6 Navigating the conflict between science and intuition

Andrew Shtulman

Scientific misconceptions have a long and stubborn life. They form early in development, as we learn about the world through casual observation and informal education (Shtulman, 2017; Vosniadou, 1994; Wandersee et al., 1994). They block the learning of more accurate ideas taught through formal instruction (Carey, 2000; Shtulman & Walker, 2020). And they linger in our minds even after we have acquired scientific ideas that directly contradict those misconceptions (Mason & Zaccoletti, 2020; Shtulman & Lombrozo, 2016). Not even professional expertise in science is sufficient for erasing childhood misconceptions. Biologists harbor the misconception that plants are not alive (Goldberg & Thompson-Schill, 2009). Physicists harbor the misconception that everything exists for a purpose (Kelemen et al., 2013). And chemists harbor misconceptions about the structure of matter (Potvin et al., 2020).

In this chapter, I will explore the tension between scientific ideas and intuitive ideas across the lifespan. Over a decade of research demonstrates that conflict between science and intuition is immediate and automatic, spurred by distinct yet coexisting representations of the same phenomena (Allaire-Duquette et al., 2019; Barlev et al., 2017; Kelemen et al., 2013; Merz et al., 2016; Potvin et al., 2020; Shtulman & Valcarcel, 2012; Vosniadou et al., 2018). Here, I explore the dynamics of this conflict, its cognitive underpinnings and its implications for theories of conceptual change. In particular, I review evidence that scientific reasoning can be improved with priming and training but still remains onerous. Scientific ideas that conflict with earlier-developed intuitions take additional time and effort to access, implying that the tension between science and intuition is never fully resolved. I also review evidence that the reason intuitive ideas survive the acquisition of scientific theories is that intuitive ideas are continually reinforced by everyday language and everyday perception. These findings suggest that science education should not attempt to erase intuitive ideas but rather help students prioritize science over intuition – a task aided by executive function and cognitive reflection.

DOI: 10.4324/9781003189930-6

Conflicting theories

As children interact with the natural world, they construct "intuitive theories" of the phenomena they observe (Carey, 2009; Shtulman, 2017; Gopnik & Wellman, 2012). These ideas are termed *theories* because they function similarly to scientific theories. They help us explain past events, predict future events, intervene on present events and imagine counterfactual events. These ideas are termed *intuitive* because they are less precise and less accurate than scientific theories of the same phenomena. Consider everyday phenomena like motion, heat and illness. Intuitively, we understand motion as caused by a force, transferred from an agent into an object, when in reality forces are interactions between objects, causing changes in motion rather than motion itself (McCloskey, 1983). We intuitively understand heat as an invisible substance that flows in and out of objects rather than the collective motion of a system's molecules (Reiner et al., 2000). And we intuitively understand illness as the consequence of imprudent behavior, like getting cold or getting wet, rather than the transmission of microbes and their replication inside our body (Au et al., 2008).

Intuitive theories are constructed in the absence of scientific theories and constitute our earliest understanding of a domain. The process of transitioning from an intuitive theory to a scientific theory requires learning a new set of concepts and is thus known as conceptual change. Most psychological models of conceptual change assume that intuitive theories are erased in the process (see Potvin, Chapter 7, this volume), as the concepts that comprise an intuitive theory are coalesced, reanalyzed, recategorized or further differentiated (Chi, 2008; Carey, 2009; Vosniadou, 1994). This assumption is now known to be false. Research in cognitive development, cognitive neuroscience and science education has consistently shown that conceptual change yields new concepts while still preserving the old ones. Scientific theories come to reside alongside intuitive theories rather than replace them, resulting in a kind of representational plurality.

Individuals who have undergone conceptual change show evidence of retaining their intuitive theories across a wide variety of measures, including structured interviews (Rosengren et al., 2014; Shtulman et al., 2016), explanation endorsement (Evans et al., 2010; Legare & Gelman, 2008; Lombrozo et al., 2007), object classification (Babai et al., 2010; Järnefelt et al., 2015; Vosniadou et al., 2018), forced-choice comparison (Potvin & Cyr, 2017; Toyama, 2019), contextual priming (Harris & Gimenez, 2005; Preston et al., 2013), lexical priming (Preston & Epley, 2009), mouse tracking (Murray et al., 2020) and neuroimaging (Allaire-Duquette et al., 2019; Brault Foisy et al., 2015; Masson et al., 2014). This evidence has been documented across a wide range of ages, from children to adolescents to elderly adults, as well as a wide range of cultures, from

South Asia to South Africa to the Pacific Islands (see Legare & Shtulman, 2018, for a review), suggesting that the conflict between science and intuition is universal. This chapter will focus on one method in particular: statement verification.

Measuring the conflict

Participants in a statement-verification task judge whether scientific statements are true or false as quickly as possible. Some statements are consistent with an intuitive theory of the domain, and others are inconsistent with that theory. For instance, the statement "the moon revolves around the earth" is true both intuitively and scientifically, whereas the statement "the sun revolves around the earth" is intuitively true but scientifically false. Both the moon and the Sun appear to revolve around the Earth, but only the moon actually does. In contrast, the statement "the earth revolves around the sun" is scientifically true but intuitively false, because the Earth does not appear to move at all. If scientific theories replace intuitive theories, then people who have learned the relevant science should be able to verify counterintuitive scientific statements as quickly and as accurately as intuitive ones. But several studies, reviewed here, show that verifying counterintuitive statements come with a cost. People take longer to verify counterintuitive statements, and they make more errors when doing so.

