outside the scope of the present study (see instead Vosniadou, 2010). Still, there are at least two findings that favor a theory-based interpretation of the data over a knowledge-in-pieces interpretation.

First, participants demonstrated the effect of interest across most to all concepts within any given domain of knowledge. That is, they exhibited slower response times for intuition-inconsistent statements relative to intuition-consistent statements for four of the five concepts in the domains of astronomy, germs, genetics, matter, mechanics, thermodynamics, and waves, and for all five concepts in the domains of evolution, fractions, and physiology. The consistency of this effect across multiple concepts within the same domain suggests that those concepts are interrelated (as has been shown in many other studies as well, e.g., Au et al., 2008; Mazens & Lautrey, 2003; McCloskey, 1983; Shtulman, 2006; Smith, 2007). Second, even scientists exhibited the effect, and it would be a stretch to claim that scientists' beliefs about natural phenomena are fragmented and incoherent. Indeed, a corollary of the knowledge-in-pieces view of folk beliefs is that science learning serves to connect and unify those beliefs, yielding a knowledge base increasingly devoid of internal inconsistencies. Yet such inconsistencies do not appear to go away, and only a theory-based view of the observed effects provides a means of explaining them.

In sum, the present results demonstrate that intuitive theories are highly resilient across age, occupation, and domain. They do not tell us, however, *why* intuitive theories are so resilient or *how* intuitive theories affect our reasoning outside the laboratory. Below we address each question in turn, speculating on how the observed effects alter our understanding of the acquisition and representation of scientific knowledge.

5. Why are intuitive theories so resilient?

One explanation for why intuitive theories seem to persist across the lifespan is that they may be represented in the brain in a cognitively impenetrable format, similar to the seemingly impenetrable representations of language (Coltheart, 1999) and vision (Pylyshyn, 1999). In vision, for instance, we can be well aware that our eyes deceive us when viewing the Muller-Lyer illusion or the Ponzo illusion, but we perceive the illusion nonetheless. The visual biases that give rise to such illusions constitute a stable backdrop against which all new visual information is interpreted, and those biases operate even when we are aware of their fallibility. Intuitive theories might be represented in the brain in a similar fashion, though this explanation begs the question as to what constitutes an intuitive theory and why such representations are impervious to revision. Part of the appeal of describing folk beliefs as "theories," after all, is that such beliefs are presumably open to revision (Gopnik & Wellman, 2012).

Another explanation for why intuitive theories seem to persist across the lifespan is that intuitive theories are actively reinforced by how we talk about natural phenomena in everyday discourse and how we perceive natural phenomena in everyday situations. Much of our colloquial language seems to be predicated on intuitive conceptions. The terms "sunrise" and "sunset," for instance, imply that day and night are caused by movements of the sun rather than movements of the earth. More accurate terms would be "sun accretion" and "sun occlusion." Likewise, the terms "warm coat" and "cold wind" imply that heat is an intrinsic property of objects rather than something that is transferred across physical systems. More accurate terms would be "insulating coat" and "disequilibrating wind."

Our perceptual experience is no less misleading. Coats *feel* as if they produce heat, and the sun *looks* as if it moves across the sky. Recent models of theory change have begun to emphasize the role of a theory's cognitive utility in catalyzing that process (e.g., Gopnik & Wellman, 2012; Ohlsson, 2009), and these models predict that an intuitive theory should be maintained alongside a scientific theory so long as that theory provides a sufficiently useful interpretation of the data at hand—that is, the data of everyday linguistic practices and the data of everyday perceptual experiences. The challenge to those who would explain the resilience of intuitive theories in terms of their cognitive utility is to clarify the dimensions along which cognitive utility is calculated and the situations in which intuitive theories trump scientific theories along those particular dimensions.

6. How do intuitive theories affect scientific reasoning?

In the present study, participants of all levels of science expertise exhibited cognitive conflict when verifying intuition-inconsistent statements. This conflict demonstrates that all participants continued to hold onto their intuitive theories (at some level of representation), but it does not tell us how those theories influence scientific reasoning "in the wild." We suspect that intuitive theories influence scientific reasoning in different ways at different points in the development of scientific expertise. Early on, intuitive theories undoubtedly interfere with the acquisition of scientific theories. Such interference has been documented previously, in many domains using many methods (see Vosniadou, 2010), but the discovery that intuitive theories are never truly overwritten by scientific theories suggests that this interference is more pervasive and more pernicious than originally thought. Students must contend with their intuitive theories not only at the outset of learning, but also throughout the process of learning.

The resilience of intuitive theories may also help to explain why *knowledge* of science has often been found to be unrelated to *acceptance* of science—for example, acceptance that humans have caused climate change (Kahan et al., 2012) or that humans evolved from non-human ancestors (Sinatra, Southerland, McConaughy, & Demastes, 2003). Intuitive theories likely cause conflict or confusion when attempting to engage with scientific issues that run counter to those theories, creating opportunities for religious views (Jelen & Lockett, 2014; Shtulman, 2013) or political views (Kahan, Jenkins-Smith, & Braman, 2011) to sway one's opinions on the matter. On the other hand, knowledge of science and acceptance of science are not always unrelated; some studies have found that the former does indeed predict the latter (Ingram & Nelson, 2006; Ranney & Clark, 2015; Rutledge & Warden, 1999; Shtulman & Calabi, 2012). The discrepancy between these studies may be due, in part, to the types of knowledge that they have assessed. Studies that have found a relation between science knowledge and science acceptance have typically assessed participants' *mechanistic* understanding of the relevant science, whereas those that have not found such a relation have typically assessed participants' *factual* understanding. The better we understand the mechanisms of a phenomenon, the more likely we accept that phenomenon as true. Still, future research is needed to determine whether mechanistic knowledge produces this effect on its own or whether it mediates the effect by shielding participants from other, non-scientific considerations, such as religious views, political views, or—most pertinent to the present study—one's own intuitive theories.

