Title
Cognitive Reflection Predicts Science Understanding

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Shtulman, Andrew
McCallum, Kate

Publication Date
2014

Cognitive Reflection Predicts Science Understanding

Andrew Shtulman (shtulman@oxy.edu)
Department of Psychology, Occidental College
1600 Campus Road, Los Angeles, CA 90041

Kate McCallum (kate.e.mccallum@gmail.com)
Department of Psychology, Sam Houston State University
1901 Avenue I, Huntsville, TX 77341

Abstract

Understanding scientific theories like evolution by natural selection, classical mechanics, or plate tectonics requires knowledge restructuring at the level of individual concepts, or conceptual change. Here, we investigate the role of cognitive reflection (Frederick, 2005) in achieving conceptual change. College undergraduates (n = 184) were administered a 45-question survey probing their understanding of six domains of science requiring conceptual change – astronomy, evolution, geology, mechanics, perception, and thermodynamics – as well as (a) their ability to analyze covariation-based data, (b) their understanding of the nature of science (NOS), and (c) their disposition towards cognitive reflection. Cognitive reflection was a significant predictor of science understanding in all domains, as well as an independent predictor, explaining significantly more variance in science understanding than that explained by covariation analysis ability and NOS understanding combined. These results suggest that cognitive reflection may be a prerequisite for changing certain cognitive structures, namely, concepts and theories.

Keywords: conceptual change, science education, CRT

Introduction

Scientific discoveries come in two forms: those that can be understood in terms of a pre-existing paradigm and those that require the adoption of a new paradigm altogether. A prime example is the difference between the discovery of Neptune and the discovery of heliocentrism. Neptune was predicted to exist many decades before it was discovered, on account of certain unexplained perturbations in the orbit of Uranus. Nineteenth century astronomers thus sought observational confirmation of an eighth planet with the same basic properties as those of the seven planets already known to exist (Littmann, 2004). Neptune's existence was thus readily assimilated into astronomers' preexisting model of the solar system. That model itself, however, was hard won, as astronomers prior to Copernicus typically subscribed to geocentric models of planetary motion. Accepting the sun as the center of planetary motion required revision of the most basic astronomical assumptions of the time, including the causes of celestial motion and the very ontology of celestial objects (Kuhn, 1957).

Parallel to scientific discovery, the process of learning scientific concepts also comes in two forms: learning that can be accomplished in terms of one's preexisting concepts, termed *knowledge enrichment*, and learning that requires the adoption of new concepts via the revision or restructuring of one's preexisting concepts, termed *conceptual change* (Carey, 2009; Shtulman, 2009; Vosniadou, 1994). In biology, for instance, learning the traits of an unfamiliar animal would constitute knowledge enrichment, whereas learning how novel traits emerge through the process of natural selection would constitute conceptual change (Shtulman, 2006). In physics, learning that objects fall with an acceleration of 9.8 m/s^2 would constitute knowledge enrichment, whereas learning that weight is a relational property between an object and a gravitational field (as opposed to an intrinsic property of the object itself) would constitute conceptual change (Halloun & Hestenes, 1985).

Thus, a primary challenge facing science educators across different levels of schooling and different domains of science is helping students achieve conceptual change. To meet this challenge, many researchers have devised curricula that are informed by research on students' preconceptions (or misconceptions) within a domain and that explicitly aim to bridge the gap between those preconceptions and a correct, scientific understanding of the domain. This approach has been adopted with great success in domains ranging from microbiology (Ah et al., 2008) to evolution (Shtulman & Calabi, 2012) to thermodynamics (Slotta & Chi, 2006) to material science (Smith, 2007). Another, less common approach is to identify domain-general factors – i.e., skills, abilities, or dispositions – that correlate with science understanding across a variety of domains in the hopes of addressing those factors prior to, or during, instruction (e.g., Kloos, 2007; Zaitchik, Iqbal, & Carey, 2014). This approach is not only for pedagogically informative but is also theoretically informative, as correlations between domain-general competencies and domain-specific knowledge shed light on how that knowledge is represented in the mind and by what mechanisms it might be acquired or changed.

