Cognition 243 (2024) 105680

Contents lists available at ScienceDirect

Cognition

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

Tempering the tension between science and intuition☆

Andrew Shtulman[a,*], Andrew G. Young[b]

[a] Department of Psychology, Occidental College, 1600 Campus Road, Los Angeles, CA 90041, USA
[b] Department of Psychology, Northeastern Illinois University, 5500 North St. Louis Avenue, Chicago, IL 60625, USA

ARTICLE INFO

Keywords:
Scientific reasoning
Intuitive theories
Explanatory coexistence
Conceptual change

ABSTRACT

Scientific ideas can be difficult to access if they contradict earlier-developed intuitive theories; counterintuitive scientific statements like “bubbles have weight” are verified more slowly and less accurately than closely-matched intuitive statements like “bricks have weight” (Shtulman & Valcarcel, 2012). Here, we investigate how context and instruction influences this conflict. In Study 1, college undergraduates ($n = 100$) verified scientific statements interspersed with images intended to prime either a scientific interpretation of the statements or an intuitive one. Participants primed with scientific images verified counterintuitive statements more accurately, but no more quickly, than those primed with intuitive images. In Study 2, college undergraduates ($n = 138$) received instruction that affirmed the scientific aspects of the target domain and refuted common misconceptions. Instruction increased the accuracy of participants’ responses to counterintuitive statements but not the speed of their responses. Collectively, these findings indicate that scientific interpretations of a domain can be prioritized over intuitive ones but the conflict between science and intuition cannot be eliminated altogether.

1. Introduction

Does air have weight? Does air take up space? Is air composed of atoms? A chemist would answer “yes,” “yes,” and “yes.” Air is a gas; gases are a form of matter; and all matter has weight, volume, and an atomic structure. Intuitively, though, air seems to be nothing more than empty space. Air can neither be seen nor felt, and it betrays no sign of its particulate nature, striking us as continuous and homogenous.

This tension between thinking of air as matter and thinking of it as empty space is one of many tensions between scientific and intuitive interpretations of the natural world (Carey, 2009; Thagard, 2014). For instance, most adults today believe that the earth orbits the sun, but this idea defies deep-seated intuitions. The earth betrays no evidence of motion yet the sun moves across the sky every day, leading children to believe that the sun orbits the earth (Vosniadou & Brewer, 1994). When we eventually learn that the opposite is true, the earlier belief does not appear to be erased. Adults who are asked to judge, as quickly as possible, whether “the earth orbits the sun” is true or false often respond “false.” And those who respond “true” take longer to make their judgment than to judge closely-matched statements that do not conflict with earlier-acquired beliefs, such as “the moon orbits the earth” (Shtulman & Valcarcel, 2012).

Under time pressure, adults have revealed many other childhood misconceptions, such as that plants are not alive (Goldberg & Thompson-Schill, 2009), that heavy objects fall faster than light ones (Foisy, Potvin, Riopel, & Masson, 2015), that fractions with large denominators are greater than fractions with small denominators (Vamvakoussi, Van Dooren, & Verschaffel, 2012), and that natural kinds, like geysers and earthworms, exist to fulfill a purpose (Kelemen & Rosset, 2009). These misconceptions are not mere factual errors; they are grounded in intuitive theories that carve the world into entities and processes that have no scientific counterpart (Carey, 2009).

Intuitive theories are well-documented among children, who construct these theories prior to learning scientific theories of the same phenomena (Shtulman, 2017). Intuitive theories have long been assumed to be replaced by their scientific counterpart (see Shtulman & Lombrozo, 2016), but statement-verification findings like those reviewed above suggest that intuitive theories remain largely intact. When adults verify scientific statements under time pressure or cognitive load, they are slower and less accurate for statements that conflict

☆ We would like to thank the James S. McDonnell Foundation for supporting this research through an Understanding Human Cognition Scholar Award to Andrew Shtulman. We would also like to thank Giovanna Dieffenbach, Eushrah Hossain, Nicholas Hung, Jasper Laca, Devon Mann, Onyekachi Nwabueze, Blaire Sellers, and Lea Theodorou for their assistance with data collection.

* Corresponding author at: 1600 Campus Road, Los Angeles, CA 90041, USA.
E-mail address: shtulman@oxy.edu (A. Shtulman).

https://doi.org/10.1016/j.cognition.2023.105680
Received 30 April 2023; Received in revised form 3 November 2023; Accepted 22 November 2023
Available online 9 December 2023
0010-0277/

with intuitive theories than for statements that accord with those theories. This pattern has been documented with a wide variety of statements and across a wide variety of domains, including evolution, chemistry, astronomy, geometry, algebra, mechanics, and thermodynamics (Allaire-Duquette et al., 2021; Barlev, Mermelstein, & German, 2017; Merz, Dietsch, & Schneider, 2016; Shtulman & Valcarcel, 2012; Stricker et al., 2021; Vosniadou et al., 2018).

When scientific theories conflict with intuitive theories, the resolution of such conflict appears to require inhibition. Adults who have undergone neuroimaging while answering counterintuitive scientific questions show increased activation in areas of the brain linked to error monitoring and inhibitory control, such as the anterior cingulate cortex and the dorsolateral prefrontal cortex, when they answer those questions correctly (Allaire-Duquette, Bélanger, Grabner, Koschutnig, & Masson, 2019; Masson, Potvin, Riopel, & Foisy, 2014). For instance, judging that a large object will fall at the same rate as a small object activates inhibitory control networks, implying that participants who make this judgment must inhibit the misconception that large objects fall faster than small ones (Foisy et al., 2015). Convergently, individuals who have lost inhibitory control abilities, such as Alzheimer's patients, also lose the ability to prioritize scientific theories over intuitive ones; they default to intuitive theories even when given ample time to respond (Lombrozo, Kelemen, & Zaitchik, 2007; Zaitchik & Solomon, 2008).

Tensions between science and intuition thus appear ubiquitous, but are such tensions intractable? Certainly, we can learn counterintuitive scientific ideas and how to apply them. The key to such learning is addressing the intuitive misconceptions that clash with a scientific understanding of the domain (Chi, 2009; Nersessian, 1989; Vosniadou, 1994). Learning that air has weight, for instance, requires addressing the intuitive misconception that gases are empty space (Smith, 2007). Learning that the earth orbits the sun requires addressing the intuitive misconception that the earth does not move (Vosniadou & Brewer, 1994). Similar patterns have been observed in learning about force (Clement, 1993), energy (Wiser & Amin, 2001), physiology (Slaughter & Lyons, 2003), illness (Au et al., 2008), inheritance (Springer, 1995), and evolution (Shtulman & Calabi, 2012). The most effective way to facilitate conceptual change, or knowledge restructuring at the level of individual concepts, is to help students bridge the gap between intuitive and expert understandings of the domain.

