TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 12 (2020) 1337–1362

ISSN:1756-8765 online
DOI: 10.1111/tops.12483

This article is part of the topic "Levels of Explanation in Cognitive Science: From Molecules to Culture," Matteo Colombo and Markus Knauff (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

Competing Explanations of Competing Explanations: Accounting for Conflict Between Scientific and Folk Explanations

Andrew Shtulman,[a] Cristine H. Legare[b]

[a]*Department of Psychology, Occidental College*
[b]*Department of Psychology, The University of Texas at Austin*

Received 25 January 2019; received in revised form 7 October 2019; accepted 29 October 2019

Abstract

People who hold scientific explanations for natural phenomena also hold folk explanations, and the two types of explanations compete under some circumstances. Here, we explore the question of why folk explanations persist in the face of a well-understood scientific alternative, a phenomenon known as *explanatory coexistence*. We consider two accounts: an associative account, where coexistence is driven by low-level associations between co-occurring ideas in experience or discourse, and a theory-based account, where coexistence reflects high-level competition between distinct sets of causal expectations. We present data that assess the relative contributions of these two accounts to the cognitive conflict elicited by counterintuitive scientific ideas. Participants (134 college undergraduates) verified scientific statements like "air has weight" and "bacteria have DNA" as quickly as possible, and we examined the speed and accuracy of their verifications in relation to measures of associative information (lexical co-occurrence of the statements' subjects and predicates) and theory-based expectations (ratings of whether the statements' subjects possess theory-relevant attributes). Both measures explained a significant amount of variance in participants' responses, but the theory-based measures explained three to five times more. These data suggest that the cognitive conflict elicited by counterintuitive scientific ideas typically arises from competing theories and that such ideas might be made more intuitive by strengthening scientific theories or weakening folk theories.

Keywords: Explanation; Explanatory coexistence; Scientific reasoning; Intuitive theories

Correspondence should be sent to Cristine H. Legare, Department of Psychology, The University of Texas at Austin, 108 E. Dean Keeton Street, Austin, TX 78712. E-mail: legare@austin.utexas.edu

1. Introduction

Natural phenomena can be explained in many ways, by appeal to folk beliefs, religious beliefs, or scientific beliefs. Sunrises can be explained as the sun orbiting the earth, a god pulling the sun across the sky, or the earth rotating on its axis. Illness can be explained as a consequence of getting cold, as divine punishment for wrongdoing, or as a viral infection. Biological adaptation can be explained as the inheritance of acquired traits, the design of a divine creator, or the outcome of natural selection. Death can be explained as the loss of vital energy, the departure of the soul, or the failure of internal organs.

There are individual differences in the kinds of explanations people favor, but recent research in cognitive development, cognitive neuroscience, and science education suggests that people frequently consider multiple kinds of explanations for the same outcome, event, or process (Busch, Watson-Jones, & Legare, 2017; Legare & Shtulman, 2018; Potvin, Sauriol, & Riopel, 2015; Shtulman & Lombrozo, 2016; Watson-Jones, Busch, & Legare, 2015). This tendency to represent multiple explanations of the same phenomenon is known as *explanatory coexistence* (Evans et al., 2010; Legare & Gelman, 2008; Legare & Visala, 2011). A religious individual might favor a creationist explanation for the origin of species but also understand an evolutionary explanation, and a secular individual might favor an evolutionary explanation but also understand a creationist one (Järnefelt, Canfield, & Kelemen, 2015; Legare, Evans, Rosengren, & Harris, 2012; Tracy, Hart, & Martens, 2011). Alternative explanations are not merely known to the explainer; they shape the inferences we draw and the information we accept as true. Few people, for instance, reason about the origin of species in purely evolutionary terms or purely creationist terms. Rather, most evoke both types of explanations (Shtulman, Neal, & Lindquist, 2016; Weisberg, Landrum, Metz, & Weisberg, 2018) and accept both types of claims (Evans et al., 2010).

In this paper, we seek to explain the phenomenon of competing explanations, focusing on competition between scientific explanations and folk explanations. Previous research indicates that the two types of explanations not only coexist in people's minds but actively compete to provide interpretations of the same information or outcome. That is, people not only switch between explanations depending on the context (Harris & Giménez, 2005; Lane, Zhu, Evans, & Wellman, 2016; Legare & Gelman, 2008; Watson-Jones, Busch, Harris, & Legare, 2017) but also experience conflict over which explanation for an outcome, event, or process is correct. This conflict manifests itself as differences in how quickly and how accurately people verify scientific ideas that are consistent with folk explanations relative to those that are inconsistent with them (Goldberg & Thompson-Schill, 2009; Kelemen & Rosset, 2009; Potvin & Cyr, 2017; Shtulman & Valcarcel, 2012). For instance, people are slower and less accurate when verifying "oaks are alive" than when verifying "owls are alive" because the former is consistent with both scientific and folk explanations of biological observations but the latter is consistent with scientific explanations but not folk explanations.

Here, we explore two factors that may be responsible for such difficulty: low-level associations and high-level theories. Cognitive conflict may arise because the words used to express counterintuitive scientific ideas are less associated with one another than the words used to express more intuitive ideas (e.g., the word "alive" is less associated with plants than animals), but it might also arise because counterintuitive scientific ideas evoke alternative theories of the domain whereas more intuitive ideas do not (e.g., intuitive theories of life treat animals as alive but not plants). Associations are a pervasive consequence of how we experience and talk about the world, but so are intuitive theories, which we form early in development as a way of organizing and interpreting our perceptions of natural phenomena (Shtulman, 2017). Here, we attempt to tease apart the influence of low-level associations and high-level theories by measuring the speed and accuracy with which people verify a wide range of scientific statements, from highly intuitive statements to highly counterintuitive ones, and comparing those data to measures of (a) how strongly the statements' subjects and predicates are associated in everyday language and (b) how strongly the statements accord with an alternative, nonscientific theory of the domain. To preview our results, we find that cognitive conflict correlates with both measures but is more strongly correlated with the latter, theory-based measure.

1.1. Explanatory coexistence

The psychological coexistence of qualitatively distinct explanations has been documented in several domains, including illness (Legare & Gelman, 2008; Nguyen & Rosengren, 2004), death (Harris & Giménez, 2005), consciousness (Preston, Ritter, & Hepler, 2013), and cosmology (Preston & Epley, 2009). It has also been documented in people of various ages, from children (Vosniadou et al., 2018) to adolescents (Babai, Sekal, & Stavy, 2010) to elderly adults (Barlev, Mermelstein, & German, 2018), and people from diverse populations, from South Africa (Legare & Gelman, 2008) to South Asia (Raman & Gelman, 2004) to the Pacific Islands (Busch, Watson-Jones, & Legare, 2017; Watson-Jones et al., 2017; Watson-Jones, Busch, & Legare, 2015). Explanatory coexistence appears to be an inherent feature of how humans represent and reason about the natural world.

Sometimes people manage to reconcile the conflict between seemingly incompatible explanations, by treating one explanation as a proximate cause and the other as a distal cause (e.g., treating evolution as the proximate cause of biological adaptation and God as the distal cause) or by relegating each explanation to different aspects of the target phenomena (e.g., relegating evolution to the origin of "lower organisms" and relegating God to the origin of humans; Gelman & Legare, 2011). But this type of metacognitive activity may not be common across people or domains, as it requires reflection on one's explanatory practices and the wherewithal to reconcile discrepancies between them (Legare & Shtulman, 2018).

Here, we focus on the coexistence of scientific and folk explanations, as opposed to scientific and religious explanations or religious and superstitious explanations (Shtulman

& Young, 2020), because this type of coexistence is particularly difficult to reconcile. People reliably experience cognitive conflict when reasoning about phenomena for which scientific explanations and folk explanations diverge. For instance, when deciding whether something is alive, people are slower and less accurate at classifying plants as alive than at classifying animals as alive (Babai et al., 2010; Goldberg & Thompson-Schill, 2009), presumably because science explains life as metabolic processing but intuition explains life as the capacity for self-directed motion (Carey, 1985; Hatano & Inagaki, 1994). Animals are alive on both explanations but plants are alive only on the scientific explanation. Tensions between scientific and folk explanations also lead to conflict when classifying things as "not alive"; people are slower and less accurate at classifying moving objects, like rivers and clouds, as not alive than at classifying nonmoving objects, like rocks and boulders, as not alive.