While intuitive theories may affect the credence that science novices place on scientific findings, they are unlikely to affect the credence that science experts place on such findings. Science experts, after all, have likely established strong boundaries between their intuitive theories of a domain and their scientific theories of a domain, particularly for their own domains of expertise. Given the right contextual cues, even non-scientists readily partition mutually incompatible pieces of knowledge into separate mental parcels (Lewandowsky, Kalish, & Ngang, 2002; Yang & Lewandowsky, 2004). Those partitions may break down, however, in the absence of the contextual cues upon which they were first established. For scientists, those cues are likely embedded in the practices of their trade—that is, running experiments, analyzing data, writing up findings, conversing with colleagues. Outside those contexts, scientists may be no more likely to honor the boundary between science and intuition than non-scientists. Indeed, there is evidence that scientists are prone to error when applying familiar principles to unfamiliar problems (Dunbar, 1995), when evaluating familiar information in unfamiliar formats (Eddy, 1982), and when making empirical projections that run counter to prevailing beliefs (Brysse, Oreskes, O'Reilly, & Oppenheimer, 2013). Whether those errors are caused by a breakdown in the partitioning of science and intuition, as opposed to some other cognitive limitation, is a question in need of further research.

7. Conclusion

Open questions aside, our results indicate that intuitive theories persist across the lifespan, influencing scientific reasoning for decades beyond the acquisition of a mutually incompatible scientific theory. The resilience of intuitive theories may be partially responsible for public skepticism toward science, though future research is needed to determine whether and how the implicit representation of intuitive theories affects our explicit attitudes toward science. Still, developing an awareness of those theories may provide some immunity to their sway over attitudes and decisions better informed by science alone.

Acknowledgments

We thank the National Science Foundation for supporting this research with grant DRL-0953384 awarded to Andrew Shtulman. We also thank Victoria Halote, Alexander Levin, Cara Neal, and Linneen Warren for their assistance with data collection.

References

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

Babai, R., Sekal, R., & Stavy, R. (2010). Persistence of the intuitive conception of living things in adolescence. *Journal of Science Education and Technology*, *19*, 20–26.

Brysse, K., Oreskes, N., O'Reilly, J., & Oppenheimer, M. (2013). Climate change prediction: erring on the side of least drama? *Global Environmental Change*, *23*, 327–337.

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

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

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*, 115–120.

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

Dunbar, K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In J. E. Davidson & R. J. Sternberg (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: The MIT Press.

Dunbar, K., Fugelsang, J., & Stein, C. (2007). Do naïve 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.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, UK: Cambridge University Press.

Evans, J. St. B. T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, *11*, 295–306.

Feil, A., & Mestre, J. P. (2010). Change blindness as a means of studying expertise in physics. *Journal of the Learning Sciences*, *19*, 480–505.

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

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108.

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

Hatano, G., & Inagaki, K. (1994). Young children's naïve theory of biology. *Cognition*, *50*, 171–188.

Ingram, E. L., & Nelson, C. E. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. *Journal of Research in Science Teaching*, *43*, 7–24.

Jelen, T. G., & Lockett, L. A. (2014). Religion, partisanship, and attitudes toward science policy. *SAGE Open*, *4*, 1–8. doi:10.1177/2158244013518932.

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

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, *14*, 147–174.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*, 732–735.

Kelemen, D. (1999). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Sciences*, *3*, 461–468.

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*, 1074–1083.

Lewandowsky, S., Kalish, M., & Ngang, S. K. (2002). Simplified learning in complex situations: Knowledge partitioning in function learning. *Journal of Experimental Psychology: General*, *131*, 163–193.

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.

Masson, S., Potvin, P., Riopel, M., & Foisy, L. M. B. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, *8*, 44–55.

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.

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.

Potvin, P., Turmel, E., & Masson, S. (2014). Linking neuroscientific research on decision making to the educational context of novice students assigned to a multiple-choice scientific task involving common misconceptions about electrical circuits. *Frontiers in Human Neuroscience*, *8*, 1–13.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–365.

Ranney, M. A., & Clark, D. (2015). Climate change conceptual change: scientific information can transform attitudes. *Topics in Cognitive Science*, *8*, 49–75.

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

Rutledge, M. L., & Warden, M. A. (1999). The development and validation of the measure of acceptance of the theory of evolution instrument. *School Science and Mathematics*, *99*, 13–18.

Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, *81*, 75–86.

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

Shtulman, A. (2013). Epistemic similarities between students' scientific and supernatural beliefs. *Journal of Educational Psychology*, *105*, 199–212.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 47–65). Cambridge, UK: Oxford University Press.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: the new frontier for theories of numerical development. *Trends in Cognitive Sciences*, *17*, 13–19.

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching*, *40*, 510–528.

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories: using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction*, *25*, 337–398.

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

Vosniadou, S. (Ed.) (2010). *International handbook of research on conceptual change*. New York: Routledge.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, *18*, 123–183.

Yang, L. X., & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1045–1064.

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.