Still, the persistence of intuitive misconceptions suggests that conceptual change is necessary but not sufficient for sound scientific reasoning; additional resources are required to access and deploy counterintuitive scientific ideas. Even professional scientists show signs of cognitive conflict when reasoning about counterintuitive scientific ideas. Although scientists verify such ideas more quickly and more accurately than non-scientists, they still exhibit lags in speed and accuracy for counterintuitive ideas relative to intuitive ones (Allaire-Duquette et al., 2021; Masson et al., 2014). For instance, under speeded conditions, professional biologists are slower and less accurate at verifying that plants are alive relative to animals (Goldberg & Thompson-Schill, 2009), and professional physicists are inclined to accept unwarranted teleological explanations, such as "moss forms around rocks to stop soil erosion" or "the earth has an ozone layer to protect it from UV light" (Kelemen, Rottman, & Seston, 2013). Indeed, Shtulman and Harrington (2016) found that science professors are slower and less accurate at verifying counterintuitive scientific statements (e.g., "air is composed of matter") relative to closely-matched intuitive statements (e.g., "rocks are composed of matter") across ten domains of science.

An important caveat is that the expertise of the scientists who participated in these studies may not have been aligned with the tasks they were asked to complete. In Shtulman and Harrington's (2016) study, for instance, participants were asked to evaluate statements about everything from optics to illness. While scientists were consistently more accurate than non-scientists, the scientists' professional expertise extended to only a subset of those domains. Similar concerns arise for the biologists in Goldberg and Thompson-Schill's (2009) study and the physicists Kelemen et al. (2013) study. A biologist who studies intracellular reactions may ponder the life status of plants no more often than a non-biologist, and a physicist who studies string theory may ponder the origins of moss and ozone no more often than a non-physicist.

For these reasons, we sought a more direct test of the malleability of the conflict between science and intuition. Rather than study individuals who are practiced at accessing counterintuitive scientific ideas, we attempted to manipulate the accessibility of such ideas through context and instruction. We measured accessibility using Shtulman and Valcarcel's (2012) statement-verification task, where conflict between intuitive and scientific interpretations of scientific statements is revealed through decreased accuracy and increased response times. In Study 1, participants completed this task while exposed to images intended to prime a scientific interpretation of the target statements. In Study 2, participants completed the task before and after a tutorial designed to refute intuitive interpretations of the target statements and reinforce scientific ones. In both studies, we sought to reduce the cognitive conflict between science and intuition.

There are at least two ways that priming and training could reduce cognitive conflict. One way is by steering participants toward scientific interpretations of counterintuitive statements before they entertain an intuitive interpretation, thereby bypassing the conflict altogether. Bypassing this conflict would increase participants' accuracy as well as their speed, as participants would no longer need to suppress a competing intuition.

Another way that priming and training could reduce cognitive conflict is by helping participants privilege scientific interpretations of counterintuitive statements over intuitive ones when both interpretations have been activated. Statements like "air has weight" might elicit an intuitive response ("no") alongside a scientific response ("yes") even after participants have been primed or trained to think of weight scientifically, but these manipulations could still dispose participants to choose the latter over the former. This pathway would yield increased accuracy but might not yield increased speed, as participants must still grapple with competing response options before making their final judgment. Priming and training should strengthen the salience of the scientific response relative to the intuitive one, but the competition between them should continue to delay judgments for counterintuitive statements to at least some extent. Indeed, this pathway might be better characterized as helping participants resolve the conflict between science and intuition rather than reducing it outright.

To that end, we modified Shtulman and Valcarcel's (2012) original task by asking participants to verify a wider variety of statements in a smaller number of domains, namely, the domains of life and matter. These domains are foundational to learning higher-order concepts in biology and physics, respectively, and our participants—college undergraduates—could be expected to have acquired a scientific understanding of both domains, as they are covered early and often in the K-12 science curriculum (National Science Teachers Association, 2013). These domains were also ideal for expanding our stimuli, as we could apply a handful of domain-specific predicates (e.g., "has weight") to a large number of domain-specific subjects (e.g., air, bubbles, clouds, dust, foam, fog, smoke, snowflakes), as we discuss below.

In addition to methodological reasons for focusing on life and matter, there are also empirical ones. Several studies have shown that adults harbor misconceptions about life and matter that conflict with the scientific conceptions they have acquired through formal instruction. Life, from a scientific perspective, is a metabolic state—the consumption of energy to further an organism's survival and reproduction—but we initially conceive of life as the capacity for self-directed motion (Hatano & Inagaki, 1994). Accordingly, young children classify moving but nonliving entities, like the sun and the clouds, as alive, and they classify living but nonmoving objects, like plants and trees, as not alive (Carey, 1985; Stavy & Wax, 1989). College undergraduates make the same kinds of mistakes when classifying entities as "alive" or "not alive" as quickly as possible (Goldberg & Thompson-Schill, 2009). They classify plants as not alive, and they classify moving but nonliving entities, like rivers and airplanes, as alive. Elderly adults make these mistakes as well, even without the burden of time pressure (Tardiff, Bascandziev, Sandor, Carey, & Zaitchik, 2017; Zaitchik & Solomon, 2008).

Likewise, matter is initially understood as something that can be seen or touched rather than something composed of atoms (Nakhleh, Samarapungavan, & Saglam, 2005). Young children mistakenly classify substances they cannot perceive, like vapors and gases, as immaterial, and they mistakenly classify forms of energy they can perceive, like lightning and rainbows, as material (Smith, 2007). Adults make the same mistakes if instructed to decide whether something is material or nonmaterial as quickly as possible (Shtulman & Legare, 2020). They also make mistakes when quickly deciding whether an object will sink or float, focusing on its size rather than its material. When shown two balls of equal size, one made of wood and one made of lead, adults judge that the wood ball is more likely to float than the lead one. But when shown a large ball of wood and a small ball of lead, they take reliably longer to make the same judgment (Potvin & Cyr, 2017; Potvin, Masson, Lafortune, & Cyr, 2015).

The domains of life and matter thus provided an ideal opportunity for examining whether, and how, context and instruction can increase the accessibility of counterintuitive ideas relative to intuitive ones. We expected that providing participants with input directly relevant to counterintuitive scientific ideas would increase the accuracy of their verifications, but it was unclear whether it would increase the speed of their verifications as well, given the robust lags in response time observed across concepts, domains, and populations, including professional scientists. If the tension between science and intuition is an inevitable byproduct of science learning, then this tension should emerge whenever we reason about ideas relevant to both, even if we can successfully suppress an intuitive interpretation in favor of a scientific one.

2. Study 1

Our first attempt at modifying the conflict between science and intuition involved priming. Previous research suggests that priming can shift the balance between scientific reasoning and intuitive reasoning in the context of religion. Scientific primes increase the endorsement of scientific explanations for the origin of humans (Tracy, Hart, & Martens, 2011) and the origin of the universe (Preston & Epley, 2009). That is, priming people to consider the explanatory power of science leads them to evaluate scientific ideas, like evolution, more positively than religious alternatives, like creationism. Scientific primes also increase the use of scientific explanations over religious explanations when reasoning about illness or death (Busch, Watson-Jones, & Legare, 2017; Harris & Gimenez, 2005; Lane, Zhu, Evans, & Wellman, 2016; Legare & Gelman, 2008). For instance, medical contexts prime people to think of death as the cessation of bodily functions, whereas religious contexts prime people to think of death as a spiritual transformation (Harris, 2011).