Similar results have been found in the domain of matter (Young et al., 2018). People are slower and less accurate at classifying gases as matter than at classifying solids as matter, presumably because science explains matter as something composed of atoms but intuition explains matter as something that can be seen or felt (Nakhleh, Samarapungavan, & Saglam, 2005; Smith, 2007). Solids are material on both explanations, but gases are material only on the scientific explanation. In this same vein, people are reluctant to classify nonmaterial entities as "not matter" if those things can be seen or felt. Forms of energy like rainbows, echoes, and lightning are classified as "not matter" more slowly and less accurately than abstract concepts, like thoughts, numbers, and words.

At the neural level, the endorsement of counterintuitive scientific ideas is associated with higher activity in areas of the brain associated with error monitoring and inhibitory control—namely, the dorsolateral prefrontal cortex and the anterior cingulate cortex (Foisy, Potvin, Riopel, & Masson, 2015; Mareschal, 2016; Masson, Potvin, Riopel, & Foisy, 2014). People who are able to distinguish functional electric circuits from plausible yet nonfunctional circuits (physics experts) show more activation in these areas than people who are unable to distinguish the two (physics novices). Similar results have been found for people who are able to distinguish accurate motion events from plausible yet inaccurate events (a heavy ball falling to the ground faster than a light ball). Privileging a scientific interpretation of the situation over a folk interpretation requires actively noticing and inhibiting the latter.

These findings indicate that people experience cognitive conflict when reasoning about phenomena interpretable by both scientific and folk explanations, but they leave open the question of what these explanations are like, either in form or function. Colombo (2017), drawing on the philosophy of science, identifies three models of explanation: a deductive-nomological model, where explanations demonstrate how the explanandum follows from general laws or empirical regularities; a unificationist model, where explanations unify the explanandum with a general pattern that applies to many instances of the phenomenon; and a causal-mechanical model, where explanations point to factors that brought about the explanandum or caused it to occur.

Folk explanations are primarily concerned with causality (Gelman & Legare, 2011; Gopnik & Wellman, 2012; Shtulman, 2017) and thus accord with the causal-mechanical model better than the other models. It is an open question which model best describes everyday scientific explanations—the scientific explanations deployed outside of formal science. We suspect the kinds of scientific considerations that a nonscientist brings to bear on questions about the natural world have a causal-mechanical basis as well, though it's possible that scientific explanations are more deductive in nature. While the study reported here does not address whether scientific and folk explanations overlap in form, it does provide evidence that they overlap in function, consistent with other research (Gelman & Legare, 2011; Legare & Shtulman, 2018). When considering whether something is alive or whether something is material, people appear to draw on both folk explanations and scientific explanations; the less the two converge, the longer it takes people to respond to the query.

1.2. Explaining explanatory coexistence: High-level theories and low-level associations

To date, studies of explanatory coexistence have focused on demonstrating that scientific and folk explanations coexist in the same mind, but they have not addressed the question of why. One possibility is that folk explanations are grounded in higher-level representations that are themselves difficult to dislodge and displace: intuitive theories. Intuitive theories are coherent sets of domain-specific causal expectations (Carey, 2009; Keil, 1992; Vosniadou, 1994; Wellman & Gelman, 1992). They arise early in development and take similar forms across cultures, helping us understand phenomena as diverse as heat, motion, gravity, growth, inheritance, and ancestry (Shtulman, 2017). Intuitive theories serve the same function as scientific theories, supporting explanation, prediction, intervention, and counterfactual reasoning (Gopnik & Wellman, 2012; Murphy & Medin, 1985), but they are less accurate and less precise.

Strong evidence for the stability of intuitive theories comes from Alzheimer's patients (Lombrozo, Kelemen, & Zaitchik, 2007; Zaitchik & Solomon, 2008). Like other adults, Alzheimer's patients are less accurate at verifying counterintuitive scientific ideas relative to intuitive ones, but their errors are more pronounced and more pervasive. When quizzed about biological phenomena, Alzheimer's patients define life in terms of motion rather than metabolic activity. They cite animals as examples of living things but rarely cite plants. And they judge nonliving things that move on their own—wind, rain, fire, sun, clouds—as alive. This response pattern is more characteristic of a four-year-old child than an age-matched adult without Alzheimer's disease, implying that Alzheimer's patients are unable to access scientific theories of life and default to intuitive theories instead.

A different reason why folk explanations may continue to exist alongside scientific explanations is that folk explanations are grounded in environmentally pervasive and psychologically entrenched patterns of association, such as the association between "animal" and "alive" or the association between "object" and "matter." Histories of association may reflect the same conceptual regularities captured by intuitive theories, but they may also reflect lower-level regularities—regularities instantiated by mere co-occurrence. We describe associations as "low level" because they occur between the individual components of an explanation—the words they contain or ideas they connote—whereas we describe theories as “high level” because they constitute a broad system of causal relations from which particular explanations are derived. Consider the association between cows and milk. This association is so strong that most people judge “cows drink milk” as true, even though cows actually drink water (Young, Powers, Pilgrim, & Shtulman, 2018). This judgment does not arise from an intuitive theory of cows or an intuitive theory of milk but rather a history of association between cows and milk, established through experience and discourse.

Histories of association may underlie, at least in part, the statement-verification effects noted above. People may be slow to verify “the earth revolves around the sun,” relative to “the moon revolves around the earth,” because the earth is associated with stability and the sun and moon are associated with motion. The very terms “sunrise” and “sunset” imply that the sun is moving, not the earth. Likewise, people may be slow to verify “mushrooms are alive” relative to “tigers are alive” because the words “tiger” and “alive” are more strongly associated than the words “mushroom” and “alive.” The difference in response times would reflect differences in the efficiency of linguistic processing—differences that may operate independently of the conflict between scientific and intuitive theories of the relevant phenomenon (in this case, life).

High-level theories and low-level associations provide competing explanations of the cognitive conflict elicited by counterintuitive scientific ideas, as indexed by lower accuracy and longer response times when verifying such ideas. The theory-based account assumes that this conflict arises from an inferential process, where two or more abstract representations (such as geocentrism and heliocentrism) are triggered by a common idea (such as the claim that the earth revolves around the sun) and then compete to provide an interpretation or judgment. The theory-based account assumes this competition originates from learning a new theory, and its resolution would be to strengthen the new theory or weaken the old one. The associative account, on the other hand, assumes that cognitive conflict arises from distinct histories of association, with some histories reinforcing a scientific interpretation of the target phenomenon and others reinforcing a folk interpretation. Knowledge of nonscientific associations (such as that between “sun” and “motion”) would pull a person toward favoring a folk interpretation, whereas knowledge of scientific associations (such as that between “earth” and “motion”) would pull in the opposite direction.

The theory-based account and the associative account are not mutually exclusive. A person could hold distinct theories of the same phenomenon and also be exposed to distinct patterns of association, and the two influences could be mutually reinforcing. The persistence of intuition-laden terms like “sunrise” and “sunset” may reflect the persistence of an intuitive theory that licenses those terms (geocentrism). Intuitive theories that assume the sun moves around the earth may support, and be supported by, discourse about the sun moving, even if we tacitly recognize that such discourse is metaphoric. That said, it is unclear whether the signature of explanatory coexistence explored here—the decreased accuracy and increased response time associated with evaluating counterintuitive scientific ideas—is better explained by contradictory theories or contradictory associations. Are people worse at verifying counterintuitive scientific statements relative to intuitive ones because they are grappling with the statements' meaning? Or because the statements' words are just less associated with one another in everyday discourse? It is an empirical question which kind of information is more responsible for the effect, even if both are available.