The statements in a statement-verification task have to meet several criteria for the results to be meaningful. The correct response – "true" or "false" – has to be balanced across statements, to prevent participants from developing response biases, such as always responding "true." Truth value must, in turn, be crossed with intuitiveness such that intuitive statements are sometimes true ("the moon revolves around the earth") and sometimes false ("the earth revolves around the sun"), and counterintuitive statements are sometimes true ("the earth revolves around the sun") and sometimes false ("the earth revolves around the moon"). Crossing truth value and intuitiveness ensures that verification behavior is not confounded with familiarity. Intuitive ideas are more familiar than scientific ideas, by virtue of when they are acquired, but this design ensures that not all scientific ideas are unfamiliar ("the moon revolves around the earth") and not all unfamiliar ideas are scientific ("the earth revolves around the moon"). Finally, the statements must be similarly complex in their wording and grammar, so neither truth value nor intuitiveness is confounded with how long it takes to process the statement itself.

Guided by these criteria, my colleagues and I began our investigation of the conflict between science and intuition by creating 200 statements: four for each of five concepts in each of ten domains. The domains were selected to exemplify areas of knowledge where students are known to construct intuitive theories at odds with the scientific theories they learn later in life. These domains included astronomy, evolution, fractions, heat, illness, inheritance, life, light/sound, matter and motion. Within each domain, we selected five concepts that exemplify the domain's unique causal structure, and then for each concept, we created four closely matched statements that might probe for conflicting interpretations of that concept: a statement that was unambiguously true, a statement that was unambiguously false, a statement that was intuitively true but scientifically false and a statement that was scientifically true by intuitively false. For example, the four previous statements about celestial motion were intended to probe conflict between intuitive and scientific conceptions of the solar system within the broader domain of astronomy.

We instructed college-educated adults to verify our statements as quickly as possible and found, as expected, that they were more accurate at verifying statements that accorded with intuitive theories than those that conflicted with them. They also took longer to verify these statements (Shtulman & Valcarcel, 2012). This finding held for all ten domains and for the majority of concepts within each domain. It held for positive misconceptions (intuitive statements that are actually false) as well as negative misconceptions (counterintuitive statements that are actually true), and it held regardless of students' overall accuracy.

These findings indicate that scientific theories do not replace intuitive ones; intuitive theories persist, causing conflict when we reason about phenomena where the two theories diverge. Still, the participants in our first study were college undergraduates, and it's possible that the conflict we observed was a by-product of the recency of their science education. Might older adults exhibit less conflict? Older adults have more time to consolidate their scientific knowledge and more practice deploying it. We explored this possibility by administering the same task to adults between the ages of 50 and 87, averaging 66 years old (Shtulman & Harrington, 2016). Some were recruited from retirement communities, and others were recruited from the faculty at Occidental College. We found that older adults showed the same effects as younger adults – namely, decreased speed and decreased accuracy when evaluating counterintuitive statements. While the lag in accuracy for older adults was equivalent to that for younger adults, the lag in speed was nearly twice as large, indicating that time and experience do not decrease the conflict between science and intuition. If anything, they increase it.

A subset of the participants recruited from Occidental College were science professors. These participants performed more accurately than other older adults, but they too showed a response lag when verifying counterintuitive statements. Scientific expertise reduced the difference in accuracy between intuitive and counterintuitive statements, but it did not reduce the difference in speed. It appears that even scientists take longer to verify statements like "the earth revolves around the sun" relative to statements like "the moon revolves around the earth," though the difference is often smaller for experts (see Goldberg & Thompson-Schill, 2009; Potvin & Cyr, 2017).

Refining our measures

In more recent studies, described later, we have continued to use statement verification as our measure of conflict between science and intuition but we narrowed our focus from ten domains to two, creating statements exclusively about life and matter. We focused on these domains for several reasons. First, both domains are foundational to scientific thought. The properties of life are foundational to higher order concepts in cellular biology, evolutionary biology and immunology, and the properties of matter are foundational to higher order concepts in physical chemistry, organic chemistry and thermodynamics. Second, these domains are addressed in science education early and often, and we could be relatively confident that our participants had acquired a scientific understanding of the target domains, during elementary school in the case of life (Hatano & Inagaki, 1994; Solomon & Zaitchik, 2012) and middle school in the case of matter (Nakhleh et al., 2005; Smith, 2007).

Third, and most important, the domains of life and matter were ideal for expanding our stimuli, as a handful of domain-specific predicates could be applied to a large number of subjects. Predicates like "reproduces," "needs nutrients" and "grows and develops" are true of all living things and can be paired with entities that vary in whether science and intuition agree on their life status. Similarly, predicates like "has weight," "takes up space" and "is composed of atoms" apply to all material things and can be paired with items that vary in whether science and intuition agree on their material status.