In the current study, we sought to determine whether the ability to achieve conceptual change across various domains of science is correlated with a disposition towards *cognitive reflection*, or a disposition towards deliberating on one's reasoning prior to accepting the first answer or solution that pops into one's mind. Cognitive reflection was first measured, as an independent construct, by Frederick (2005). Frederick's "Cognitive Reflection Test," or CRT, consists of three items like the following: "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, given that the lily pads must have covered half the lake one day prior to covering the entire lake, but the question is designed to elicit an intuitive response of 24, derived simply by dividing 48 in half. In general, the items on the CRT are designed to elicit an intuitive, yet erroneous, response that is readily recognized *as* erroneous upon further reflection. At issue is whether one engages in that reflection or simply “goes with their gut.”

The CRT has been shown to be a strong predictor of many conceptually independent forms of reasoning, including syllogistic reasoning, Bayesian reasoning, causal reasoning, covariation detection, and temporal discounting (Frederick, 2005; Toplak, West, & Stanovich, 2011). Reflective responses to CRT items predict accurate or normative patterns of reasoning, whereas intuitive responses predict inaccurate or fallacious patterns of reasoning. Here, we sought to determine whether reflective responses predict evidence of science understanding and intuitive responses predict a lack thereof. In particular, we sought to determine whether reflective responses predict science understanding in domains for which *conceptual change* has been implicated as a prerequisite to understanding, i.e., the domains of astronomy, evolution, geology, mechanics, perception, and thermodynamics. Our rationale for targeting knowledge acquired through conceptual change is that conceptual change is a protracted and hard-won cognitive achievement (Carey, 2009), and, as such, may require a fair amount of cognitive reflection to complete.

In addition to cognitive reflection, we measured two other forms of reasoning arguably associated with science understanding: the ability to analyze covariation-based data (Fugelsang & Thompson, 2003; Smedslund, 1963) and an understanding of the nature of science, or NOS, as a method of inquiry (Carey, Evans, Honda, Jay, & Unger, 1989; Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). We expected all three factors – cognitive reflection, covariation analysis, and NOS understanding – to correlate with science understanding, as each constitutes a unique form of reasoning potentially important to the acquisition of domain-specific knowledge, but we hypothesized that cognitive reflection would be (a) a stronger predictor of science understanding and (b) an independent predictor, explaining variance in science understanding above and beyond that explained by the other two factors. Both hypotheses were confirmed, as described below.

Method

Participants

The participants were 184 undergraduates at Occidental College who were recruited from psychology courses and compensated with course credit. The majority (71%) were female, and they reported having completed an average of 4.0 college-level math and science courses (*SD* = 4.2).

Procedure

Participants answered 45 questions in a survey administered via MediaLab v2012 software. The questions were broken into nine blocks of five questions each. The first six blocks assessed participants’ understanding of various concepts within the domains of astronomy, evolution, geology, mechanics, perception, and thermodynamics (one block per domain). The last three blocks covered more domain-general forms of reasoning: covariation analysis, NOS understanding, and cognitive reflection. Questions were selected from preexisting instruments in the science education and cognitive psychology literatures and had thus been vetted for validity and reliability in prior research.

The ordering of the blocks was randomized across participants, as was the ordering of the questions within each block, with the stipulation that the six blocks of domain-specific questions were administered before the three blocks of domain-general questions. One final block of questions was used to collect demographic information from participants, namely, gender, class standing, major, and number of college-level math and science courses taken (referred to henceforth as STEM courses).

Participants selected one of several multiple-choice options for all questions except those on the CRT, for which they entered a numerical value instead. Each multiple-choice question was, in turn, followed by a prompt to provide an explanation. We requested explanations mainly to discourage participants from making their selections at random, and the explanation data are not analyzed here. Accuracy was gauged by multiple-choice responses alone (or, in the case of the CRT, participants’ numerical entries).