The question of whether reasoning is better explained by associative knowledge or structured knowledge has a long history in cognitive psychology (Evans, 2008; Rogers & McClelland, 2004; Sloman, 1996; Sloutsky & Fisher, 2008). The answer is rarely one or the other but a combination of the two, depending on the type of reasoning and the context in which it is deployed. For instance, Bright and Feeney (2014) asked people to evaluate arguments like "mice have disease 3dfT; therefore, squirrels have disease 3dfT" under speeded and unspeeded conditions. They compared ratings of argument strength to ratings of how strongly the relevant categories are associated with one another (associative knowledge) and whether the two categories come from the same taxonomic group (taxonomic knowledge). They found that argument strength was better predicted by associative knowledge than by taxonomic knowledge when the arguments were evaluated under time pressure but was better predicted by taxonomic knowledge than by associative knowledge when there was no time pressure. Similar results were found when participants were placed under cognitive load, by performing a secondary task: Associative knowledge was a better predictor of argument strength under a heavy load, and structured knowledge was a better predictor under a light load.

We expected people's judgments of the truth of scientific statements to vary both by how strongly the statements' subjects and predicates are associated in everyday discourse (associative knowledge) and by how strongly the statements conform to intuitive theories of the relevant domain (structured knowledge), but it was an open question whether one form of knowledge would be more predictive than the other and by how much.

1.3. Comparing the influence of theories and associations

If both theories and associations contribute to conflict between explanations, which contributes more? We address that question in a parametric fashion, by asking adults with multiple years of college-level science education to verify a wide range of scientific statements and then comparing the speed and accuracy of their verifications to properties of the statements that correspond to the associative account and properties that correspond to the theory-based account. Our task, adapted from Shtulman and Valcarcel (2012), was to verify scientific statements as quickly as possible. Some statements were consistent with intuition, like "frogs reproduce" and "bricks have weight," whereas others were inconsistent with intuition, like "algae reproduce" and "snowflakes have weight." The intuition-consistent statements reflect situations in which scientific explanations and folk explanations coexist without competition, because both lead to the same judgment, whereas the intuition-inconsistent statements reflect situations in which the two compete, because they lead to different judgments. Science dictates that algae reproduces (because all living things reproduce) and snowflakes have weight (because all matter has weight), but intuition dictates that they do not.

Consistency with intuition was crossed with whether the statement was true from a scientific point of view, such that some statements were intuitively true ("logs are composed of matter"), some were intuitively false ("numbers are composed of matter"), some were counterintuitively true ("fog is composed of matter"), and some were counterintuitively false ("heat is composed of matter"). Participants judged each statement as true or false, and we refer to those judgments collectively as "verifications." We predicted that intuitive statements would be verified correctly more often than counterintuitive statements and that correct verifications for intuitive statements would also require less time.

Half of our statements pertained to life, and half pertained to matter. We chose these domains because they are foundational to scientific reasoning in general. Properties of life are foundational to higher-level concepts in cellular biology, evolutionary biology, and immunology, and properties of matter are foundational to higher-level concepts in physical chemistry, organic chemistry, and thermodynamics. Our participants had likely acquired a basic scientific understanding of these domains many years earlier, during elementary school in the case of life (Carey, 1985; Hatano & Inagaki, 1994; Stavy & Wax, 1989) and middle school in the case of matter (Nakhleh, Samarapungavan, & Saglam, 2005; Smith, 2007).

The statements were constructed by pairing a handful of scientific properties with dozens of different items, and the variance in participants' responses across statements was compared to two statement-specific measures: how often the statements' subjects co-occur with their predicates in a large corpus of English-language documents (our measure of lexical association) and ratings of how well the statements' subjects embody core properties of a domain-relevant intuitive theory (our measure of theory-based expectations).

These measures tap distinct kinds of information that may drive the cognitive conflict elicited by counterintuitive scientific ideas. Our first measure taps co-occurrence information, or how often the subjects of our statements are discussed in conjunction with science-relevant predicates. We expected that greater co-occurrence would be associated with a stronger pull to judge the statement as true, which would facilitate correct verifications for true statements but impede correct verifications for false statements. 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." On the other hand, if "matter" co-occurs with "heat" more often than "number," then "heat" and "matter" should be more closely associated in participants' minds, and "heat is composed of matter" should be judged false *less* quickly and *less* accurately than "numbers are composed of matter."

Our second measure taps conceptual information, or how participants routinely conceptualize the subjects of our statements. For biological statements, we collected ratings on how well their subjects accord with intuitive theories of life—whether they move on their own, whether they have goals, and whether they sense their surroundings. For physical statements, we collected ratings on how well their subjects accord with intuitive theories of matter—whether they can be seen, whether they can be felt, and whether they can be lifted. We expected that higher ratings would be associated with a stronger pull to judge the statement as true, which, once again, would facilitate correct verifications for true statements but impede correct verifications for false statements. If logs are rated as more tangible than fog, then logs should be more readily conceptualized as matter and "logs are composed of matter" should be verified more quickly and more accurately than "fog is composed of matter." On the other hand, if heat is rated as more tangible than numbers, then heat should be more readily conceptualized as matter, and "heat is composed of matter" should be judged false *less* quickly and *less* accurately than "numbers are composed of matter."

We expected that our measure of lexical association and our measure of theory-based expectation would both explain a significant amount of variance in the speed and accuracy of participants' statement verifications, but it was an open question whether one measure would explain more than the other.

2. Method

2.1. Participants

The participants were 134 college undergraduates recruited from the campus of Occidental College. They were compensated either monetarily or with extra credit in a psychology course. Their average age was 20.3 years, and they were predominately female (76% female). Fifty-six percent were majoring in a social science; 22%, a natural science; and 22%, the arts and humanities. Prior to the study, participants had taken an average of 4.3 college-level math and science courses.

2.2. Materials and procedure

Participants decided whether scientific statements like "bacteria have DNA" and "air contains atoms" were true or false as quickly as possible. Their instructions were as follows: "You will be shown 10 blocks of 40 statements, with each block pertaining to a different scientific question. Your task is to determine whether each statement is true or false as quickly as possible without sacrificing accuracy for speed; your judgments will be timed. Some of the statements you are about to see may be difficult to verify, and we do not expect you to know the correct answer for all 400. We simply urge you to try your best."

As noted above, we limited our statements to the domains of life and matter. Statements were created by pairing a handful of domain-specific predicates to a large number of subjects. For statements about life, we used the predicates "is alive," "has cells," "has DNA," "excretes waste," "respires," "reproduces," "needs nutrients," "needs water," and "is adapted to the environment." For statements about matter, we used the predicates "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." We thus used nine predicates per domain. Each predicate was paired with 80 items. Twenty of the items were classified as part of the domain by both science and intuition; 20 were classified as part of the domain by science but not intuition; 20 were classified as part of the domain by intuition but not science; and 20 were classified as part of the domain by neither science nor intuition. Sample items are presented in Table 1.

With respect to criteria for 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 that inanimate objects are not, albeit for different reasons, but disagree about the status of living but nonmoving objects (e.g., flowers, trees) and moving but nonliving objects (e.g., wind, 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 agree that solids are material and abstractions are immaterial but disagree about the status of material entities that cannot be perceived (e.g., gases, vapors) and immaterial entities that can be perceived (e.g., light, sound).

Nine predicates paired with 80 items yielded 720 statements per domain. To make the task manageable, we asked participants to verify only 200 statements per domain: five predicates paired with 40 items (10 of each type). The pairings were counterbalanced across participants to ensure that each unique pairing appeared as often as the others. Participants verified statements involving the same predicate in a block, but statements within each block were randomized, as were the blocks themselves. Participants completed a total of 10 blocks (five per domain) and thus verified a total of 400 statements. The entire study was run using MediaLab v1.21.