With respect to life, science classifies entities that engage in metabolic activity as alive, but intuition classifies entities that move on their own as alive. Science and intuition thus agree that animals are alive and inanimate objects are not, but they disagree about the status of living objects that do not appear to move on their own, like flowers and trees, and objects that move on their own but are not alive, like wind and fire. With respect to matter, science classifies entities composed of atoms as material, but intuition classifies entities that can be seen or felt as material. Science and intuition thus agree that solid objects are material and ideas are immaterial, but they disagree about the status of material entities that cannot be perceived, like gases and vapors, and immaterial entities that can be perceived, like lightning and rainbows.

By following this logic, we created a large battery of statements that fulfill the statement-verification criteria listed earlier. Within each domain, a quarter of our statements were about entities classified as part of the domain by both science and intuition and are thus unambiguously true, such as "tigers need nutrients" and "bricks take up space." A quarter were about entities that are not classified as part of the domain by either science or intuition and are thus unambiguously false, such as "forks need nutrients" and "dreams take up space." A quarter were about entities classified as part of the domain by science but not intuition, making them scientifically true but intuitively false, such as "moss needs nutrients" and "air takes up space." And a quarter was about entities classified as part of the domain by intuition but not science, making them intuitively true but scientifically false, such as "robots need nutrients" and "rainbows takes up space." In the studies described here, we looked at factors that influence the speed and accuracy of participants' ability to verify statements where science and intuition disagree relative to those where there is agreement.

The role of priming

A counterintuitive statement like "air has weight" is difficult to verify because we represent two senses of weight: a scientific sense (the product of mass and gravity) and an intuitive sense (heaviness or heft). The intuitive sense is most likely our default sense, given that we are concerned with how heavy objects feel more often than how they interact within the Earth's gravitational field. But could we be primed to access the scientific sense first? Are there contexts in which counterintuitive scientific ideas are as easy to access as intuitive ones? Presumably, scientists reason about science with relative ease while engaged in their domain of expertise, even if they default to intuitive conceptions outside the lab (Lewis & Linn, 1994; Kozhevnikov & Hegarty, 2001).

To prime a scientific interpretation of our statements, we interspersed them with images that might connote a more formal or abstract understanding of the relevant predicates. In the domain of life, statements about reproduction ("[entity] reproduces") were interspersed with diagrams of cell division. Statements about nutrition ("[entity] needs nutrients") were interspersed with diagrams of cellular transport. And statements about respiration ("[entity] respires") were interspersed with diagrams of gas exchange. In the domain of matter, statements about weight ("[entity] has weight") were interspersed with force diagrams, depicting weight as a vector. Statements about temperature ("[entity] has a temperature") were interspersed with diagrams of molecular motion. And statements about spatial extent ("[entity] occupies space") were interspersed with ball-and-stick models of molecular structure.

A different group of participants received images that primed an intuitive interpretation of the same predicates. These images depicted everyday objects or events intended to connote observable experiences and were typically photographs rather than diagrams. In the domain of life, statements about reproduction were interspersed with images of childbirth; statements about nutrition, with people eating; and statements about respiration, with people breathing. In the domain of matter, statements about weight were interspersed with images of barbells and dumbbells; statements about temperature, with thermometers and thermostats, and statements about spatial extent, with tape measurers and rulers.

Participants in each condition verified a total of 480 statements: 80 for each of three predicates in each of two domains. The 80 statements per predicate were composed of 20 statements that were unambiguously true, 20 statements that were unambiguously false, 20 statements true by science but false by intuition and 20 statements true by intuition but false by science.

Replicating previous findings, participants verified counterintuitive statements – where science and intuition disagreed – more slowly and less accurately than intuitive ones, and these effects were observed in both priming conditions (scientific and intuitive), in both domains (life and matter), and for all predicates (reproduces, respires, needs nutrients, has weight, has a temperature and occupies space). Extending previous findings, we found that participants who saw scientific primes responded more accurately than those who saw intuitive primes, but the effect was small, yielding only a 3% boost in accuracy. Moreover, scientific primes did not increase the speed of participants' responses. The gap in response times between intuitive and counterintuitive statements was equivalent across priming conditions, indicating that scientific primes induced participants to resolve conflicting interpretations of the same statement (like "air has weight") in favor of a scientific interpretation ("air has mass, pulled by gravity") but did not diminish the conflict itself. Interestingly, participants who saw intuitive primes responded equivalently to a third group of participants who saw no prime at all. It would thus appear that, in the absence of contextual primes, participants approach the task with an intuitive mindset (see Potvin, Chapter 7, this volume, for discussion of how context can be used to alter such mindsets in an instructional setting).

The role of training

Priming a scientific mindset increased accuracy, but only by a small amount and with no concomitant effects on speed. Might a more direct manipulation yield stronger effects? In a follow-up study (Young et al., 2018), we administered a tutorial on the very properties participants were asked to verify and looked for differences in the speed and accuracy of those verifications from before the tutorial to after.