Astronomy Understanding

While most college-educated adults know that the moon orbits the Earth and that the Earth orbits the sun, few understand how those orbits give rise to astronomical phenomena like the tides, the seasons, or the phases of the moon (Trundle, Atwood, & Christopher, 2007; Tsai & Chang, 2005). We probed participants’ understanding of such phenomena with questions like the following: “The Earth’s shadow on the moon is responsible for which of the following phenomena? (a) A lunar eclipse; (b) A crescent moon; (c) A new moon; (d) All of the above.” The correct response is (a); however, most participants (69%) chose (b), (c), or (d), symptomatic of a misunderstanding of the role of the observer’s perspective in our perception of the moon’s degree of illumination.

Evolution Understanding

Most adults construe evolution not as a selection-based process but as an entire population holistically transforming over time, with each generation somehow guaranteed to be born with the traits they need in order to survive (Bishop & Anderson, 1990; Shtulman, 2006; Shtulman & Calabi, 2013). We probed for this alternative, need-based view of evolution with such questions as: “Imagine that biologists discover a new species of woodpecker that lives in isolation on a secluded island. These woodpeckers have, on average, a 1.0 inch beak, and their only food source is a tree-dwelling insect that lives, on average, 1.5 inches under the tree bark.

Compared to its parents, the offspring of any two woodpeckers should develop which of the following traits? (a) A longer beak; (b) A shorter beak; (c) Either a longer beak or a shorter beak; neither is more likely." The correct answer is (c), since parent-offspring differences are random and unpredictable, but most participants (65%) chose (a), endorsing the idea that offspring will be born more adapted to the environment than their parents were at birth.

Geology Understanding

The Earth is a dynamic system of interacting processes, but most students view the Earth as an inert object. Consequently, they have difficulty appreciating the causes of geological phenomena, like earthquakes and volcanoes, as well as changes to the Earth that occur over a geologic time scale (Libarkin, Anderson, Dahl, Beilfuss, Boone, & Kurdziel, 2005; Trend, 2000). We assessed students' understanding of geologic systems and geologic time with such questions as: "Where are tectonic plates located? (a) At the Earth's surface; (b) At the Earth's core; (c) Between the Earth's surface and the Earth's core; (d) Different plates are located at different positions within the Earth." The correct answer is (a), but most participants (62%) chose (b), (c), or (d), reflecting the misconception that there is a discontinuity between the seemingly static ground we stand on and the dynamic components of the Earth's structure.

Mechanics Understanding

The concept of inertia plays no role in the average adult's understanding of object motion. Motion and rest are seen as fundamentally distinct states, with motion implying the presence of a force and rest implying the absence of one. Indeed, forces are conceptualized not as *acting on* objects but as *imparted to* objects – i.e., as an internal impetus propelling objects forward or upward until dissipated (Halloun & Hestenes, 1985; McCloskey, 1983). To tap into this alternative, impetus-based view of motion, we asked questions like the following: "A ball is thrown into the air. What forces act on the ball on its way up? (a) Its weight, vertically downward; (b) A force that maintains the ball's motion, vertically upward; (c) Both of the above; (d) Neither of the above." The correct answer is (a), but most participants (55%) chose (c), ostensibly reasoning that upward motion requires an upward "force" or impetus.

Perception Understanding

Colors and sounds are typically conceived of as intrinsic properties of the environment, whereas, in reality, they are relational properties between the environment and the perceiver (Mazens & Lautrey, 2003; Eaton, Anderson, & Smith, 1984). We assessed participants' understanding of the relational aspects of perception with questions like the following: "Red objects are perceived as red rather than blue because they do what? (a) Absorb more red light than blue light; (b) Reflect more red light than blue light; (c) Both absorb and reflect more red light than blue light." The correct answer is (b), because color is the perception of light waves reflected off an object and into the eye, but around half of the sample (52%) chose (a), seemingly treating color as an intrinsic property of the object itself.