Table 1
Sample biological items and physical items, grouped by whether they are classified as part of the domain by intuition or by science

Domain	Intuition: Yes Science: Yes	Intuition: Yes Science: No	Intuition: No Science: Yes	Intuition: No Science: No
Biology	Crocodiles	Clocks	Algae	Caves
	Frogs	Comets	Bacteria	Forks
	Pelicans	Geysers	Coral	Hammers
	Rabbits	Rivers	Grass	Mittens
	Sharks	Robots	Mold	Mountains
	Snails	Satellites	Moss	Sand
	Spiders	The sun	Mushrooms	Shells
	Zebras	Tornadoes	Seaweed	Shovels
Physics	Bricks	Echoes	Air	Dreams
	Coal	Flames	Bubbles	Feelings
	Concrete	Heat	Clouds	Hours
	Diamonds	Lightning	Dust	Numbers
	Dumbbells	Rainbows	Fog	Songs
	Logs	Shadows	Smoke	Stories
	Rocks	Starlight	Snowflakes	Thoughts
	Steel	Thunder	Spores	Words

2.3. Measures of lexical association

To assess whether the variance in participants' responses tracked the variance in public discourse about the stimuli, we turned to the EBSCOhost research database. This database indexes millions of English-language documents, including journals, books, magazines, and newspapers. This database was selected for its breadth, as well as the precision of its search results. Searches in EBSCOhost provide the exact number of documents containing the search terms, as opposed to Google, which provide only an estimate. We began with Google, but our searches yielded contradictory results. For some subject–predicate pairs, the number of Google hits for both terms exceeded the number of hits for just the subject. This problem never surfaced in EBSCOhost.

EBSCOhost covers published documents, which are written more formally than the text available on most webpages and social media feeds. These documents may not capture the exact statistics of everyday speech, but they are a reasonable place to start, given the wide variety of subjects and predicates used as search terms. Searches that yielded zero hits would have been as unhelpful as searches that yielded contradictory results, and we encountered such searches only 2% of the time (57 out of 2,880 searches). EBSCOhost covers dozens of subjects, thousands of sources, and millions of documents, which made it large enough and diverse enough to sample language patterns relevant to our idiosyncratic goals.

For each of the 720 statements that participants verified, 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 multi-word predicates, we used only the main content term. The predicate "is composed of matter," for instance, was reduced to "matter." The predicate "needs nutrients" was reduced to "nutrients." Table 2 provides sample results for the predicate "is alive." These results are consistent with the conceptual status of the statement's subject. Animals co-occur with "alive" more than plants, and nonliving entities that move co-occur with "alive" more than nonliving entities that do not.

2.4. Measures of theory-based expectations

To assess the role of intuitive theories in participants' statement verifications, we collected ratings of how well the subjects of our statements embody the core attributes of a domain-relevant theory. For biological items, we asked participants to rate whether each item moves on its own, has goals, and senses its surroundings—three properties true of animals but not all living things (at least not perceptively). For physical items, we asked participants to rate whether each item can be seen, can be felt, and can be lifted—three properties true of solid objects but not all material substances. Note that these attributes differ substantively from the predicates used in the statement-verification task. They probe how the items are perceived rather than whether they embody properties true of the domain as a whole.

Table 2
Co-occurrence data for a sample of “is alive” statements: the number of EBSCOhost documents containing the statements’ subject, the number containing the predicate (“alive”) as well, and the proportion of the latter to the former

Statement	Subject	Subject + Predicate	Proportion
Crocodiles are alive.	14,218	196	0.0138
Tigers are alive.	277,152	2,550	0.0092
Turtles are alive.	24,145	201	0.0083
Owls are alive.	32,046	258	0.0081
Petunias are alive.	2,516	19	0.0076
Fire is alive.	881,091	5,404	0.0061
Willows are alive.	27,633	147	0.0053
Grass is alive.	176,970	921	0.0052
Geysers are alive.	2,597	13	0.0050
Oaks are alive.	162,783	777	0.0048
Comets are alive.	29,942	134	0.0045
Glaciers are alive.	24,377	103	0.0042
Stones are alive.	487,175	1,729	0.0036
Mittens are alive.	4,050	14	0.0035
Boulders are alive.	79,979	269	0.0034
Tables are alive.	354,057	619	0.0018

Ratings were collected on a 5-point scale, from “definitely” (5) to “definitely not” (1). Participants made these ratings following the statement-verification task and only for items they had not seen previously—that is, items assigned to other participants via counterbalancing. Ratings were averaged across the three attributes to derive a composite measure of how well the items accord with intuitive expectations in each domain. The internal reliability of these scales was near ceiling. Chronbach’s alpha for the biological items were 0.97 and 0.98 for each of the two counterbalancing sets. For the physical items, they were 0.89 and 0.90.

Table 3 displays ratings for the biological items included in Table 2. The ratings track the co-occurrence estimates: Animals were assigned higher ratings than plants, and nonliving entities that move were assigned higher ratings than nonliving entities that do not. The question explored below is which type of data—co-occurrence estimates or attribute ratings—explains the most variance in participants’ statement verifications.

3. Results

3.1. Main effects of statement type

Participants verified two types of statements—those where science and intuition yield the same truth-value (intuitive statements) and those where science and intuition yield different truth-values (counterintuitive statements)—for each of nine predicates in each of two

Table 3
Mean attribute ratings for the biological items from Table 2, ordered by their overall mean

Item	Moves On Its Own	Has Goals	Senses Its Surroundings	Overall Mean
Owls	5.0	4.6	5.0	4.9
Crocodiles	5.0	4.6	4.9	4.8
Turtles	4.9	4.6	4.9	4.8
Tigers	4.8	4.5	4.9	4.8
Petunias	3.6	3.1	3.7	3.5
Oaks	3.5	3.4	3.3	3.4
Willows	2.7	3.0	3.5	3.1
Grass	2.7	2.5	3.3	2.8
Fire	3.4	1.6	1.7	2.2
Glaciers	2.8	1.6	1.9	2.1
Geysers	2.8	1.7	1.7	2.1
Comets	3.1	1.4	1.4	2.0
Boulders	1.6	1.5	1.4	1.5
Stones	1.4	2.0	1.0	1.5
Mittens	1.1	1.5	1.3	1.3
Tables	1.1	1.3	1.0	1.1

domains. Mean accuracy for statement verifications was greater than 80% for 16 of the 18 predicates, and comparisons to chance (50%) revealed that statements with all 18 predicates were verified more accurately than not (all $t(79) > 2.8$, $p < .01$). This level of accuracy indicates that participants interpreted the task as intended and seem to have taken it seriously.

While overall accuracy was high, participants still verified intuitive statements more accurately than they verified counterintuitive ones, as shown in the top panel of Fig. 1. For each participant, we computed their mean accuracy for each type of statement (intuitive vs. counterintuitive) in each domain (biology vs. physics) and submitted those scores to a repeated-measures analyses of variance (ANOVA). This analysis confirmed that accuracy varied by statement type ($F(1, 133) = 501.0$, $p < .001$, partial $\eta^2 = 0.79$), as expected. Accuracy also varied by domain ($F(1, 133) = 164.0$, $p < .001$, partial $\eta^2 = 0.55$) and by the interaction of statement type and domain ($F(1, 133) = 164.0$, $p < .001$, partial $\eta^2 = 0.55$). The latter two effects indicate that statements about matter were more difficult for participants to verify than statements about life, particularly when those statements were counterintuitive. Participants may have been less knowledgeable about matter, or the statements about matter may have involved less familiar terminology. Either way, the simple effect of statement type was present in both domains (biology: $t(133) = 10.3$, $p < .001$, $d = 0.89$; physics: $t(133) = 23.2$, $p < .001$, $d = 2.00$).