Our participants were college undergraduates. We assigned them to complete a tutorial on life or matter but not both. Each tutorial began with definitions of key characteristics of the domain, followed by a brief video that illustrated those characteristics with examples. The tutorials then addressed common misconceptions about the domain, followed by videos that explained why those misconceptions are false. 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 explained that being alive is not the same as being able to move and provided examples of living things that do not move (e.g., moss) and moving things that are not alive (e.g., comets). The tutorial on matter emphasized that all matter occupies space, has weight, is made of atoms and can undergo phase transitions. It explained that being able to see something or feel something is not evidence that it is composed of matter and provided examples of matter that cannot be perceived (e.g., air) and perceptible phenomena that are not material (e.g., lightning). Tutorials took approximately seven minutes to complete.

Prior to the tutorial, participants verified 120 statements about life and 120 statements about matter. Following the tutorial, they verified another 120 statements in each domain. The posttest statements involved the same predicates but were paired with different subjects. Statements were generated by pairing three predicates in each domain with 80 entities. The predicates pertaining to life were "reproduces," "needs nutrients" and "grows and develops," and the predicates pertaining to matter were "has weight," "takes up space" and "is made of atoms." Statements involving the same predicate were presented in blocks but were randomized within the block. Statements were also randomized across tests, to ensure that the pretest and the posttest were comparably difficult.

We predicted that participants would verify counterintuitive scientific statements more quickly and more accurately after receiving a tutorial but only within the domain of instruction. Consistent with this prediction, participants who received a tutorial on life verified counterintuitive statements about life more accurately at posttest, and participants who received a tutorial on matter verified counterintuitive statements about matter more accurately at posttest, but no effect was found for the opposite domain. In other words, tutorials selectively reduced the gap in accuracy between intuitive and counterintuitive statements for the domain they targeted. They did not, however, reduce the gap in speed. This gap remained constant from pretest to posttest, regardless of instruction. Taken together, these findings indicate that instruction did not reduce the immediate conflict elicited by counterintuitive statements, as revealed by response lags, but did help participants favor scientific responses over intuitive ones.

In a follow-up study (Young & Shtulman, 2020b), we extended this line of research to elementary-school-aged children (ages 5–12). Our motivation was threefold. First, children are in the earliest stages of learning science, and it's unclear whether their nascent scientific theories would pose a demonstrable challenge to their intuitive theories of the same phenomena. Second, any conflict that children experience between science and intuition may be more malleable than that experienced by adults, either because children's scientific theories are less developed (and more easily bolstered) or because their intuitive theories are less entrenched. Third, adapting our task for use with children may have pedagogical value if it provides useful information about early science learning or early scientific reasoning.

The child version of our training study followed the same pre-post design as the adult version, though we reduced the number of statements from 480 to 192 (32 statements for each of three predicates in each of two domains). Children verified 48 statements in each domain at pretest, and then another 48 at posttest. Because children might not know the meanings of certain predicates, we defined each predicate on first introduction. "Reproduce" was defined as "things that can make more things like themselves," for instance, and "made of atoms" was defined as "things that are made of up of tiny pieces." Children completed the task on an iPad, responding via touch screen. Approximately half opted to listen to audio recordings of the statements, obviating the need to read them. Children completed the same tutorials as our adult participants because we designed them to be comprehensible to participants of all ages.

On the whole, children performed more slowly and less accurately than adults, but the same signatures of cognitive conflict were observed. They verified counterintuitive statements more slowly and less accurately than intuitive ones in both domains (life and matter) and at both tests (pretest and posttest). Children's accuracy at verifying counterintuitive statements improved from pretest to posttest but only within the domain of instruction. The speed of their responses did not change. They were slower to verify counterintuitive statements than closely matched intuitive ones regardless of instruction, similar to adults. These findings indicate that, despite minimal levels of science instruction, elementary schoolers experience the same conflict between science and intuition as experienced by adults. This conflict can be resolved in favor of scientific ideas with targeted instruction, but the conflict itself appears unmalleable. Counterintuitive scientific ideas elicit slower responses even among those whose intuitive theories are much less entrenched.

The role of language and perception

If the conflict between science and intuition persists across concepts, contexts, age and experience, what sustains it? Why are intuitive ideas so robust? One possibility is that they are better aligned with how we talk about natural phenomena in everyday contexts and are thus reinforced by such language. We describe coats as "warm" even though the warmth we experience when wearing them comes from our own bodies; a better label would be "insulating." We describe wind as "cold" even though its coldness is just a disruption of our own thermal equilibrium; a better label would be "disequilibrating." When we see meteors burn up in the Earth's atmosphere, we mislabel them as "shooting stars," and when the Sun recedes from view at the end of the day, we label the event a "sunset" even though "sun occlusion" would be more accurate. Phrases like "sun occlusion" and "disequilibrating wind" are unlikely to infiltrate common discourse, let alone eclipse their intuitive counterparts.

Scientific terms often elude common discourse because their referents elude perception. Intuitive ideas are usually better aligned with perception than scientific ones. It's not mere happenstance that we call coats warm; they feel warm. Wind feels cold. Stars appear to shoot across the sky. And the Sun appears to set behind the horizon. We may know full well that the Earth is moving, not the Sun, but we don't feel the Earth's motion, nor can we easily adopt the perspective of being situated on a revolving sphere (Jee & Anggoro, 2019).