Thermodynamics Understanding

Heat is an emergent property of kinetic energy at the molecular level, but many adults do not view heat as energy. Rather, they view heat as a kind of immaterial substance that passes from one object to another. On this view, heat is seen as trappable, containable, non-additive (like temperature), and fundamentally distinct from coldness (Clark, 2006; Wiser & Amin, 2001). To probe for this alternative, substance-based views of heat, we asked questions like the following: "An ice cube is placed inside the pocket of a thick coat and left at room temperature over night. Compared to an ice cube on an open counter, the ice cube in the coat will do what? (a) Melt faster; (b) Melt slower; (c) Melt at the same rate." The correct answer is (b), because the coat would insulate the ice from the ambient heat of the room, but most participants (63%) chose (a) or (c), reasoning either that coats are intrinsically warm or that coats trap heat but not cold.

Cognitive Reflection

Frederick's (2005) Cognitive Reflection Task (CRT), described above, was administered with one modification: two additional items were included so that the range of possible scores was equivalent across sections (0 to 5). The additional items were: (1) "A house contains a living room and a den that are perfectly square. The living room has 4 times the square footage of the den. If the walls in the den are 10 feet long, how long are the walls in the living room?" and (2) "A store owner reduced the price of a $100 pair of shoes by 10 percent. A week later, he reduced the price of the shoes by another 10 percent. How much do the shoes cost now?" For item 1, 59% of participants provided the correct, reflective response of 20, and 26% provided the incorrect, intuitive response of 40. For item 2, 80% provided the correct, reflective response of 81, and 11% provided the incorrect, intuitive response of 80. Preliminary analyses revealed that our 5-item CRT (henceforth referred to as the "CRT-5") was a stronger predictor of science understanding than the original 3-item CRT across all six domains, so we retained the two new items in our final analyses.

Covariation Analysis

Determining whether a correlation can be inferred from covariation-based data is a notoriously difficult task, particularly when the data are at odds with prior beliefs about whether the candidate variables are causally related (Fugelsang & Thompson, 2003; Smedslund, 1963). We assessed participants' ability to analyze covariation-based data with questions like the following: "Imagine you are a scientist who is trying to determine the cause of a recent increase in lung cancer. You hypothesize that the lung cancer may be due to taking a new type of iron supplement. To test this hypothesis, you investigate 10 patients who took the new iron supplement and 50 patients who did not and discover that 8 of the 10 patients who took the supplement developed lung cancer and 8 of the 50 patients who did not take the supplement developed lung cancer. Which of the following can be inferred from these data? (a) There is a strong relationship between lung cancer and the new iron supplement; (b) There is a weak relationship between lung cancer and the new iron supplement; (c) There is no relationship between lung cancer and the new iron supplement." The correct answer is (a); however, around half of the sample (47%) chose (b) or (c), presumably swayed by the correspondence in absolute frequencies across conditions (8 vs. 8) despite vast difference in relative frequencies (0.80 vs. 0.16).

Nature of Science Understanding

In addition to measuring participants' understanding of specific science concepts, we also measured their understanding of science as a method of inquiry (Carey et al., 1989; Lederman et al., 2002). Our questions covered the nature of an experiment, the nature of a theory, the role of empirical data in testing scientific claims, and the role of inference in generating scientific knowledge. A sample question is as follows: "Which of the following best describes the nature of a scientific theory? (a) A well supported explanation; (b) A well educated guess; (c) A well documented finding; (d) A well respected assumption; (e) An irrefutable idea." The correct answer is (a), given that theories are more explanatory than descriptive and are well substantiated but not irrefutable. However, a substantial minority (34%) chose one of the alternative options.

Results

Scores by Section

Each section of the survey consisted of five questions, the answers to which were scored as either correct (1) or incorrect (0). Total scores therefore ranged from 0 to 5, with mean scores and standard deviations displayed in Table 1. Also displayed in Table 1 are the proportion of participants who answered more questions incorrectly than correctly (thus earning a score between 0 and 2) and the proportion who answered more questions correctly than incorrectly (thus earning a score between 3 and 5).

