Can Science Beat Out Intuition? Increasing the Accessibility of Counterintuitive Scientific Ideas

Andrew G. Young (ayoung2@oxy.edu), Jasper Laca, Giovanna Dieffenbach, Eushrah Hossain, Devon Mann, & Andrew Shtulman (shtulman@oxy.edu)
Department of Psychology, Occidental College
1600 Campus Road, Los Angeles, CA 90041

Abstract

Scientific ideas can be difficult to affirm if they contradict earlier-developed intuitive theories. Here, we investigated how instruction on counterintuitive scientific ideas affects the accessibility of those ideas under time pressure. Participants (138 college undergraduates) verified, as quickly as possible, statements about life and matter before and after a tutorial on the scientific properties of life or matter. Half the statements were consistent with intuitive theories of the domain (e.g., "zebras reproduce") and half were inconsistent (e.g., "mushrooms reproduce"). Participants verified the latter less accurately and more slowly than the former, both before instruction and after. Instruction did, however, increase accuracy for counterintuitive statements within the domain of instruction, but changes in accuracy were not accompanied by changes in speed. These results confirm the conclusion drawn from studies with professional scientists that scientific ideas can be prioritized over intuitive ones but the conflict between science and intuition cannot be eliminated altogether.

Keywords: conceptual development, scientific reasoning, explanatory coexistence, intuitive theories

Introduction

Does air have weight? Does air take up space? Is air composed of atoms? Any chemist will tell you "yes," "yes," and "yes." Air is a gas; gases are a form of matter; and all matter has weight, volume, and an atomic structure. From an intuitive point of view, 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 homogenous and indecomposable.

This tension between conceiving of air as matter and conceiving of it as empty space is just one of many examples of the conflict between scientific and intuitive representations of the natural world (Carey, 2009; Thagard, 2014). Most adults today believe that the earth orbits the sun, but this idea defies deep-seated intuitions. The earth betrays no evidence of motion, whereas the sun appears to move daily, and these observations lead children to believe that the sun orbits the earth before they learn the opposite is true (Vosniadou & Brewer, 1994). Learning the opposite does not appear to erase the earlier belief, however. When adults are asked to verify that the earth orbits the sun as quickly as possible, many mistakenly claim that it does not. And those who correctly verify that the earth orbits the sun take longer to do so than to verify that the moon orbits the earth—a statement that is scientifically true but also intuitively true (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 large objects are more likely to sink in water than small objects (Potvin & Cyr, 2017), that fractions with large denominators are greater than fractions with small denominators (Vamvakoussi, Van Dooren, & Verschaffel, 2012), and that the origins of natural kinds, like geysers and earthworms, are explained by the functions they serve within an ecosystem (Rottman, Zhu, Wang, Seston Schillaci, Clarke, & Kelemen, 2015). These misconceptions are not mere factual errors; they are grounded in intuitive theories of the domain that make fundamentally different assumptions about the entities and processes within the domain.

Intuitive theories are well-documented among children, as children construct these theories prior to learning scientific theories of the relevant phenomena (Shtulman, 2017). Scientific theories, once learned, have long been assumed to replace intuitive theories (see Shtulman & Lombrozo, 2016), but findings like those reviewed above suggest that the 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 with intuitive theories than for statements that accord with those theories. This finding has been documented in several studies across several domains of science, from evolution to mechanics to thermodynamics (Barlev, Mermelstein, & German, 2017; Merz, Dietsch, & Schneider, 2016; Shtulman & Valcarcel, 2012; Vosniadou et al., 2015).

When scientific theories conflict with intuitive theories, the resolution of such conflict appears to require inhibitory control processes. 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— the anterior cingulate cortex and the dorsolateral prefrontal cortex—when they answer those questions correctly (Foisy, Potvin, Riopel, & Masson, 2015; 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. In this same vein, 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 across domains and across the lifespan, but are such tensions intractable? Can adults be trained to privilege science over intuition, even when responding under time pressure? Research with professional scientists suggests not. 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 become more likely 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). Likewise, Shtulman and Harrington (2016) found that science professors are uniformly slower and less accurate at verifying counterintuitive scientific statements (e.g., "air is composed of matter") relative to closely-matched intuitive ones (e.g., "rocks are composed of matter").