To explore whether the conflict between science and intuition is driven by everyday language or everyday perception, we compared the speed and accuracy of participants' statement verifications to two measures of the statements' properties: (1) ratings of how strongly the statements' subjects embody perceptual properties intuitively associated with the domain and (2) estimates of how often the statements' subjects co-occur with their predicates in English-language documents (Shtulman & Legare, 2020). We sought to determine whether the signatures of cognitive conflict in a statement-verification task – decreased accuracy and increased response time – are better explained by how we perceive the statements' subjects or how we talk about those subjects in relation to their predicates.

This study required expanding our battery of scientific statements to increase the variance associated with each predicate. We constructed 720 statements about life and matter by pairing 80 subjects with 9 predicates. Our predicates pertaining to life were "is alive," "has cells," "has DNA," "excretes waste," "respires," "reproduces," "needs nutrients," "needs water" and "is adapted to the environment." Our predicates pertaining to matter were "is composed of matter," "occupies space," "contains atoms," "has weight," "has momentum," "has a density," "has a temperature," "has a molecular structure" and "can be put in a container." Of the 1,440 statements created in this manner, we had participants verify a random subset of 400 (200 per domain). We calculated participants' mean accuracy and mean response time for each statement and then compared those means to our measures of how well the statements resonate with everyday language and everyday perception.

To assess how language influences verification behavior, we searched the EBSCOhost database for our statements' subjects and predicates. This database indexes millions of English language documents, including journals, books, magazines and newspapers. It was selected for its breadth, as well as the precision of its search results. Google indexes a larger database, but its search results are only estimates and can lead to logical contradictions. For each of our 1,440 statements, we computed the proportion of EBSCOhost records containing the statement's subject that also contained its predicate. That is, for each statement, we divided the number of EBSCOhost records containing both the subject and the predicate by the number containing only the subject. We used proportions rather than absolute frequencies to control for differences in the total number of records on a given subject. For instance, EBSCOhost indexed 14,218 records containing the word "crocodile," and 196 of those records also contained the word "alive." Dividing the latter by the former yielded a co-occurrence estimate of .014, indicating that "alive" co-occurs with "crocodile" in 1.4% of documents that reference crocodiles.

Co-occurrence estimates were roughly consistent with intuitive conceptions of the statements' subjects. For instance, the word "alive" co-occurred with animals (such as crocodiles, turtles and owls) more often than plants (such as willows, oaks and grass) and co-occurred with non-living entities that move on their own (such as fire, geysers and comets) more often than nonliving entities that do not (such as mittens, boulders and tables), consistent with the intuition that life is synonymous with motion.

To assess how perception influences verification behavior, we collected ratings from independent judges on how well the subjects of our statements embody perceptual attributes intuitively associated with the domain. For biological items, we asked participants to rate whether each item appears to move on its own, have goals and sense its surroundings – three properties true of animals but not all living things. For physical items, we asked participants to rate whether each item can be seen, felt or lifted – three properties true of solid objects but not all material substances. Ratings were collected on a five-point scale, from "definitely" to "definitely not." They were averaged across the three attributes to derive a single composite measure. The internal reliability of these composites was near ceiling, indicating that there is strong agreement as to whether an entity resembles an animal (the core ontology of an intuitive theory of life) or a solid object (the core ontology of an intuitive theory of matter).

We expected that lexical co-occurrence would pull participants toward judging the statement as true. If "matter" co-occurs with "log" more often than "fog," then "matter" and "log" should be more closely associated in participants' minds, and "logs are composed of matter" should be verified more quickly and more accurately than "fog is composed of matter." We expected that higher perceptual ratings would also pull participants toward judging a statement as true. If logs are rated as more tangible than fog, then logs should be viewed as more material and "logs are composed of matter" should be verified more quickly and more accurately than "fog is composed of matter." The reverse logic held for false statements. Lexical co-occurrence and perceptual ratings would assist in evaluating unambiguously false statements ("numbers are composed of matter") relative to statements that are scientifically false but intuitively true ("heat is composed of matter") because the subject of the latter ("heat") would be more closely associated with the corresponding predicate ("matter"), both lexically and perceptually.

Replicating previous findings, we found that participants verified counterintuitive statements less accurately and more slowly than intuitive statements involving the same predicate. Extending those findings, we found that the speed and accuracy of participants' verifications from one statement to another correlated with our measures of everyday language (co-occurrence estimates) and everyday perception (attribute ratings). Co-occurrence estimates explained 10% of the variance in how accurately participants verified different versions of the same statement and 5% of the variance in how quickly they made their verifications. Attribute ratings explained 35% of the variance in participants' accuracy and 26% of the variance in their speed. These effects were comparable for statements about life and statements about matter.

These results suggest that the reason a statement like "coral is alive" is more difficult to evaluate than a statement like "crocodiles are alive" is that we perceive crocodiles as alive and talk about crocodiles as alive but are less inclined to do the same for coral. In our minds, coral is more closely associated with nonliving things than living things, both perceptually and linguistically. That said, our measure of everyday perception explained several times more variance in the speed and accuracy of participants' verifications than our measure of everyday language, suggesting that language and perception do not stand on equal footing. Both may play a role in sustaining intuitive theories – their influence is not mutually exclusive – but perception appears to play a much larger role, possibly because many of our perceptual biases are innate (Carey, 2009).