Similar results were found for response latency, shown in the bottom panel of Fig. 1. We computed the mean latency for each type of statement (intuitive vs. counterintuitive) in each domain (biology vs. physics) and submitted those scores to a repeated-measures ANOVA. Only correct responses were included in this analysis, though the results do not change if incorrect responses are included as well. The ANOVA revealed significant effects of statement type ($F(1, 133) = 309.7$, $p < .001$, partial $\eta^2 = 0.70$) and domain

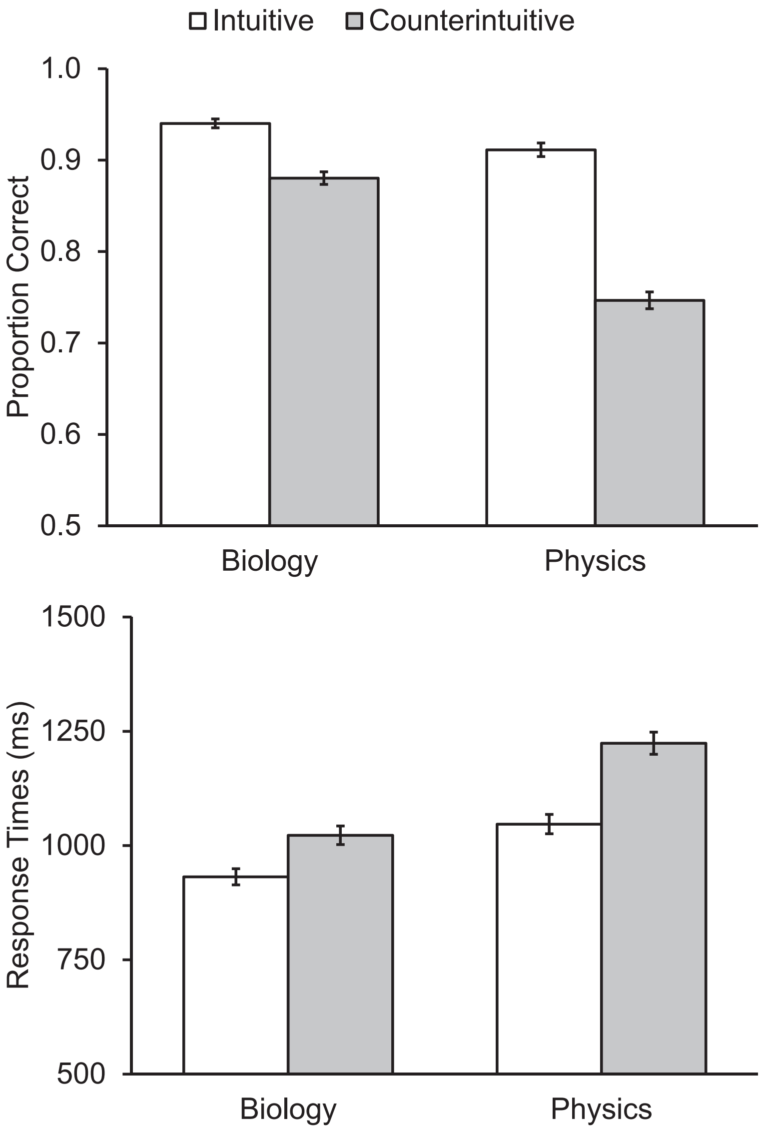

Fig. 1. Mean accuracy and mean response times for verifying intuitive and counterintuitive statements in the domains of biology and physics.

(F(1, 133) = 227.8, $p < .001$, partial $\eta^2 = 0.63$), as well as a significant interaction between them (F(1, 133) = 58.1, $p < .001$, partial $\eta^2 = 0.30$). The simple effect of statement type was again present in both domains (biology: t(133) = 11.9, $p < .001$, $d = 1.03$; physics: t(133) = 16.1, $p < .001$, $d = 1.39$).

The effect of statement type was large, both for accuracy (partial $\eta^2 = 0.79$) and latency (partial $\eta^2 = 0.70$), and was comparable to the effect of statement type documented in previous studies using the same method. Shtulman and Valcarcel (2012) documented effect sizes of 0.91 and 0.74 for accuracy and latency, respectively, and Shtulman and Harrington (2016) documented effect size of 0.86 and 0.59. The present study thus replicates the finding that counterintuitive scientific statements elicit cognitive conflict using an entirely different set of statements.

In one final analysis, we looked at whether participants responded more accurately or more quickly if they had received more science education. We did not ask participants to report their major, but we did ask them whether they were majoring in a natural science, a social science, or a humanities. Participants majoring in the natural sciences did not perform differently from those majoring in the social sciences or the humanities, though this classification is too broad to capture the domain-specific knowledge tapped by our biological statements and our physical statements. Nevertheless, we did find that performance varied by prior STEM coursework. Participants who had taken more STEM courses verified counterintuitive statements about matter more accurately ($r = .20$, $p = .02$) and more quickly ($r = .20$, $p = .02$), and they verified counterintuitive statements about life more quickly ($r = .19$, $p = .02$) though not more accurately ($r = .12$, $p = .16$). STEM coursework was unassociated with statement verifications for the intuitive statements, which all participants verified quickly and accurately. These findings suggest that participants who knew more about science were more competent at evaluating counterintuitive scientific statements, similar to previous findings with science professors (Shtulman & Harrington, 2016).

3.2. Correlations with co-occurrence estimates and attribute ratings

The above analyses indicate that participants experienced the cognitive conflict we anticipated they would experience when verifying counterintuitive statements. We now turn to the question of what explains that conflict. These analyses were conducted at the level of the predicate (“is alive,” “has cells,” “reproduces,” etc.) rather than the level of the participant. For each statement, we calculated the mean accuracy and mean latency across participants. We then correlated these means with their respective attribute ratings and co-occurrence estimates across statements. Table 4 provides a snapshot of such data, for the predicate “is alive.” Correlations were computed between (a) mean accuracy scores and mean attribute ratings, (b) mean accuracy scores and co-occurrence estimates, (c) mean latency scores and mean attribute ratings, and (d) mean latency scores and co-occurrence estimates.

True statements (like “crocodiles are alive” and “grass is alive”) appear in the top half of Table 4, and false statements (like “tables are alive” and “fire is alive”) appear in the bottom. Note that the correlations of interest run in opposite directions for true and false statements. Accuracy was positively correlated with attribute ratings and co-occurrence estimates for true statements but negatively correlated with these variables for false statements. Latency, on the other hand, was negatively correlated with attribute ratings and co-occurrence estimates for true statements but positively correlated with these variables for false statements. The correlations switch directions because the information captured

Table 4
Mean response accuracy (proportion correct) and response latency (in milliseconds) for the biological statements from Table 2, along with mean attribute ratings for the statements' subjects and co-occurrence estimates for the subjects and predicates. True statements are listed at the top and false statements at the bottom, with each list ordered by accuracy

Statement	Accuracy	Latency	Attribute Rating	Co-occurrence Estimate
Crocodiles are alive.	1.00	745	4.8	0.0138
Owls are alive.	1.00	860	4.9	0.0081
Turtles are alive.	1.00	868	4.8	0.0083
Tigers are alive.	1.00	877	4.8	0.0092
Oaks are alive.	0.94	912	3.4	0.0048
Willows are alive.	0.91	958	3.1	0.0053
Petunias are alive.	0.89	1,064	3.5	0.0076
Grass is alive.	0.88	824	2.8	0.0052
Tables are alive.	1.00	903	1.1	0.0018
Stones are alive.	0.94	911	1.5	0.0036
Boulders are alive.	0.94	799	1.5	0.0034
Mittens are alive.	0.91	875	1.3	0.0035
Geysers are alive.	0.88	1,010	2.1	0.0050
Glaciers are alive.	0.88	1,131	2.1	0.0042
Comets are alive.	0.86	1,053	2.0	0.0045
Fire is alive.	0.77	1,191	2.2	0.0061

by attribute ratings and co-occurrence estimates pulls participants toward responding "true." This information facilitates performance for true statements but hinders performance for false statements. Notably, truth-value was crossed with intuitiveness, meaning that attribute ratings and co-occurrence estimates were expected to track the intuitiveness of each statement among statements of the same truth-value.