That said, the expertise of the scientists recruited in these studies may not have been well aligned with the tasks they were asked to complete. In Shtulman and Harrington's (2016) study, for instance, participants were asked to evaluate materials culled from ten different domains. While scientists were consistently more accurate at the task 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.'s (2013) study. A biologist who studies intra-cellular reactions may ponder the life status of plants no more often than a non-biologist, and a physicist who studies string theory may have ponder the origins of natural kinds, like moss and ozone, no more often than a non-physicist.

For these reasons, we sought a more direct test of how science instruction influences the accessibility of counterintuitive scientific ideas. To do so, we administered a version of Shtulman and Valcarcel's (2012) statement-verification task before and after a tutorial targeting the content of those statements. We adapted the task by focusing exclusively on the domains of life (basic physiology) and matter (basic chemistry). Both domains are foundational to scientific reasoning; basic physiology is necessary for learning higher-level concepts in cellular biology, evolutionary biology, and immunology, and basic chemistry is necessary for learning higher-level concepts in optics, thermodynamics, and electromagnetism. Because these domains are foundational, our participants—college undergraduates—could be expected to have acquired a scientific understanding of these domains many years earlier, during elementary school in the case of life (Carey, 1985; Hatano & Inagaki, 1994) and during middle school in the case of matter (Nakhleh, Samarapungavan, & Saglam, 2005; Smith, 2007). The domains of matter and life were also ideal for expanding our stimuli, as a handful of domain-specific predicates could be applied to a large number of domain-specific objects, as discussed below.

In addition to crafting new test materials, we crafted new tutorials for teaching participants' about the scientific properties of life and matter. These tutorials targeted ideas instantiated in the statements under verification, such as the idea that all matter has weight or that all organisms need nutrients. Participants received either the life tutorial or the matter tutorial but not both, which allowed us to disentangle the domain-specific effects of instruction from domain-general effects of practice with the task or familiarity with the materials. Our prediction was that instruction would increase the accuracy of participants' statement verifications for those within the domain of instruction and for which science and intuition conflict. Statements for which science and intuition agree were predicted to be less affected by instruction because participants should have verified those statements correctly from the start. We also predicted that instruction would decrease how long it took participants to make their verifications, though this prediction was more tentative given that prior studies have found either a negative relationship between speed and accuracy for counterintuitive statements (Shtulman & Valcarcel, 2012) or no relationship at all (Shtulman & Harrington, 2016).

Method

Participants

The participants were 138 college undergraduates, recruited through campus advertisements or visits to psychology classes. They were compensated with extra credit or a small stipend. Five additional participants were tested but dropped for various reasons (two did not complete the tutorial; two were non-native English speakers and reported difficulty with the task; one took three times longer than usual to complete the task). Most participants were female (74%), and they came from a variety of majors: 30% from the natural sciences, 45% from the social sciences, and 25% from the humanities. They had taken an average of 6.1 college-level math and science courses, though some had taken as many as 21. With this level of STEM education, participants could reasonably be expected to know the scientific properties of both life and matter.

Materials

Our measure of the conflict between science and intuition was a 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., "otters need nutrients"); some were false both from both perspectives ("boulders need nutrients"); some were true from a scientific perspective but false from an intuitive perspective ("bacteria need nutrients"), and some were false from a scientific perspective but true from an intuitive perspective ("robots need nutrients"). The first two types of statements will be referred to as *intuitive* and the latter two types as *counterintuitive*.

This task has several advantages over other measures of explanatory coexistence (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.

Our statements were generated by pairing one of three predicates in each domain with one of 80 entities. In the domain of life, the predicates were "reproduces," "needs nutrients," and "grows and develops." In the domain of matter, the predicates were "has weight," "takes up space," and "is composed of atoms." The biological predicates apply to all living things, but we predicted that participants would be inclined to apply them only to entities that appear to move on their own. Likewise, the physical predicates apply to all material things, but we predicted that participants would be inclined to apply them only to entities that can be seen or felt. These predictions were derived from the extensive literatures on intuitive theories of life (see Hatano & Inagaki, 1994) and intuitive theories of matter (see Smith, 2007).