The role of cognitive reflection

Accurately verifying a counterintuitive scientific idea not only requires knowing the relevant science, but it also requires identifying and suppressing a conflicting intuition. The latter ability has been studied as a form of cognitive reflection, or the disposition to reflect on a cognitively intuitive response to derive a more accurate response in its place. Cognitive reflection is a domain-general disposition that predicts several domain-specific skills, including causal reasoning (Don et al., 2016), moral reasoning (Royzman et al., 2014) and vigilance against falsehoods (Pennycook & Rand, 2018). The standard measure of cognitive reflection is Frederick's (2005) Cognitive Reflection Test, or CRT, which consists of three problems designed to elicit an erroneous intuition that can be corrected upon further reflection, including this one:

> In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

The correct answer is 47, since the lake would be half covered one day prior to being fully covered, but most people divide 48 by two and provide the incorrect answer of 24.

Performance on the CRT predicts adults' understanding of difficult science topics (Shtulman & McCallum, 2014). Cognitively reflective adults perform more accurately on tests of astronomy, evolution, geology, mechanics, perception and thermodynamics than their less reflective peers, and this effect holds even when controlling for prior science and math coursework, statistical reasoning ability and explicit understanding of the nature of science. Cognitive reflection may thus facilitate conceptual change. All of us reason *with* our concepts, but conceptual change requires reasoning *about* our concepts, and cognitive reflection may be required for the latter. Research on a related construct – executive function – indicates that students with higher executive function skills, particularly inhibition and cognitive flexibility, are more likely to undergo conceptual change when learning counterintuitive concepts in biology (Bascandziev et al., 2018; Tardiff et al., 2020). Still, the finding that cognitively reflective adults exhibit superior scientific reasoning does not necessarily mean that cognitive reflection facilitates science learning; it may instead facilitate the deployment of scientific concepts once learned, especially when those concepts conflict with pre-existing intuitions.

My colleagues and I sought to disentangle these possibilities by assessing whether cognitive reflection facilitates performance in our training studies (Young & Shtulman, 2020b). We focused on children, who are in the midst of revising their intuitive theories and might thus show the greatest effects of cognitive reflection on learning. Children's cognitive reflection cannot be measured with the original CRT (Frederick, 2005) because it involves mathematical operations beyond the grasp of early elementary schoolers. We instead used a measure of cognitive reflection designed specifically for children – the CRT-D ("D" for developmental) – which uses verbal brain teasers instead of math problems (Young & Shtulman, 2020a). Each item on this nine-item test is designed to elicit an erroneous response that even children can correct upon further reflection. A sample item is "What do cows drink?" The correct answer is water, but cow's association with dairy leads many children (and some adults) to answer milk instead.

We administered the CRT-D prior to the statement-verification task and found that the former strongly predicted the latter. Children with higher CRT-D scores were faster and more accurate at verifying counterintuitive scientific statements than those with lower scores, at both pretest and posttest. They also exhibited significant pre–post gains in accuracy, indicating they had learned more from the tutorial. This learning gain held even when controlling for children's age. Cognitive reflection thus appears to play a dual role in scientific reasoning. It facilitates learning counterintuitive scientific concepts as well as deploying those concepts once learned.

Similar results have been obtained in research on executive function. Executive function skills facilitate both science learning (Bascandziev et al., 2018; Tardiff et al., 2020) and scientific reasoning (Vosniadou et al., 2018) in domains where the relevant science conflicts with intuition. The similarity is not coincidental. Cognitive reflection draws upon many executive function skills, including inhibition and cognitive flexibility. Responding accurately to a CRT brain teaser requires inhibiting an intuitive response, as well as shifting between an intuitive response and an analytic one. But cognitive reflection also requires the insight that these skills are necessary, the motivation to use them and the ability to coordinate them in real time.

Indeed, cognitive reflection is empirically distinct from executive function, predicting aspects of conceptual development that executive function does not (Young & Shtulman, 2020a), but more research is needed to determine what makes cognitive reflection unique. Cognitively reflective individuals may be better at identifying gaps in their understanding, or they may be better at filling those gaps with new information. They may be more receptive to instruction, or they may be better at monitoring and resolving response conflicts. We suspect that cognitive reflection is valuable because it fosters metaconceptual awareness. Students who are able to recognize the limitations of their intuitive theories, along with the advantages of a scientific alternative, may be more motivated to learn and use the latter (as Bélanger also argues in this volume, when considering the metacognitive value of different explanatory relationships). Pedagogically, our findings imply that instructors could use the CRT-D as a diagnostic for determining who is likely to profit from instruction and who may need additional support. Instructors might also use the CRT-D to teach cognitive reflection, as an important skill unto itself.