Computational simulations of these effects imply that they are pervasive, reflecting an attempt to establish coherence between incompatible causal principles, such as viruses vs. witchcraft, and the situations that evoke them, such as infectious disease (Friedman & Goldwater, 2023). These effects are limited, however, in that they pit scientific ideas against supernatural ones in populations where people typically endorse both ideas to some degree. Primes that pit scientific ideas against intuitive, yet naturalistic, ideas may hold less sway given that the conflict between scientific theories and intuitive theories is more implicit and perhaps, then, less tractable.

In Study 1, we explored whether priming might help students prioritize scientific ideas over intuitive ones by asking them to verify counterintuitive scientific statements while viewing scientific images relevant to the content of those statements. We predicted that such primes would increase the accuracy of participants' verifications and possibly also their speed. We assessed the effect of scientific primes against the effect of intuitive primes, or images intended to evoke intuitive interpretations of the same statements. Whether intuitive primes would impact participants' verifications was less clear. If intuitive theories are a default mode of reasoning, then priming them may not interfere with scientific reasoning any more than usual (i.e., in the absence of primes).

2.1. Method

Our study employed a 2 × 3 factorial design, where statement type (intuitive vs. counterintuitive) was varied within participants, and prime type (intuitive vs. scientific vs. none) was varied between participants.

Participants. One-hundred undergraduate students completed the study for extra credit in a psychology class. Participants' average age was 20.2, and most (72%) were female. They had taken an average of 4.9 college-level math and science courses, with a median of 4.

Materials. We probed the conflict between science and intuition using Shtulman and Valcarcel's (2012) statement-verification task. Participants were presented with four types of scientific statements and asked to judge those statements as "true" or "false" as quickly as possible. Some statements were true from both a scientific perspective and an intuitive perspective (e.g., "bricks occupy space"); some were false both from both perspectives ("numbers occupy space"); some were true from a scientific perspective but false from an intuitive perspective ("air occupies space"), and some were false from a scientific perspective but true from an intuitive perspective ("rainbows occupy space"). The first two statement types will be referred to as intuitive and the latter two as counterintuitive. Sample statements are displayed in Fig. 1.

This task has several advantages over other measures of the conflict between science and intuition, such as explanation endorsement or object categorization (for a review, see Shtulman & Lombrozo, 2016). First, by comparing statements involving the same predicates, the linguistic complexity of intuitive and counterintuitive statements is equated across stimuli. Second, by including an equal number of objectively true and objectively false statements, the possibility of participants developing response biases is minimized. Third, by crossing truth-value (true vs. false) with intuitiveness (intuitive vs. counterintuitive), the effects of each factor are empirically distinguishable.

Participants verified 240 statements about life and 240 statements about matter. Statements about life covered the concepts of reproduction, respiration, and nutrition, and.

statements about matter covered the concepts of weight, temperature, and spatial extent. Each concept was expressed with a predicate, and each predicate was paired with one of 80 subjects. The subjects were chosen in accordance with the logic described above, so as to create 20 intuitively true statements, 20 intuitively false statements, 20 counterintuitively true statements, and 20 counterintuitively false statements.

Participants were randomly assigned to one of three priming conditions: scientific priming ($n = 34$), intuitive priming ($n = 32$), or no priming ($n = 34$). The primes were images presented for two seconds between each statement and were intended to convey either a scientific interpretation of the relevant predicate or an intuitive one. Scientific primes consisted of models, diagrams, magnifications, or dissections, similar to those found in science textbooks, whereas intuitive primes were photographs of everyday situations, typically involving people. For instance, scientific primes for statements about weight ("[x] has weight") were force diagrams, where weight was represented as a downward-pointing vector, consistent with the scientific sense of weight as the product of mass and gravity. Intuitive primes for these same statements were images of barbells, dumbbells, and scales, consistent with the intuitive sense of weight as heft.

We selected primes that corresponded to canonically intuitive or canonically scientific interpretations of the target predicates, in line with the contrast between intuitive and scientific theories of the domain,

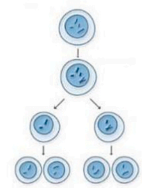
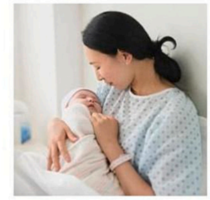
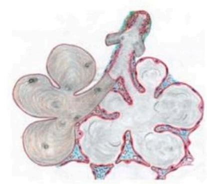
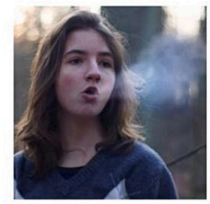
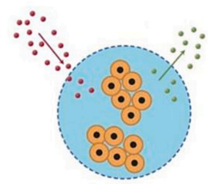
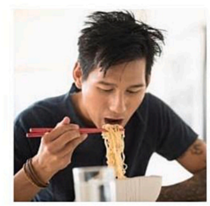
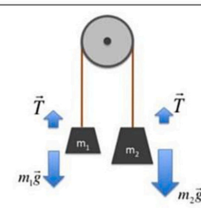
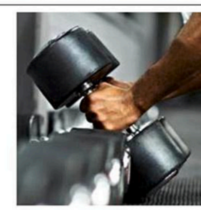
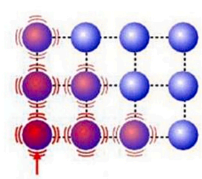
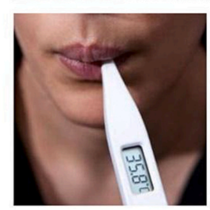
Concept	Type	Sample statement	Sample scientific prime	Sample intuitive prime
Reproduction	1	Spiders reproduce		
	2	Tables reproduce		
	3	Coral reproduces		
	4	Fire reproduces		
Respiration	1	Pelicans respire		
	2	Forks respire		
	3	Grass respires		
	4	Tornados respire		
Nutrition	1	Zebras need nutrients		
	2	Rugs need nutrients		
	3	Algae need nutrients		
	4	Robots need nutrients		
Weight	1	Gold has weight		
	2	Minutes has weight		
	3	Clouds have weight		
	4	Heat has weight		
Temperature	1	Steel has a temperature		
	2	Dreams have a temperature		
	3	Dust has a temperature		
	4	Shadows have a temperature		
Spatial extent	1	Bricks occupy space		
	2	Numbers occupy space		
	3	Air occupies space		
	4	Rainbows occupy space		

Fig. 1. Sample statements and primes, organized by domain and statement type: (1) intuitively true, (2) intuitively false, (3) counterintuitively true, (4) counter-intuitively false.

and that differed qualitatively in appearance. The intuitive primes depicted objects observable from daily life, whereas the scientific images depicted unobservable objects (in the case of dissections and magnifications) or abstractions (in the case of models and diagrams). Participants in the no-priming condition were shown a fixation cross for two seconds between each statement.

Procedure. The task was administered in 6 blocks of 80 statements, for a total of 480 statements. All statements within a block contained the same predicate, such as "respires" or "has weight." Statements were presented in a random order, as were the blocks. Participants were not given a time limit, but they were encouraged to answer as quickly as possible. The task was administered using MediaLab v1.21.