We created our four types of statements by pairing our predicates with four types of entities, as shown in Table 1. In the domain of life, those entities were animals (deemed alive by both science and intuition), inanimate artifacts and inanimate natural kinds (deemed alive by neither science nor intuition), plants and microorganisms (deemed alive by science but not by intuition), and animate artifacts and animate natural kinds (deemed alive by intuition but not science). In the domain of matter, those entities were physical objects (deemed material by both science and intuition), abstract ideas (deemed material by neither science nor intuition), gases and other bulk-less or heft-less objects (deemed material by science but not by intuition), and the visible or tangible components of energy transfer (deemed material by intuition but not science).

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 (e.g., algae) and entities that move on their own but are not alive (e.g., 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 (e.g., vapors) and entities that can be seen or felt but are not material (e.g., 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 the general principles (e.g., "Which criteria can you use to know something is made of matter?"), and four questions assessed their attentiveness to the 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 only one question (14%) or two (1%), so all participants were included in the final analysis.

Table 1: Sample items used in the biological statements (top) and physical statements (bottom), organized by their role in scientific and intuitive views of the domain.

Is it alive?	Intuition: Yes	Intuition: No
Science: Yes	Pigs Turtles Snails	Kelp Oaks Mold
Science: No	Geysers Tornadoes Fire	Tables Pebbles Shells

Is it matter?	Intuition: Yes	Intuition: No
Science: Yes	Bricks Dumbbells Logs	Smoke Clouds Methane
Science: No	Rainbows Shadows Heat	Dreams Songs Numbers

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. Last, they verified 120 additional statements about life and 120 additional statements about matter (the posttest). Half the participants received the tutorial on life and half the tutorial on matter.

Participants completed the pretest and posttest in blocks. They saw a screen introducing a particular predicate (e.g., "does it grow and develop?"), followed by 40 statements derived from that predicate (e.g., "seaweed grows and develops"). Ten of the statements were scientifically and intuitively true; ten were scientifically and intuitively false; ten were scientifically true but intuitively false; and ten were scientifically false but intuitively true. The statements were randomly ordered within a block, and the blocks were randomly ordered within the testing phase, meaning that biological and physical predicates were intermixed. Participants saw the same predicates at pretest and posttest, but those predicates were paired with 40 new entities. The entities 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 on matter to ensure that the effects of the tutorial were not confounded with the effects of particular pretest items or particular posttest items.

Results

The statement-verification task yielded two measures: response accuracy and response latency. We analyzed each with a repeated-measures analysis of variance (ANOVA), in which statement type (intuitive vs. counterintuitive), assessment period (pretest vs. posttest), and instruction (present vs. absent) were all treated as within-participants factors. Our content domains were life and matter, but we collapse that distinction here for lack of space and focus instead on whether the statements were targeted by instruction or not. Responses were not identical across domains; participants verified the biological statements more accurately and more quickly than the physical statements (accuracy: 92% vs 86%; speed: 906 ms vs. 1038 ms), but the effects of the tutorials were largely the same.

Response Accuracy

Participants verified intuitive statements more accurately than counterintuitive statements in both the target domain and the nontarget domain and at both assessment periods (see Figure 1). A repeated-measures ANOVA confirmed that there was an effect of statement type ($F(1,137) = 351.33$, $p < .001$, $\eta_p^2 = .719$). It also revealed an effect of assessment period ($F(1,137) = 105.32$, $p < .001$, $\eta_p^2 = .435$) but no effect of instruction ($F(1,137) = 2.44$, $p = .12$, $\eta_p^2 = .018$). Participants increased their overall accuracy from pretest (87%) to posttest (92%), but this effect was qualified by a three-way interaction between statement type, assessment period, and instruction ($F(1,137) = 16.77$, $p < .001$, $\eta_p^2 = .109$).

We explored this interaction by calculating the difference in response accuracy between intuitive statements and counterintuitive statements at each assessment period and comparing those differences across assessment periods using paired-samples t tests. In the nontarget domain, the difference in response accuracy remained essentially the same from pretest to posttest (11% vs. 9%, $t(137) = 1.71$, $p = .09$, $d = 0.15$), whereas in the target domain, this difference was significantly attenuated (10% vs. 4%, $t(137) = 7.03$, $p < .001$, $d = 0.60$). Thus, instruction increased the accuracy of participants' verifications for counterintuitive statements relative to intuitive ones within the domain of instruction but not within the other domain.