Conclusions

Conflict between science and intuition is a seemingly universal outcome of learning science. It emerges early in the acquisition of scientific knowledge (Young & Shtulman, 2020b), pervades scientific reasoning in different contexts and content domains (Shtulman & Valcarcel, 2012), and persists across targeted training (Young et al., 2018) and the development of scientific expertise more generally (Shtulman & Harrington, 2016). The resilience of intuitive ideas appears to be grounded in everyday language and everyday perception (Shtulman & Legare, 2020), which resonate with intuition more than science.

Our findings indicate that priming and training can help reasoners favor scientific ideas over intuitive ones, but the conflict elicited by counterintuitive scientific ideas cannot be eliminated. Counterintuitive statements like "yeast needs nutrients" or "clouds have weight" elicit contradictory responses – science says "true" but intuition says "false" – and it takes us appreciably longer to select the correct, scientific response than in situations where science and intuition agree. Priming and training increase the likelihood that participants will select the correct response, but it does not change how quickly that response is selected, implying that both responses are elicited automatically. These results parallel those obtained with professional scientists (Goldberg & Thompson-Schill, 2009; Kelemen et al., 2013; Potvin et al., 2020), who are more accurate than non-scientists at affirming counterintuitive scientific ideas but still exhibit reliable response lags when evaluating counterintuitive ideas. While the conflict between science and intuition may be inevitable, appropriate resolution of that conflict can be facilitated by reflecting on gut responses (Young & Shtulman, 2020b). Fostering cognitive reflection may thus be critical for improving science education and scientific literacy. Cognitively reflective individuals are not only better equipped to learn science but also more likely to use that knowledge in lieu of naïve intuitions, developed in childhood but retained across the lifespan.

Acknowledgment: This research was supported by National Science Foundation grant DRL-0953384 and James S. McDonnell Foundation grant 220020425.

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*(9), 1163–1178. https://doi.org/10.1002/jnr.24435

Au, T. K.-F., Chan, C. K. K., Chan, T.-K., Cheung, M. W. L., Ho, J. Y. S., & Ip, G. W. M. (2008). Folkbiology meets microbiology: A study of conceptual and behavioral change. *Cognitive Psychology*, *57*(1), 1–19. https://doi.org/10.1016/j.cogpsych.2008.03.002

Babai, R., Sekal, R., & Stavy, R. (2010). Persistence of the intuitive conception of living things in adolescence. *Journal of Science Education and Technology*, *19*(1), 20–26. https://doi.org/10.1007/s10956-009-9174-2

Barlev, M., Mermelstein, S., & German, T. C. (2017). Core intuitions about persons coexist and interfere with acquired Christian beliefs about God. *Cognitive Science*, *41*(S3), 425–454. https://doi.org/10.1111/cogs.12435

Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018). The role of domain-general cognitive resources in children's construction of a vitalist theory of biology. *Cognitive Psychology*, *104*, 1–28. https://doi.org/10.1016/j.cogpsych.2018.03.002

Brault Foisy, L.-M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, *4*(1–2), 26–36. https://doi.org/10.1016/j.tine.2015.03.001

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology*, *21*(1), 13–19. https://doi.org/10.1016/S0193-3973(99)00046-5

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Chi, M. T. H. (2008). 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. 61–82). Routledge.

Don, H. J., Goldwater, M. B., Otto, A. R., & Livesey, E. J. (2016). Rule abstraction, model-based choice, and cognitive reflection. *Psychonomic Bulletin & Review*, *23*, 1615–1623. https://doi.org/10.3758/s13423-016-1012-y

Evans, E. M., Spiegel, A. N., Gram, W., Frazier, B. N., Tare, M., Thompson, S., & Diamond, J. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching*, *47*(3), 326–353. https://doi.org/10.1002/tea.20337

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental 'roots' in mature biological knowledge. *Psychological Science*, *20*(4), 480–487. https://doi.org/10.1111/j.1467-9280.2009.02320.x

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Harris, P. L., & Gimenez, M. (2005). Children's acceptance of conflicting testimony: The case of death. *Journal of Cognition and Culture*, *5*(1–2), 143–164. https://doi.org/10.1163/1568537054068606

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, *50*(1–3), 171–188. https://doi.org/10.1016/0010-0277(94)90027-2

Järnefelt, E., Canfield, C. F., & Kelemen, D. (2015). The divided mind of a disbeliever: Intuitive beliefs about nature as purposefully created among different groups of non-religious adults. *Cognition*, *140*, 72–88. https://doi.org/10.1016/j.cognition.2015.02.005

Jee, B. D., & Anggoro, F. K. (2019). Relational scaffolding enhances children's understanding of scientific models. *Psychological Science*, *30*(9), 1287–1302. https://doi.org/10.1177/0956797619864601

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*(4), 1074–1083. https://doi.org/10.1037/a0030399

Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, *8*, 439–453. https://doi.org/10.3758/BF03196179

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

Legare, C. H., & Shtulman, A. (2018). Explanatory pluralism across cultures and development. In J. Proust & M. Fortier (Eds.), *Metacognitive diversity: An interdisciplinary approach* (pp. 415–432). Oxford University Press.