2.2. *Results*

We explored the effect of priming across domains and then assessed the consistency of this effect within domains. Participants verified 86% of statements correctly overall, and their mean response time was 1145.2 ms ($SD = 1077.0$ ms). Response times greater than two standard deviations above the mean were excluded from analysis, as were response times less than 250 ms—a duration too short for participants to have read the statement and provided a deliberate response. For all analyses, statement type (intuitive vs. counterintuitive) was collapsed across truth-value, as truth-value was deliberately balanced across statement type (see Fig. 1). For our analysis of response latency, we further excluded response times for incorrect responses, though the results do not change if those times are included.

We used (generalized) linear mixed models (GLMMs) to analyze participants' response accuracies and latencies. Models were fit with maximal by-participant and by-item random effects structures when they allowed for satisfactory model convergence. In cases of poor convergence, we followed suggestions of Barr, Levy, Scheepers, and Tily (2013) and Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017) to find maximal converging random effects structures. Inference for fixed effects was carried out using Type II likelihood ratio test model comparisons. Data and R scripts to reproduce all analyses in Study 1 and Study 2 are available at the Open Science Framework: https://osf.io/3phmt.

Response Accuracy. We fit a binomial GLMM on correct responses with statement type, prime type, and their interaction as fixed effects. Participants' accuracy at verifying intuitive and counterintuitive statements by priming condition is displayed in Fig. 2. Accuracy varied by statement type, LRT $X^2(1) = 123.55$, $p < .001$. Participants were more accurate for intuitive statements than counterintuitive statements, odds-ratio (OR) = 6.30, 95% CI [4.66, 8.42]. Accuracy also varied by prime type, LRT $X^2(2) = 7.42$, $p = .025$. Participants' were more accurate in the scientific prime condition than the intuitive prime condition, OR = 1.49, 95% CI [1.04, 2.14], and no-prime condition, OR = 1.57, 95% CI [1.10, 2.24]. Accuracy in the intuitive prime and no-prime conditions was similar, OR = 1.06, 95% CI [0.74, 1.53]. There was no interaction between statement type and prime type, LRT $X^2(2) = 1.58$, $p = .454$.

Response Latency. We fit a gaussian LMM on correct response times with statement type, prime type, and their interaction as fixed effects. Participants' speed at verifying intuitive and counterintuitive statements by priming condition is displayed in Fig. 2.

Speed varied by statement type, LRT $X^2(1) = 95.44$, $p < .001$. Participants were slower to correctly verify counterintuitive statements than intuitive statements (1064 ms vs. 940 ms, $b = 124$ ms, 95% CI [100 ms, 148 ms]). Speed did not vary by prime type, LRT $X^2(2) = 0.64$, $p = .725$. Indeed, a follow-up Bayes Factor analysis found very strong evidence for null effects of the scientific prime (BF = 0.019) and intuitive prime (BF = 0.024) conditions relative to the no-prime condition (see OSF files for details). There was, however, an interaction between the statement type and prime type, LRT $X^2(2) = 6.85$, $p = .033$, such that the difference in response times.

between counterintuitive and intuitive statements was smaller in the intuitive prime condition than the no-prime condition (103 ms vs. 140 ms, $b = 37$ ms, 95% CI [9 ms, 65 ms]), possibly because participants in the no-prime condition had no cues for privileging one interpretation of a counterintuitive statement over another and thus deliberated longer.

Effects By Domain. Scientific primes increased the accuracy but not the speed of participants' responses. We next explored whether these

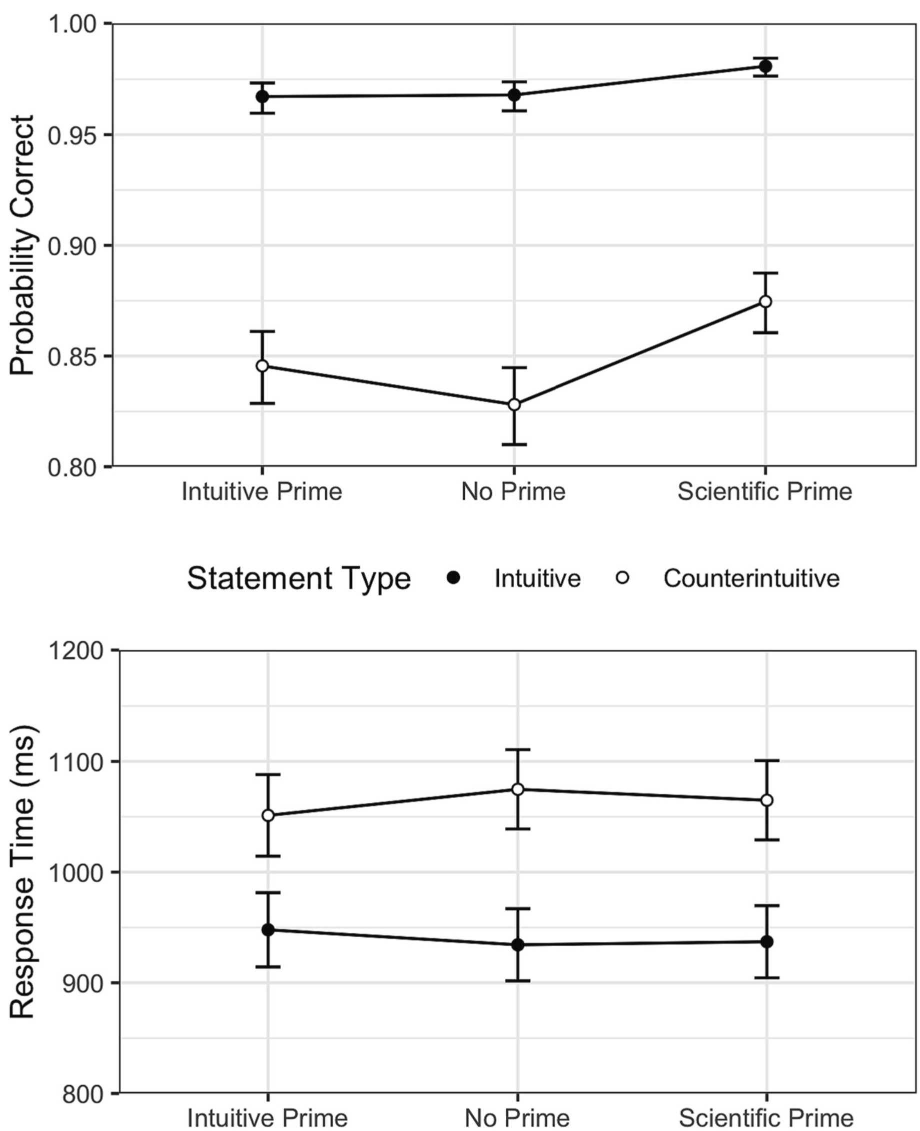


Fig. 2. Estimated probability of correct response (top) and mean response time (bottom) for intuitive by prime type. Error bars represent +/− *SE*.

effects held for each domain or were driven by one domain in particular. In the biological domain, response accuracy varied both by statement type (LRT $X^2(1) = 51.98$, $p < .001$) and by prime type (LRT $X^2(2) = 8.94$, $p = .011$), but response latency varied only by statement type (LRT $X^2(1) = 42.98$, $p < .001$). Likewise, in the physical domain, response accuracy varied both by statement type (LRT $X^2(1) = 105.38$, $p < .001$) and by prime type (LRT $X^2(2) = 5.16$, $p = .076$), but response latency varied only by statement type (LRT $X^2(1) = 114.42$, $p < .001$). No interaction effects were observed in either domain for either measure. In sum, priming affected accuracy in both domains but did not affect speed on the whole.