Response Latency

Before submitting response latencies to an ANOVA, we calculated the mean response latency across participants and statements ($M = 1099$ ms) and removed latencies more than two standard deviations above the mean (i.e., latencies greater than 2947 ms). We also removed latencies shorter than 250 ms, as responses produced that quickly were unlikely to have been deliberate. We further culled the

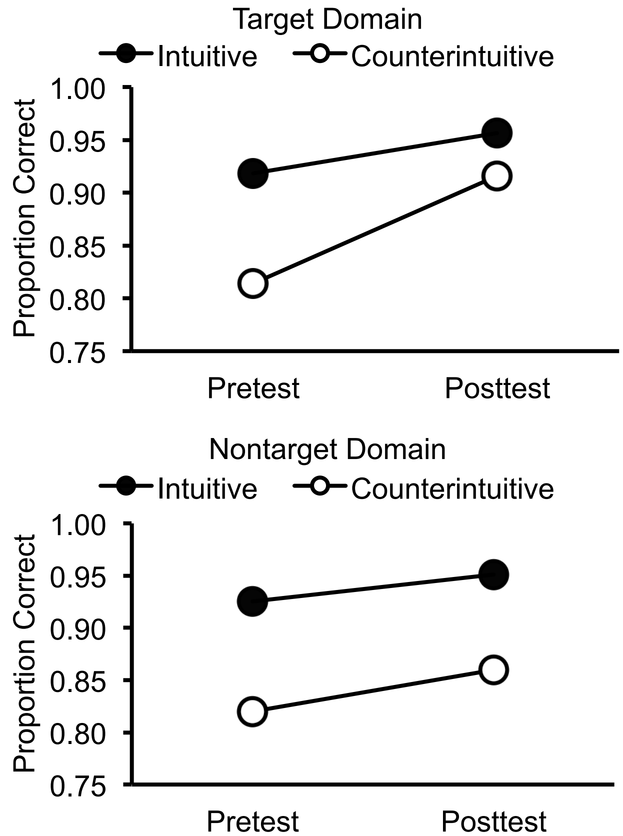

Figure 1: The proportion of intuitive and counterintuitive statements verified correctly before and after instruction, as a function of whether the statements were from the domain targeted by instruction (top) or not (bottom). All $SE < .011$.

dataset by removing latencies associated with incorrect responses. We then computed the average latency for each predicate, separating intuitive statements from counterintuitive statements and pretest statements from posttest statements.

When participants correctly verified a statement, they did so more slowly for counterintuitive statements than for intuitive ones (see Figure 2). A repeated-measures ANOVA confirmed this effect ($F(1,137) = 372.93$, $p < .001$, $\eta_p^2 = .731$). It also revealed an effect of assessment period ($F(1,137) = 238.69$, $p < .001$, $\eta_p^2 = .635$) but no effect of instruction ($F(1,137) = 3.37$, $p = .07$, $\eta_p^2 = .024$). These effects were additionally qualified by an interaction between assessment period and statement type ($F(1,137) = 8.19$, $p < .01$, $\eta_p^2 = .056$).

We explored this interaction in the same manner that we explored the interaction relating to accuracy, calculating the difference in response latency between intuitive statements and counterintuitive statements at each assessment period and comparing those differences across assessment periods. In both the target domain and the nontarget domain, these difference decreased by a small but significant amount (target domain: 106 ms vs. 85 ms, $t(137) = 1.98$, $p < .05$, $d = 0.17$; nontarget domain: 127 ms vs. 107 ms, $t(137) = 2.30$, p