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*(6), 657–677. https://doi.org/10.1002/tea.3660310607

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*(11), 999–1006. https://doi.org/10.1111/j.1467-9280.2007.02015.x

Mason, L., & Zaccoletti, S. (2020). Inhibition and conceptual learning in science: A review of studies. *Educational Psychology Review*, *33*, 181–212. https://doi.org/10.1007/s10648-020-09529-x

Masson, S., Potvin, P., Riopel, M., & Brault Foisy, L.-M. (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*(1), 44–55. https://doi.org/10.1111/mbe.12043

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Erlbaum.

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. https://doi.org/10.1016/j.nlm.2016.08.020

Murray, G. W., Arner, T., Roche, J. M., & Morris, B. J. (2020). Using neuromyths to explore educator cognition: A mouse-tracking paradigm. *Proceedings of the 42nd Conference of the Cognitive Science Society*, 1335–1341.

Nakhleh, M. B., Samarapungavan, A., & Saglam, Y. (2005). Middle school students' beliefs about matter. *Journal of Research in Science Teaching*, *42*(5), 581–612. https://doi.org/10.1002/tea.20065

Pennycook, G., & Rand, D. G. (2018). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

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*(9), 1121–1142. https://doi.org/10.1002/tea.21396

Potvin, P., Malenfant-Robichaud, G., Cormier, C., & Masson, S. (2020). Coexistence of misconceptions and scientific conceptions in chemistry professors: A mental chronometry and fMRI Study. *Frontiers in Education*, *5*(180). https://doi.org/10.3389/feduc.2020.542458

Preston, J., & Epley, N. (2009). Science and God: An automatic opposition between ultimate explanations. *Journal of Experimental Social Psychology*, *45*(1), 238–241. https://doi.org/10.1016/j.jesp.2008.07.013

Preston, J. L., Ritter, R. S., & Hepler, J. (2013). Neuroscience and the soul: Competing explanations for the human experience. *Cognition*, *127*(1), 31–37. https://doi.org/10.1016/j.cognition.2012.12.003

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naive physics reasoning: A commitment to substance-based conceptions. *Cognition and Instruction*, *18*(1), 1–34. https://doi.org/10.1207/S1532690XCI1801_01

Rosengren, K. S., Miller, P. J., Gutiérrez, I. T., Chow, P. I., Schein, S. S., & Anderson, K. N. (2014). Children's understanding of death: Toward a contextualized and integrated account. *Monographs of the Society for Research in Child Development*, *79*, 1–141.

Royzman, E. B., Landy, J. F., & Leeman, R. F. (2014). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science*, *39*(2), 325–352. https://doi.org/10.1111/cogs.12136

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

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*(1), 118–137. https://doi.org/10.1111/tops.12174

Shtulman, A., & Legare, C. H. (2020). Competing explanations of competing explanations: Accounting for conflict between scientific and folk explanations. *Topics in Cognitive Science*, *12*(4), 1337–1362. https://doi.org/10.1111/tops.12483

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 53–72). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467630.003.0004

Shtulman, A., & McCallum, K. (2014). Cognitive reflection predicts science understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 2937–2942. https://escholarship.org/uc/item/4t79p8pj

Shtulman, A., Neal, C., & Lindquist, G. (2016). Children's ability to learn evolutionary explanations for biological adaptation. *Early Education and Development*, *27*(8), 1222–1236. https://doi.org/10.1080/10409289.2016.1154418

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209–215. https://doi.org/10.1016/j.cognition.2012.04.005

Shtulman, A., & Walker, C. M. (2020). Developing an understanding of science. *Annual Review of Developmental Psychology*, *2*, 111–132. https://doi.org/10.1146/annurev-devpsych-060320-092346

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories. *Cognition and Instruction*, *25*(4), 337–398. https://doi.org/10.1080/07370000701632363

Solomon, G. E. A., & Zaitchik, D. (2012). Folkbiology. *Wires Cognitive Science*, *3*(1), 105–115. https://doi.org/10.1002/wcs.150

Tardiff, N., Bascandziev, I., Carey, S., & Zaitchik, D. (2020). Specifying the domain-general resources that contribute to conceptual construction: Evidence from the child's acquisition of vitalist biology. *Cognition*, *195*, 104090. https://doi.org/10.1016/j.cognition.2019.104090

Toyama, N. (2019). Development of integrated explanations for illness. *Cognitive Development*, *51*, 1–13. https://doi.org/10.1016/j.cogdev.2019.05.003

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3

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*(6), 1860–1886. https://doi.org/10.1111/cogs.12624

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 177–210). Macmillan.

Young, A. G., Laca, J., Dieffenbach, G., Hossain, E., Mann, D., & Shtulman, A. (2018). Can science beat out intuition? Increasing the accessibility of counter-intuitive scientific ideas. *Proceedings of the 40th Conference of the Cognitive Science Society*, 1236–1241.

Young, A. G., & Shtulman, A. (2020a). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychological Science*, *31*(11), 1396–1408. https://doi.org/10.1177/0956797620954449

Young, A. G., & Shtulman, A. (2020b). How children's cognitive reflection shapes their science understanding. *Frontiers in Psychology*, *11*, 1247. https://doi.org/10.3389/fpsyg.2020.01247