2.3. Discussion

Study 1 examined whether the conflict between science and intuition can be reduced by priming. We found that scientific primes did not generally increase the speed of participants' statement verifications but did increase their accuracy, at least slightly. Participants verified counterintuitive statements like "clouds have weight" more accurately when these statements were interspersed with scientific depictions of weight (force diagrams) than when interspersed with intuitive depictions of the same concept (images of dumbbells and scales) or no depictions at all. These findings indicate that counterintuitive scientific ideas can be primed with the right contextual cues, which could, in principle, facilitate scientific reasoning more broadly.

The finding that scientific primes increased accuracy but not speed suggests that the conflict between science and intuition can be shifted in one direction or the other but cannot be circumvented altogether. Scientific primes allowed participants to privilege scientific interpretations of a statement over intuitive ones, but both interpretations appear to have been activated upon participants' initial reading, yielding cognitive conflict. For instance, a statement like "clouds have weight" may have activated both a scientific sense of weight (which applies to all material substances, including clouds) and an intuitive sense (which does not apply to clouds) regardless of what primes participants saw. But scientific primes prompted them to endorse the scientific sense of weight and, ultimately, judge the statement "true."

It's possible that scientific primes promoted accuracy at a higher level of reasoning—by motivating participants' to pay more attention or by heightening their error-monitoring skills—but we suspect the effect was content-specific. An accuracy-oriented mindset would be useful only if participants were aware of, and attuned to, the relevant scientific interpretations of each statement. At the same time, intuitive primes did not decrease accuracy relative to no primes, suggesting that participants' default interpretation was an intuitive one. If priming had achieved its effect by fostering a general mindset, then participants in the intuitive-prime condition should have performed comparably worse, having been encouraged to adopt an informal, non-technical interpretation of the task. That said, future research could assess the scope of scientific primes by including no-prime blocks among the scientific-prime blocks, to determine whether participants would continue to perform more accurately even in the absence of content-specific cues.

The findings from Study 1 suggest that the conflict between science and intuition is malleable but not avoidable. Priming participants to adopt a scientific interpretation of weight, temperature, and space (in the case of matter) or reproduction, respiration, and nutrition (in the case of life) improved their performance but did not radically alter it. In Study 2, we attempted to improve performance even more by providing direct instruction on the scientific properties of life and matter. This instruction was intended to remind participants of the relevant science while also encouraging them to apply that science to the task at hand.

3. Study 2

Study 2 employed a pre-post design, where participants verified counterintuitive scientific statements before and after a content-specific tutorial. Participants verified statements about life and matter, similar to Study 1, but received instruction on only one of the two topics. This design allowed us to disentangle the domain-specific effects of instruction from domain-general effects of practice with the task or familiarity with the materials.

We predicted that instruction would increase the accuracy of participants' statement verifications within the domain of instruction and for statements where science and intuition conflict (counterintuitive statements). For statements where science and intuition agree (intuitive statements), we expected instruction to be less impactful because participants would be highly accurate from the start. Whether instruction would also increase the speed of participants' verifications was an open question, given that priming had no overall effect on speed in Study 1.

3.1. Method

Participants. The participants were 138 college undergraduates, compensated with extra credit or a small stipend. Five additional participants were tested but dropped for not completing the task as directed. Participants' average age was 20.2, and most (74%) were female. They had taken an average of 6.1 college-level math and science courses, with a median of 4.

Materials. As in Study 1, participants made true-or-false judgments for 240 statements about life and 240 statements about matter. The statements were generated by pairing one of three predicates in each domain with one of 80 subjects. In the domain of life, the predicates were "reproduces," "needs nutrients," and "grows and develops," and in the domain of matter, they were "has weight," "takes up space," and "is composed of atoms." The predicates were slightly different from those used in Study 1 because we intended to administer the same task to children and did not think children would know that all living things respire or that all material substances have a temperature (see Young & Shtulman, 2020).

Participants completed a tutorial on life or matter midway through the experiment. The tutorial on life emphasized that all living things need energy and nutrients, grow and develop, react to stimuli in their environment, and reproduce. It also addressed the misconception that life is synonymous with self-directed motion, providing examples of entities that do not appear to move on their own but are alive (such as algae) and entities that do appear to move on their own but are not alive (such as comets). The tutorial on matter emphasized that all matter occupies space, has weight, is composed of atoms, and can undergo phase transitions. It also addressed the misconception that matter is synonymous with visibility or tangibility, providing examples of entities that cannot be seen or felt but are material (such as vapors) and entities that can be seen or felt but are not material (such as lightning).

Both tutorials contained a mixture of text, images, and videos and took approximately seven minutes to complete. The tutorials were followed by eight multiple-choice questions intended to assess participants' engagement with the material. Four questions assessed their comprehension of general principles (e.g., "Which criteria can you use to know something is made of matter?"), and four questions assessed their attentiveness to specific examples (e.g., "What was the color of the balloon in the video?"). Most participants (85%) answered all eight questions correctly, and the rest missed either one question (14%) or two (1%), so all participants were retained for the final analysis.

Procedure. Each study session proceeded in three phases. First, participants verified 120 statements about life and 120 statements about matter (the pretest). Next, they completed a tutorial on life or matter, including the post-tutorial quiz. Lastly, they verified 120 additional statements about life and 120 additional statements about matter (the posttest). Half of the participants received the tutorial on life, and half received the tutorial on matter.

Participants completed the pretest and posttest in blocks. They saw a screen introducing a particular predicate (e.g., "Does it need nutrients?"), followed by 40 statements incorporating that predicate (e.g., "seaweed needs nutrients"). Ten of the statements were intuitively true ("otters need nutrients"); ten were intuitively false ("boulders need nutrients"); ten were counterintuitively true ("bacteria need nutrients"); and ten were counterintuitively false ("robots need nutrients"). The statements were randomly ordered within a block, and the blocks were randomly ordered within the testing phase, such that biological and physical predicates were intermixed.

Participants saw the same predicates at pretest and posttest, but those predicates were paired with 40 new subjects. The subjects presented at pretest for half the participants were presented at posttest for the other half and vice versa. This variable was crossed with whether participants received the tutorial on life or the tutorial matter to ensure that the effects of instruction were not confounded with any item effects at pretest or posttest.