To pool correlations across true statements and false statements, we squared them and treated these r-square values as our unit of analysis. The final dataset consisted of 144 r-squares: one for true statements and one for false statements for each of nine predicates in each of two domains across four measures of association: (a) accuracy and attribute ratings, (b) accuracy and co-occurrence estimates, (c) latency and attribute ratings, and (d) latency and co-occurrence estimates. R-squares correspond to the amount of variance in one measure explained by the other—in this case, the amount of variance in accuracy scores and latency scores explained by attribute ratings and co-occurrence estimates.

Fig. 2 displays r-square values, expressed as the percent of variance explained, averaged by response type (accuracy vs. latency), predictor variable (attribute ratings vs. co-occurrence estimates), and domain (biology vs. physics). All means were significantly greater than zero ($t(17) > 3.1$, $p < .01$), indicating that accuracy and latency were predicted by both variables in both domains.

The amount of variance explained by attribute ratings was consistently greater than that explained by co-occurrence estimates. We ran a repeated-measures ANOVA to assess the effects of predictor variable and domain on response accuracy. It revealed a

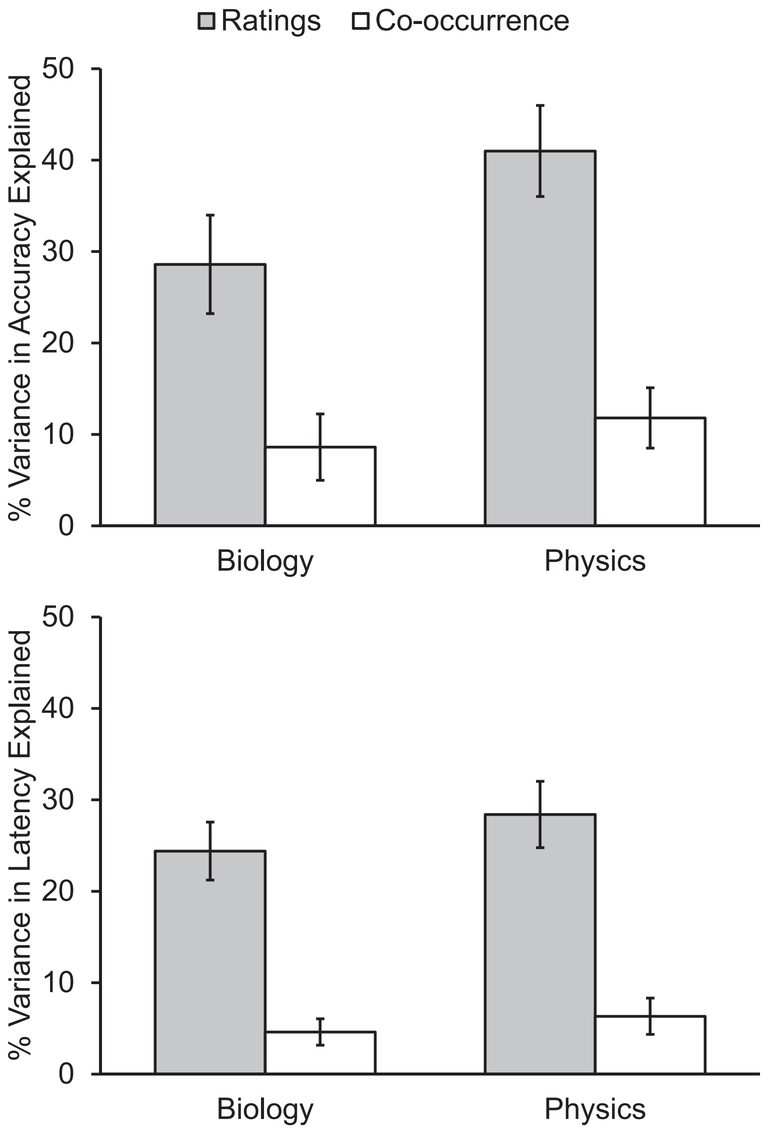

Fig. 2. The average amount of variance in participants' statement verifications explained by attribute ratings (of the statements' subjects) and lexical co-occurrence (of the statements' subjects and predicates) in each domain.

significant effect of predictor ($F(1, 34) = 43.2$, $p < .001$, partial $\eta^2 = 0.56$) but no effect of domain ($F(1, 34) = 1.1$, $p = .30$, partial $\eta^2 = 0.03$) and no interaction between predictor and domain ($F(1, 34) = 3.5$, $p = .07$, partial $\eta^2 = 0.09$). Similar findings were obtained for response latency. The effect of predictor was significant ($F(1, 34) = 68.4$, $p < .001$, partial $\eta^2 = 0.67$) but not the effect of domain ($F(1, 34) = 0.2$, $p = .66$, partial

$\eta^2 = 0.01$), nor the interaction between predictor and domain ($F(1, 34) = 1.0$, $p = .32$, partial $\eta^2 = 0.03$).

Across domains, the average amount of variance explained by attribute ratings was several times larger than that explained by co-occurrence estimates. For response accuracy, the amount of variance explained by attribute ratings was 35%, compared to the 10% explained by co-occurrence estimates. For response latency, the amount of variance explained by attribute ratings was 26%, compared to the 5% explained by co-occurrence estimates. That said, variance explained by one factor likely overlapped with variance explained by the other. To assess how much unique variance the two factors explained, we repeated the analyses described above but used partial correlations—correlations between accuracy and attribute ratings controlling for co-occurrence estimates, correlations between accuracy and co-occurrence estimates controlling for attribute ratings, and so forth.

The *r*-square values derived from partial correlations are displayed in Fig. 3 (expressed as the percent of variance explained), averaged again by response type, predictor variable, and domain. The partial correlations patterned the same as the full correlations. A comparison of the two sets of *r*-squares revealed no difference in the amount of variance explained by attribute ratings ($t(35) = 0.78$, $p = .44$) and co-occurrence estimates ($t(35) = 0.90$, $p = .38$) for response accuracy and only a slight drop in the amount of variance explained by attribute ratings ($t(35) = 2.58$, $p = .014$, mean difference = 1.6%) and co-occurrence estimates ($t(35) = 2.66$, $p = .012$, mean difference = 1.9%) for response latency. These findings indicate that attribute ratings and co-occurrence estimates explain unique portions of variance, presumably because they track unique influences on participants' behavior.

4. Discussion

When we reason about natural phenomena, we often experience cognitive conflict between scientific and folk explanations (Kelemen et al., 2013; Shtulman & Valcarcel, 2012; Vosniadou et al., 2018). This study looked at two potential factors underlying this conflict: theory-based expectations and lexical associations. In this study, we asked participants to verify a wide range of scientific statements—some intuitive and some counterintuitive—and found that lexical associations and theory-based expectations both accounted for variance in participants' speed and accuracy, but the latter accounted for three to five times as much. That is, measures of how well an item conforms to a domain-relevant intuitive theory explain three to five times more variance in participants' ability to verify statements about that item than measures of how often the item co-occurs with the statements' predicates in natural-language documents.