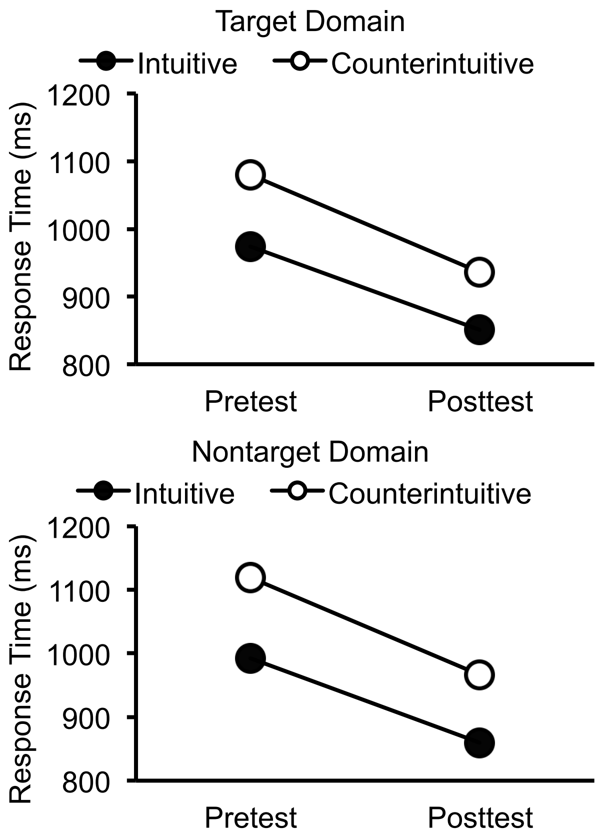

Figure 2: The mean response latency for correct verifications of intuitive and counterintuitive statements before and after instruction, as a function of whether the statements were from the domain targeted by instruction (top) or not (bottom). All *SE* < 21 ms.

< .05, $d = 0.20$). Such effects were likely due to familiarity with the task, as they were obtained in both the target and nontarget domain and did not differ across domains (as indicated by the lack of a three-way interaction).

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 those ideas are verified (Barlev, et al., 2017; Goldberg & Thompson-Schill, 2009; Kelemen et al., 2013; Shtulman & Valcarcel, 2012). The commonsense notion that scientific theories replace intuitive theories is incorrect, but it remains an open question as to how and why the two theories coexist.

Here, we attempted to increase the accessibility of counterintuitive scientific ideas by tutoring participants on the domain-specific principles that defy intuition. Although our tutorials were brief (7 minutes), they increased participants' accuracy at verifying counterintuitive scientific ideas within the domain of instruction. The gap in response accuracy between intuitive and counterintuitive statements in the target domain nearly closed, moving from 10% at pretest to 4% at posttest. And there was no concomitant decrease in the nontarget domain. Surprisingly, there was no change in how quickly participants verified the statements, at least not one that was specific to instruction. The gap in response latency between intuitive and counterintuitive statements did decrease, but this decrease was miniscule (20 ms) and occurred regardless of instruction.

These findings indicate that the conflict between science and intuition is amenable to instruction insofar that instruction helps reasoners favor scientific responses over intuitive ones, but the conflict itself cannot be eliminated. Counterintuitive statements like "yeast needs nutrients" or "clouds 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 than for statements in which science and intuition agree. Targeted instruction may increase the likelihood that participants will select the correct response, but it does not change how quickly that response is selected, implying that both response options are elicited automatically.

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 that specifically targeted the judgments at hand, in the moment those judgments were being made, but our attempt 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).

Thus, a critical question for future research is why the conflict between science and intuition appears inevitable. 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 after we become 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 is similar in nature, then this conflict is likely to appear early in science education and plateau soon after. Advanced science education might facilitate deliberate scientific reasoning, but it would not weaken the immediate conflict elicited by stimuli that evoke both scientific theories and intuitive theories.

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 early on. 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, whether that input is perceptual (e.g., observing the sun move across the sky) or linguistic (e.g., using the words "sunset" and "sunrise" to describe the sun's apparent motion). These inputs will differ across domains, and they may also differ across development. Studies of children in the earliest stages of science learning may thus provide the leverage needed to determine how the lifelong conflict between science and intuition becomes established in the first place.

Acknowledgements

We would like to thank the James S. McDonnell Foundation for supporting this research with an Understanding Human Cognition Scholar Award to Andrew Shtulman.

References

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.

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

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

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.

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*. Cambridge, MA: The MIT Press.

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

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

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.

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.

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.

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.

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

Rottman, J., Zhu, L., Wang, W., Seston Schillaci, R., Clark, K. J., & Kelemen, D. (2017). Cultural influences on the teleological stance: evidence from China. *Religion, Brain & Behavior, 7*, 17-26.

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

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

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

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, UK: Oxford University Press.

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.

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

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., Pnevmantikos, D., Makris, N., Ikospentaki, K., Lepenioti, D., Chountala, A., & Kyrianakis, G. (2015). Executive functions and conceptual change in science and mathematics. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

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