3.2. Results

Our analytic strategy followed that of Study 1. We used GLMMS to analyze participants' response accuracies and latencies. Models were fit with maximal converging by-participant and by-item random effects structures. Response times greater than two standard deviations above the mean were excluded from analysis (i.e., those greater than 3142 ms), as were response times less than 250 ms. For our analysis of response latency, we further excluded response times for incorrect responses, though the results do not change if those times are included.

Our analyses focus on whether the statements were targeted by instruction or not, collapsed across content domain. Responses were not identical across domains; participants verified statements about life more accurately and more quickly than they verified statements about matter (accuracy: 92% vs 86%; speed: 906 ms vs. 1038 ms), but the effects of the tutorials were largely the same. In the analyses below, we refer to the domain of statements targeted by instruction as the *instructed* domain and the other domain as the *uninstructed* domain. The latter served as a control for the former in that it allowed us to differentiate pre-post changes in performance due to instruction from those due to familiarity with the materials or practice with the task. Participants' verified statements from both domains under the same testing conditions and across the same delay from pretest to posttest, but they received content-relevant training in only one of the two domains (the instructed domain).

Response Accuracy. We fit a binomial GLMM on correct responses with statement type (intuitive vs. counterintuitive), assessment period (pretest vs. posttest), instruction (instructed vs. uninstructed), and their interaction as fixed effects. Fig. 3 shows participants' accuracy at

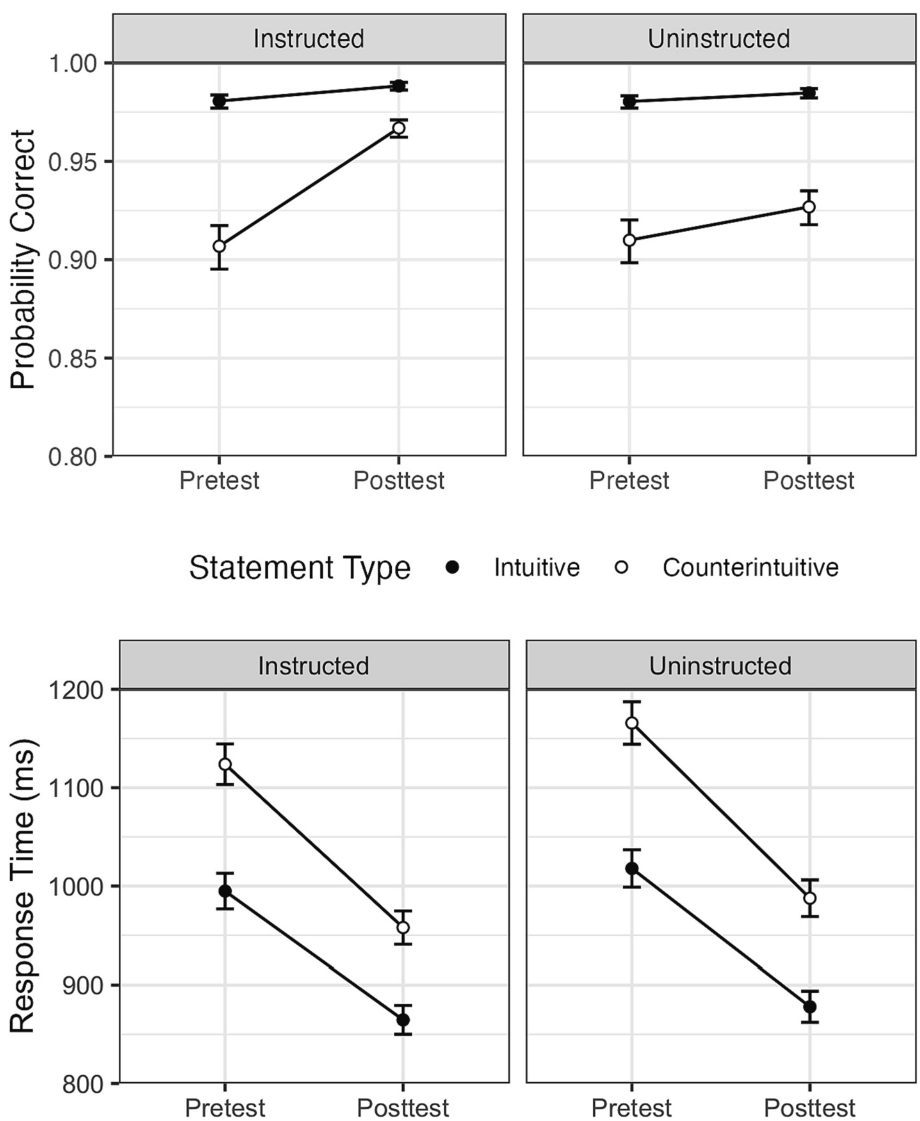


Fig. 3. Estimated probability of correct response (top) and mean response time (bottom) for intuitive and counterintuitive statements before and after instruction, as a function of whether the statements were from the instructed domain (left) or uninstructed domain (right). Error bars represent +/− *SE*.

verifying intuitive and counterintuitive statements by assessment period and instruction.

Accuracy varied by statement type, LRT $X^2(1) = 102.65$, $p < .001$. Participants were more accurate for intuitive statements than counterintuitive statements, OR = 4.44, 95% CI [3.39, 5.75]. Accuracy also varied by assessment period, LRT $X^2(1) = 35.42$, $p < .001$, such that responses were more accurate at posttest than pretest, OR = 1.69, 95% CI [1.43, 1.99]. There was also an effect of instruction, LRT $X^2(1) = 5.48$, $p < .019$, with participants responding more accurately in the instructed domain than the uninstructed domain, OR = 1.31, 95% CI [1.08, 1.59].

These effects were qualified by a three-way interaction between statement type, assessment period, and instruction, LRT $X^2(1) = 6.22$, $p < .013$. Instruction was most effective at increasing the accuracy of participants' verifications for counterintuitive statements within the targeted domain. More specifically, pre-post gains in accuracy for counterintuitive statements were greater in the instructed domain than the uninstructed domain, OR = 2.39, 95% CI [1.86, 3.08], and pre-post gains in accuracy in the instructed domain were greater for counterintuitive statements than intuitive statements, OR = 1.80, 95% CI [1.26, 2.57].

Response Latency. We fit a gaussian LMM on correct response times with statement type (intuitive vs. counterintuitive), assessment period (pretest vs. posttest), instruction (instructed vs. uninstructed), and their interaction as fixed effects. Fig. 3 shows participants' speed at verifying intuitive and counterintuitive statements by assessment period and instruction.

Speed varied by statement type, LRT $X^2(1) = 85.48$, $p < .001$. Participants were slower to correctly verify counterintuitive statements than intuitive statements (1059 ms vs. 939 ms, $b = 120$ ms, 95% CI [98 ms, 142 ms]). Speed also varied by assessment period, LRT $X^2(1) =164.16$, $p < .001$, such that participants responded faster at posttest than pretest (922 ms vs. 1076 ms, $b = 154$ ms, 95% CI [136 ms, 171 ms]). There was also an effect of instruction, LRT $X^2(1) = 7.64$, $p < .006$, with participants responding more quickly in the instructed domain than the uninstructed domain (985 ms vs. 1012 ms, $b = 27$ ms, 95% CI [9 ms, 45 ms]).