These data imply that lexical association may contribute to the conflict between folk explanations and scientific explanation, but intuitive theories are the primary contributor (at least when reasoning without time pressure or cognitive load). A statement like "dandelions are alive" is difficult for people to judge as true not because discourse about

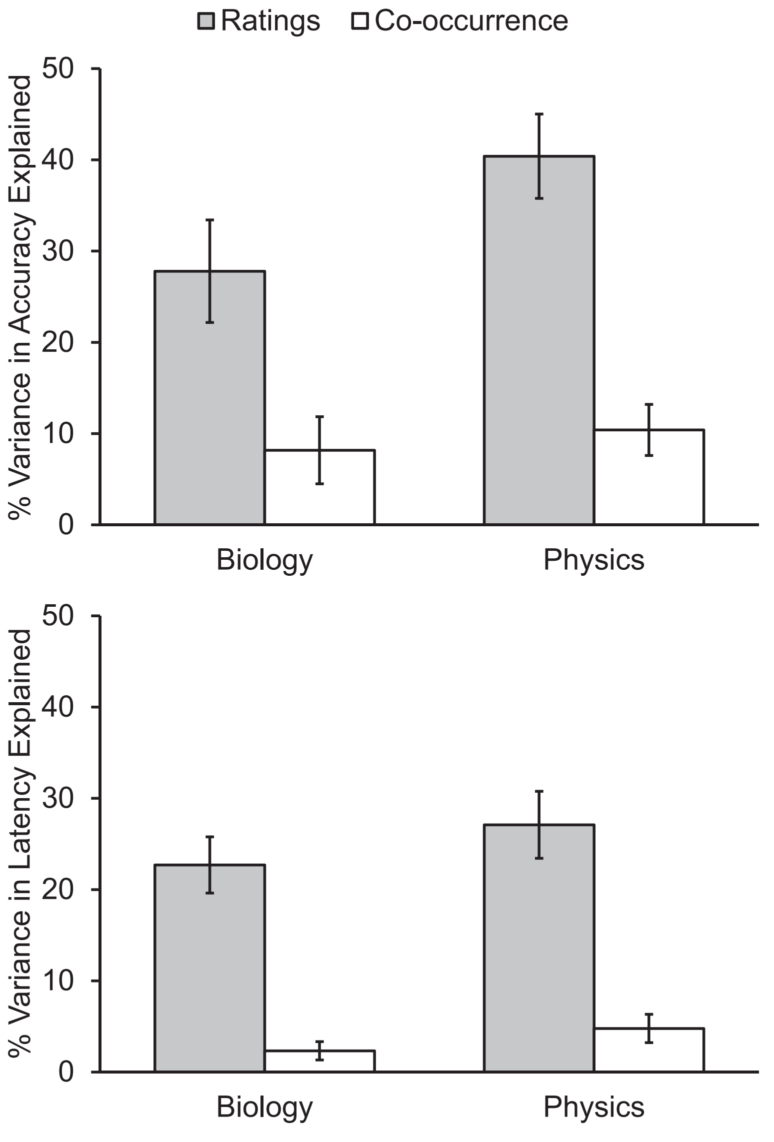

Fig. 3. The average amount of variance in participants' statement verifications explained by attribute ratings controlling for lexical co-occurrence and by lexical co-occurrence controlling for attribute ratings.

dandelions is void of biological terms but because dandelions lack the animate properties intuitively associated with living things. Conversely, a statement like "rivers are alive" is difficult for people to judge as false not because discourse about rivers is rife with biological terms but because rivers, like animals, appear animate. The sustained influence of intuitive theories on scientific reasoning makes sense in light of their origin. These theories are constructed early in development as a way of explaining everyday observations

(Carey, 2009; Shtulman, 2017; Vosniadou, 1994), and they continue to function in this way across the lifespan, even when we acquire a scientific understanding of the same observations (Carey & Spelke, 1996; Lewis & Linn, 1994). The utility of intuitive theories in everyday situations may be what explains their longevity in the face of scientific alternatives (Ohlsson, 2009; Shtulman & Lombrozo, 2016).

Lexical associations also accounted for unique variance in the speed and accuracy of participants' statement verifications. The joint contribution of lexical associations and intuitive theories echo the findings of Bright and Feeney (2014), which demonstrate that structured knowledge and associative knowledge jointly predict reasoning about natural kinds. One important difference between that study and ours is that Bright and Feeney manipulated the context in which reasoning occurred, by imposing time constraints (in Experiment 1) and cognitive load (in Experiment 2), whereas we allowed participants to respond at their own speed and without any load. It is possible that the predictive power of lexical associations and theory-based expectations might change if participants' ability to process our statements was curtailed. Under such circumstances, participants might rely more heavily on lexical associations than intuitive theories, though it is an open question whether the difference in predictive power between the two sources of information would be eliminated or even reversed.

Context aside, our data confirm that lexical associations and theory-based expectations are mutually compatible accounts of explanatory coexistence. They also confirm that explanatory coexistence is not driven solely by lexical association. In our task, conceptual properties of the items, known to participants but not articulated in the statements, had a strong influence on participants' responses. Conceptual knowledge provides the backdrop for interpreting a statement's scope and meaning, and it's difficult to imagine how lexical associations could exist in the absence of conceptual support for those associations. Language patterns that defy modern science should persist only if they remain interpretable and useful to language users. "Sunrise" and "sunset" have survived the cultural eclipse of geocentrism because we continue to represent geocentric models of the universe, at least implicitly (Shtulman, 2017). "Phlogiston," "miasma," and "telegony," on the other hand, do not correspond to any commonly held intuitive theories and have fallen by the wayside, along with their previously known associations to natural phenomena.

Stepping back from the question of what best predicts the cognitive conflict between scientific and folk explanations, it's worth considering whether "cognitive conflict" is the right characterization of this effect (i.e., decreased accuracy and increased response times for counterintuitive statements). If people recognize that they hold multiple conceptions of a domain, they may have trouble responding to a statement like "air has weight" because they have to decide which conception is appropriate for the context, the folk conception (weight as heft) or the scientific conception (weight as the product of mass and gravity). On this interpretation, conflict lies in the nature of the task rather than the minds of the respondents.

There are several problems with this interpretation. First, scientists show the same response pattern—that is, decreased accuracy and increased latency when verifying counterintuitive scientific ideas—despite a lifetime of deploying scientific explanations (Goldberg

& Thompson-Schill, 2009; Shtulman & Harrington, 2016). Second, people who are directly trained to privilege science over intuition perform more accurately on our statement-verification task but continue to show response lags for counterintuitive statements (Young et al., 2018). Third, tensions between science and intuition have been documented using other methods, including brain imaging (Allaire-Duquette, Belanger, Grabner, Koschutnig, & Masson, 2019) and masked priming (Preston & Epley, 2009), and in populations with less meta-conceptual awareness, including children (Vosniadou et al., 2018) and Alzheimer's patients (Zaitchik & Solomon, 2008). That said, it is an open question whether people are generally aware of the coexistence of scientific and folk explanations, and whether this awareness influences their ability to prioritize one over the other.

4.1. Limitations and future directions

There are several reasons to suspect that lexical associations were weak predictors of cognitive conflict because of how those associations were measured. Our measures of associative strength were derived from a diverse sample of documents, produced at different times by different people, whereas our measures of theory-based expectations were derived from the same population that produced our dependent measure (statement verifications). In addition, our measures of associative strength were nondirectional, meaning that associations between a subject and predicate would exert the same influence if the subject and predicate were reversed ("the sun revolves around the earth" vs. "the earth revolves around the sun") or otherwise modified ("the sun revolves around the earth" vs. "the sun does not revolve around the earth"). Finally, the observed rates of co-occurrence between our subjects and our predicates were low, averaging 2.1% across statements, and differences within this range of values may have paled in comparison to differences in accuracy and speed.

For these reasons, it would be useful to extend this research using other measures of associative strength. The words "sun" and "alive" may not be tightly linked in corpus data (they overlapped in only 0.6% of documents containing "sun"), but their referents might be linked through other types of mental representations, such as mental images or episodic memories. These associations could be captured by asking people to rate the global similarity between subjects and predicates (as done by Bright & Feeney, 2014), but precautions would need to be taken to avoid tapping theory-based expectations as well. Ratings provided without any guidance or constraint would likely reflect a mixture of both low-level associations and high-level expectations.

Another direction for future research would be to build upon our measure of lexical association by using different corpuses of natural-language documents or different metrics of association. We could, for instance, compute co-occurrence estimates using a database that includes nonpublished documents. EBSCOhost proved useful for our purposes, but co-occurrence statistics gathered from less formal documents may do a better job predicting the dynamics of how nonscientists evaluate scientific statements. Other metrics of association may do a better job as well. Raw co-occurrence does not always capture lexical associations evident at higher levels of linguistic processing. The association between "fireman" and "policeman," for instance, may be stronger in people's minds than in English-language documents because stories about firemen do not necessarily mention policemen and vice versa, even though both occupations are prototypical of public service professions.