These effects were additionally qualified by an interaction between assessment period and statement type LRT $X^2(1) = 25.15$, $p < .001$. The difference in response times between counterintuitive and intuitive statements was smaller at posttest than at pretest (101 ms vs. 138 ms, $b = 37$ ms, 95% CI [21 ms, 53 ms]). That is, response times for counterintuitive statements decreased more than response times for intuitive ones. But this decrease was no greater in the instructed domain than the uninstructed one (166 ms vs. 178 ms, $b = 12$ ms, 95% CI [−13 ms, 37 ms]), and there was no three-way interaction between instruction, assessment period, and statement type, LRT $X^2(1) = 0.03$, $p = .852$). Indeed, a follow-up Bayes Factor analysis found decisive evidence for a null three-way interaction (BF = 0.004; see OSF files for details).

Follow-up analyses indicate that the assessment period by statement interaction was due to familiarity with the task rather than a general effect of instruction. Prior to *any* instruction, pretest response latencies decreased more for counterintuitive statements than intuitive ones over the six statement blocks. That is, there was a block by statement type interaction in a pretest-only LMM that also included block, LRT $X^2(1) = 17.31$. $p < .001$, and there were no effects of instruction, or interactions with instruction, in the pretest-only LMM (see OSF files for details).

3.3. Discussion

Does instructing participants on domain-specific scientific principles increase the accessibility of counterintuitive scientific ideas? Yes and no. Instruction increased how accurately participants verified counterintuitive scientific ideas but not how quickly. Participants did respond faster from the beginning of the experimental session to the end, but they did so in both the instructed domain and the uninstructed one, consistent with a familiarity effect but not a training effect. Training targeted a single type of statement—counterintuitive statements in the instructed domain—so response times should have changed primarily for that type relative to the others, but no such difference was observed. While the gap in response latency between intuitive and counterintuitive statements did decrease, this decrease was miniscule (37 ms) and occurred in both the instructed and uninstructed domains.

In contrast to the absence of instructional effects on speed, the instructional effects on accuracy were substantial. Given just seven minutes of instruction, the gap in accuracy between intuitive and counterintuitive statements nearly closed, decreasing from 10% at pretest to 4% at posttest. No such decrease was observed in the uninstructed domain, indicating that improvements in accuracy were specific to instruction.

These findings parallel the findings from Study 1, where priming participants to adopt a scientific interpretation of counterintuitive statements increased the accuracy of their verifications but not the speed. They also parallel the findings from Young and Shtulman (2020), who administered an abbreviated version of Study 2 to elementary-school-aged children. Similar to adults in the present study, the children in Young and Shtulman's (2020) study verified counterintuitive scientific statements more accurately following instruction but no more quickly.

There were, of course, baseline differences between children and adults. Children were less accurate overall, verifying 65% of counterintuitive statements correctly at pretest and 75% correctly at posttest within the domain of instruction, compared to 81% and 92% for adults. Children were also substantially slower, taking around three seconds to verify counterintuitive statements compared to the one second taken by adults. It was thus possible that adults would show a different pattern of results, either because they began the task with less room for improvement or because they responded to the task with more precision. Yet both groups responded to instruction similarly, with greater accuracy for counterintuitive statements relative to intuitive ones but no concomitant changes in speed. This parallel suggests that conflict between science and intuition is ubiquitous, impacting those with ample science education as well as those with very little. While instruction helps everyone resolve this conflict in favor of science, it does not eliminate the conflict itself.

To be clear, instruction proved helpful only for statements targeted by that instruction. Participants who received instruction on the scientific properties of life verified statements about life more accurately at posttest but not statements about matter. And participants who received instruction on the scientific properties of matter verified statements about matter more accurately at posttest but not statements about life. Pre-post gains in accuracy were domain-specific; the absence of such gains in the uninstructed domain indicates that domain-general factors, such as increased familiarity with the task or increased pressure to pay attention, were not responsible for improvements in accuracy. Instruction, like priming, appears to allow reasoners to privilege scientific interpretations of counterintuitive scientific ideas over intuitive ones, though doing so still requires additional time.

4. General discussion

Scientific ideas that defy intuition are more difficult to access than those that accord with intuition, as revealed by how accurately and how quickly these ideas are verified (Goldberg & Thompson-Schill, 2009; Kelemen et al., 2013; Shtulman & Valcarcel, 2012). Counterintuitive statements like "bacteria need nutrients" or "bubbles have weight" appear to elicit contradictory responses—"false" according to intuition but "true" according to science—and it takes people appreciably longer to select the correct (scientific) response in comparison to statements where science and intuition agree.

Here, we explored how context and instruction impact such verifications. Across two domains, several concepts, and hundreds of statements, we found that context and instruction have specific yet robust effects. Scientific priming and scientific training increased the likelihood that participants would select the correct response but did not change how quickly that response was selected (relative to intuitive statements), implying that context and instruction can help participants privilege science over intuition but cannot bypass the conflict altogether. In other words, counterintuitive scientific ideas can be verified more accurately with contextual or instructional support, but the time cost remains the same. Indeed, our most direct intervention—-training—improved the accuracy of participants' verifications but not their speed, and it improved accuracy for only one type of statement: counterintuitive statements within the domain of instruction. Targeting early-acquired intuitions improves scientific reasoning at odds with those intuitions, but it does not alter the speed or scope of scientific reasoning more generally.

These results parallel those obtained with professional scientists (Goldberg & Thompson-Schill, 2009; Shtulman & Harrington, 2016). Professional scientists are more accurate than non-scientists at affirming counterintuitive scientific ideas, but they still exhibit a reliable lag in the time taken to affirm counterintuitive ideas relative to intuitive ones. We attempted to attenuate that lag directly, by providing instruction tailored to the judgments participants made in the moment they made them, but our attempts proved unsuccessful. The conflict between science and intuition may be an inevitable byproduct of holding competing representations of the same phenomena (Shtulman, 2017; Thagard, 2014)—representations elicited in parallel when we reason about those phenomena. Collectively, these findings suggest that science learning is a process of accretion rather than refinement; learners acquire additional representations of a domain rather than refine a single representation, and they must learn how to coordinate and prioritize those representations. That said, scientists do verify counterintuitive scientific ideas faster than nonscientists do even if the lag between intuitive and counterintuitive ideas persists (Allaire-Duquette et al., 2021; Masson et al., 2014). This finding implies that scientific expertise can attenuate the conflict between science and intuition, though it remains notable that decades of education and training do not eliminate the conflict altogether.

A critical question for future research is why the conflict between science and intuition is so robust. One possibility is that this conflict, once established, remains stable across development, akin to the persistence of visual illusions despite awareness of their illusory nature (Pylyshyn, 1999) or the persistence of heuristic-based inference strategies despite awareness of their suboptimality (Kahneman, 2011). Visual illusions and cognitive heuristics do not diminish in strength once we are aware of them, but they do not increase in strength either. Rather, they constitute a stable backdrop for perceiving or evaluating new information. If the conflict between science and intuition follows a similar dynamic, then this conflict is likely to appear early in science education and plateau soon after. Additional education might facilitate more elaborate or deliberate reasoning, but it would not weaken the immediate conflict elicited by ideas that evoke both scientific and intuitive interpretations.

On the other hand, the conflict between science and intuition may vary with the strength and consistency of the underlying theories. Intuitive theories are, after all, theories (Carey, 1985; Gopnik & Meltzoff, 1997). They are constructed from data, and they are open to revision, at least initially. There may be important representational differences between intuitive theories and the perceptual biases that give rise to visual illusions, rendering the former more malleable and context-dependent. On this view, the conflict between science and intuition should vary with the input that supports intuitive theories, including perceptual input, such as the observation that animals move but plants do not (Opfer & Siegler, 2004), and linguistic input, such as practice of using different terms to describe the same biological processes in animals and plants (Stavy & Wax, 1989). These inputs will differ across domains and may also differ across development. Studies of children in the earliest stages of science learning could shed light on how the lifelong conflict between science and intuition becomes established in the first place.

If this conflict is inevitable, as multiple lines of research suggest it is (Belanger, Potvin, Horst, Shtulman, & Mortimer, 2022), then those tasked with teaching science should not attempt to eradicate it. Rather, science educators might help students recognize the existence—and persistence—of such conflict, highlighting examples of attitudes or behaviors grounded in intuition rather than science. Addressing students' pre-instructional misconceptions may be beneficial not only for introducing scientific concepts but also for promoting their application beyond the science classroom, given that real-world situations which activate students' scientific concepts will also activate contradictory intuitions.

Science educators might also benefit from recognizing that cognitive conflict is a sign of instructional success rather than failure. Students should be expected to experience conflict both at the beginning of instruction and the end, as even professional scientists experience conflict between the concepts that structure their discipline and the concepts that structure everyday life (Lewis & Linn, 1994). This conflict is not driven by ignorance but by competition between two substantive theories of the same phenomena, and its emergence merits further study as a marker of science learning and a signature of mature scientific reasoning.

Credit author statement

Andrew Shtulman: conceptualization, funding acquisition, investigation, methodology, project administration, resources, software, supervision, writing- original draft, writing - revision & editing.

Andrew G. Young: conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, writing - original draft, writing - revision & editing.

Data availability

Data and analyses are available at the Open Science Framework: https://osf.io/3phmt

References

Allaire-Duquette, G., Bélanger, M., Grabner, R. H., Koschutnig, K., & Masson, S. (2019). Individual differences in science competence among students are associated with ventrolateral prefrontal cortex activation. *Journal of Neuroscience Research, 97*, 1163–1178.

Allaire-Duquette, G., Brault Foisy, L. M., Potvin, P., Riopel, M., Larose, M., & Masson, S. (2021). An fMRI study of scientists with a Ph. D. in physics confronted with naive ideas in science. *npj Science of Learning, 6*, 11.

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

Barlev, M., Mermelstein, S., & German, T. C. (2017). Core intuitions about persons coexist and interfere with acquired Christian beliefs about god. *Cognitive Science, 41*, 425–454.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Belanger, M., Potvin, P., Horst, S., Shtulman, A., & Mortimer, E. (2022). *Multidisciplinary perspectives on representational pluralism in human cognition*. Routledge.

Busch, J. T., Watson-Jones, R. E., & Legare, C. H. (2017). The coexistence of natural and supernatural explanations within and across domains and development. *British Journal of Developmental Psychology, 35*, 4–20.

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

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

Chi, M. T. H. (2009). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 89–110). Routledge.

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.

Foisy, L. M. B., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education, 4*, 26–36.

Friedman, S., & Goldwater, M. (2023). Computational modeling of representational pluralism in explanations. In M. Belanger, P. Potvin, S. Horst, A. Shtulman, & E. Mortimer (Eds.), *Multidisciplinary perspectives on representational pluralism in human cognition* (pp. 81–103). Routledge.

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

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. The MIT Press.

Harris, P., & Gimenez, M. (2005). Children's acceptance of conflicting testimony: The case of death. *Journal of Cognition and Culture, 5*, 143–164.

Harris, P. L. (2011). Conflicting thoughts about death. *Human Development, 54*, 160–168.

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

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*, 138–143.

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.

Lane, J. D., Zhu, L., Evans, E. M., & Wellman, H. M. (2016). Developing concepts of the mind, body, and afterlife: Exploring the roles of narrative context and culture. *Journal of Cognition and Culture, 16*, 50–82.

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

Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. *Journal of Research in Science Teaching, 31*, 657–677.

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.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Merz, C. J., Dietsch, F., & Schneider, M. (2016). The impact of psychosocial stress on conceptual knowledge retrieval. *Neurobiology of Learning and Memory, 134*, 392–399.

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

National Science Teachers Association. (2013). *Next generation science standards*. National Academies Press.

Nersessian, N. J. (1989). Conceptual change in science and in science education. *Synthese, 80*, 163–183.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49*, 301–332.

Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching, 54*, 1121–1142.

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education, 13*, 21–43.

Preston, J., & Epley, N. (2009). Science and god: An automatic opposition between ultimate explanations. *Journal of Experimental Social Psychology, 45*, 238–241.

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

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Basic Books.

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). Oxford University Press.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8*, 118–137.

Shtulman, A., & Legare, C. H. (2020). Competing explanations of competing explanations: Accounting for conflict between scientific and folk explanations. *Topics in Cognitive Science, 12*, 1337–1362.

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner, & A. Baron (Eds.), *Core knowledge and conceptual change* (pp. 49–67). Oxford University Press.

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

Slaughter, V., & Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology, 46*, 1–30.

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.

Springer, K. (1995). Acquiring a naive theory of kinship through inference. *Child Development, 66*, 547–558.

Stavy, R., & Wax, N. (1989). Children's conceptions of plants as living things. *Human Development, 32*, 88–94.

Stricker, J., Vogel, S. E., Schöneburg-Lehnert, S., Krohn, T., Dögnitz, S., Jud, N., & Grabner, R. H. (2021). Interference between naïve and scientific theories occurs in mathematics and is related to mathematical achievement. *Cognition, 214*, Article 104789.

Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., & Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: A case study of vitalist biology. *Cognitive Psychology, 95*, 145–163.

Thagard, P. (2014). Explanatory identities and conceptual change. *Science & Education, 23*, 1531–1548.

Tracy, J. L., Hart, J., & Martens, J. P. (2011). Death and science: The existential underpinnings of belief in intelligent design and discomfort with evolution. *PLoS One, 6*, Article e17349.

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior, 31*, 344–355.

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.

Vosniadou, S., Pnevmatikos, D., Makris, N., Lepenioti, D., Eikospentaki, K., Chountala, A., & Kyrianakis, G. (2018). The recruitment of shifting and inhibition in on-line science and mathematics tasks. *Cognitive Science, 42*, 1860–1886.

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.

Young, A. G., & Shtulman, A. (2020). How children's cognitive reflection shapes their science understanding. *Frontiers in Psychology, 11*, 1247.

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