An analytic technique that can capture such higher-level associations, like Latent Semantic Analysis (Landauer & Dumais, 1997), may yield a more predictive measure of lexical association. But this measure may be more predictive because it captures some of the theory-based expectations not captured by raw co-occurrence (Dam & Kaufmann, 2008; Sherin, 2013), which would run counter to the goal of differentiating associative knowledge from structured knowledge. While different measures of lexical association may explain more of the variance in people's statement-verification behavior, we suspect that such measures will still lag behind theory-based measures, particularly for statements about scientific ideas that people don't often discuss, like atoms, molecules, and density or nutrients, cells, and respiration. Our understanding of these ideas may elicit competing explanations even if we do not talk about them much.

Another way of advancing research on the conflict between scientific and folk explanations would be to explore the underpinnings of this conflict in other scientific domains. We selected life and matter for their suitability to our task, but competition between scientific and folk explanations has been documented in at least 10 other domains (Shtulman & Valcarcel, 2012), including domains where scientific explanations are learned much later (e.g., mechanics), domains where scientific explanations compete with religious explanations (e.g., evolution), and domains where the relevant form of reasoning is more spatial (e.g., astronomy) or more mathematical (e.g., fractions). Such domains provide opportunities for replicating the observed effects with different measures of lexical association and different measures of theory-based expectations. The persistence of folk explanations across ages, cultures, and domains implies that explanatory coexistence is an inherent feature of conceptual change, and research on the factors that preserve folk explanations promises to inform our understanding of that process.

4.2. Conclusions

Science provides explanations for a wealth of empirical phenomena—life, matter, heat, adaptation, illness—but people construct folk explanations for those phenomena before learning the scientific ones. Folk explanations appear to persist throughout the lifespan, producing cognitive conflict when incompatible with a scientific alternative. Here, we examined the dynamics of that conflict and found that it tracks two kinds of information: the strength of association between the words used to articulate scientific ideas and the degree to which those ideas conform to intuitive theories of the domain. Both kinds of information play a role in the cognitive conflict, but the latter plays a more substantial role (at least when reasoners have sufficient time and attention to access their theories).

This finding implies that folk explanations are grounded in intuitive theories and are elicited in situations where intuitive theories are particularly useful or applicable. Intuitive theories of matter are triggered by things that can be seen or felt, and intuitive theories of life are triggered by things that move on their own. If folk explanations are driven by the utility of an intuitive theory, then one way to curb them might be to strengthen the utility of their scientific competitor (Ohlsson, 2009; Young et al., 2018). A primary goal of science educators is to help students understand the conceptual structure of scientific theories, but science educators could also help students apply those theories to everyday situations—situations most readily explained by intuitive theories (Chi, Slotta, & de Leeuw, 1994; Clement, 1993; Songer & Linn, 1991). Science educators could also help students appreciate the shortcomings of intuitive theories, but such instruction may be futile, given that intuitive theories interfere with scientific reasoning even for professional scientists (Goldberg & Thompson-Schill, 2009; Kelemen et al., 2013; Shtulman & Harrington, 2016). The persistence of intuitive theories seems to be inevitable, but the dynamics of when and how they are deployed remains a fruitful topic for further investigation.

Acknowledgments

This research was supported by an Understanding Human Cognition Scholars Award from the James S. McDonnell Foundation to Andrew Shtulman and a National Science Foundation grant (1420241) to Cristine Legare. We thank Liza Comart, Alexander Flood-Bryzman, Columbia Shafer, and Stefanie Young for their assistance with data collection and data analysis.

References

Allaire-Duquette, G., Belanger, 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.

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

Barlev, M., Mermelstein, S., & German, T. C. (2018). Representational coexistence in the God concept: Core knowledge intuitions of God as a person are not revised by Christian theology despite lifelong experience. *Psychonomic Bulletin & Review*, *25*, 2330–2338.

Bright, A. K., & Feeney, A. (2014). The engine of thought is a hybrid: Roles of associative and structured knowledge in reasoning. *Journal of Experimental Psychology: General*, *143*, 2082–2102.

Busch, J. T. A., 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*. Cambridge, MA: MIT Press.

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

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515–533.

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

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

Colombo, M. (2017). Experimental philosophy of explanation rising: The case for a plurality of concepts of explanation. *Cognitive Science*, *41*, 503–517.

Dam, G., & Kaufmann, S. (2008). Computer assessment of interview data using latent semantic analysis. *Behavior Research Methods*, *40*, 8–20.

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*, 326–353.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

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.

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual Review of Anthropology*, *40*, 379–398.

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

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

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

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

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.

Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *11*, 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.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

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., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development*, *83*, 779–793.

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.

Legare, C. H., & Shtulman, A. (2018). Explanatory pluralism across cultures and development. In J. Proust & M. Fortier (Eds.), *Interdisciplinary approaches to metacognitive diversity* (pp. 415–432). Oxford, UK: Oxford University Press.

Legare, C. H., & Visala, A. (2011). Between religion and science: Integrating psychological and philosophical accounts of explanatory coexistence. *Human Development*, *54*, 169–184.

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.

Mareschal, D. (2016). The neuroscience of conceptual learning in science and mathematics. *Current Opinion in Behavioral Sciences*, *10*, 114–118.

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.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

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

Nguyen, S., & Rosengren, K. (2004). Causal reasoning about illness: A comparison between European- and Vietnamese-American children. *Journal of Cognition & Culture*, *4*, 51–78.

Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist*, *44*, 20–40.

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., Sauriol, E., & Riopel, M. (2015). Experimental evidence of the superiority of the prevalence model of conceptual change over the classical models and repetition. *Journal of Research in Science Teaching*, *52*, 1082–1108.

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

Preston, J. L., Ritter, R. S., & Hepler, J. (2013). Neuroscience and the soul: Competing explanations for the human experience. *Cognition*, *127*, 31–37.

Raman, L., & Gelman, S. A. (2004). A cross-cultural developmental analysis of children's and adults' understanding of illness in South Asia (India) and the United States. *Journal of Cognition and Culture*, *2*, 293–317.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Sherin, B. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences*, *22*, 600–638.

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., & 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.

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

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

Shtulman, A., & Young, A. (2020). Why do logically incompatible explanations seem psychologically compatible? Science, pseudoscience, religion, and superstition. In K. McCain & K. Kampourakis (Eds.), *What is scientific knowledge? An introduction to contemporary epistemology of science* (pp. 163–178). New York: Routledge.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*, 639–651.

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.

Songer, N. B., & Linn, N. B. (1991). How do students views of science influence knowledge integration? *Journal of Research in Science Teaching*, *28*, 761–784.

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

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*, e17349.

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

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

Watson-Jones, R. E., Busch, J. T. A., Harris, P. L., & Legare, C. H. (2017). Does the body survive death? Cultural variation in beliefs about life everlasting. *Cognitive Science*, *41*, 455–476.

Watson-Jones, R. E., Busch, J. T. A., & Legare, C. H. (2015). Interdisciplinary and cross-cultural perspectives on explanatory coexistence. *Topics in Cognitive Science*, *7*, 611–623.

Weisberg, D. S., Landrum, A. R., Metz, S. E., & Weisberg, M. (2018). No missing link: Knowledge predicts acceptance of evolution in the United States. *BioScience*, *68*, 212–222.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375.

Young, A., Laca, J., Dieffenbach, G., Hossain, E., Mann, D., & Shtulman, A. (2018). Can science beat out intuition? Increasing the accessibility of counterintuitive scientific ideas. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1238–1243). Austin, TX: Cognitive Science Society.

Young, A., Powers, A., Pilgrim, L., & Shtulman, A. (2018). Developing a cognitive reflection test for school-age children. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1232–1237). 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.