Scores on the six sections measuring science understanding differed significantly by domain ($F(5,915) = 19.18$, $p < .001$). Nevertheless, Bonferroni comparisons revealed that this effect was driven entirely by the difference between the thermodynamics section and all other sections. With the exception of the thermodynamics section, most participants answered most questions incorrectly. There was, however, a sizeable minority in each domain who showed evidence of having achieved conceptual change, and this minority tended to be comprised of the same participants across domains. In other words, participants who scored high in one domain tended to score high in other domains as well.

Table 1: Mean scores on each section of the survey, plus the proportion of participants who scored 0-2 vs. 3-5.

Section	*M*	*SD*	Scored 0-2	Scored 3-5
Astronomy	2.1	1.1	.65	.35
Evolution	2.1	1.3	.64	.36
Geology	2.1	1.4	.60	.40
Mechanics	1.8	1.2	.77	.23
Perception	2.0	0.9	.74	.26
Thermodynamics	2.8	1.1	.39	.61
CRT-5	2.6	1.5	.46	.54
COVAR	3.3	1.1	.24	.76
NOS	2.5	1.0	.54	.46

Correlational analyses confirmed this observation. Of the 15 pairwise correlations between the six domains, 11 were significant. Furthermore, a factor analysis of composite scores for each domain yielded a one-factor solution (at an Eigenvalue threshold of 1.0), implying that participants' understanding of science across a variety of domains was determined, in part, by a single underlying disposition. Below we attempt to characterize that disposition by comparing participants' scores on the domain-specific measures of science understanding to their scores on the tasks measuring more domain-general forms of reasoning.

Predictors of Science Understanding

Correlations between participants' science understanding scores and their scores on the CRT-5, their scores on the covariation analysis task (abbreviated COVAR), and their scores on the NOS understanding task are displayed in Table 2. While all three sets of correlations were positive and generally significant, the set pertaining to CRT-5 scores were larger and more consistent than the other two.

Table 2: Correlations between CRT-5 scores, COVAR scores, NOS scores, and science understanding. $^{*}p < .05$, $^{**}p < .01$

Section	CRT-5	COVAR	NOS
Astronomy	.16*	.10	.08
Evolution	.35**	.26**	.22**
Geology	.35**	.16*	.26**
Mechanics	.27**	.13	.09
Perception	.24**	.03	.17*
Thermodynamics	.30**	.12	.16*

To determine whether CRT-5 scores were an independent predictor of science understanding, we regressed the total number of science questions answered correctly ($M = 12.8$, $SD = 4.2$, range = 3 to 27) against CRT-5 scores in a hierarchical regression. In the first step of the regression, we entered demographic variables, namely, gender (coded "0" for female and "1" for male) and number of prior STEM courses. In the second step, we entered COVAR scores and NOS scores. In the third and final step, we entered CRT-5 scores. The results of this analysis are displayed in Table 3.

As predicted, COVAR scores and NOS scores explained significantly more variance than that explained by the demographic variables, and CRT-5 scores explained significantly more variance than that explained by COVAR scores and NOS scores (plus the demographic variables). Indeed, in the final model, CRT-5 scores emerged as the strongest predictor, explaining nearly twice as much variance in science understanding than that explained by the next strongest predictor, gender ($\beta = .36$ vs. $\beta = .19$).

Table 3: Regression analysis of composite science understanding scores by gender, STEM courses, COVAR score, NOS score, and CRT-5 score. $^*p < .05$, $^{**}p < .01$

Model	Factor	Beta	*t* value	R^2	*F* change
1	Gender	.26	3.66**	.13	13.26**
	STEM	.22	3.28**		
2	Gender	.21	3.12**	.22	12.51**
	STEM	.21	3.21**		
	COVAR	.18	2.65**		
	NOS	.23	3.41**		
3	Gender	.19	2.96**	.32	17.00**
	STEM	.16	2.48*		
	COVAR	.09	1.33		
	NOS	.13	2.07*		
	CRT-5	.36	5.25**		

Reflective vs. Intuitive Responses

The CRT-5 scores entered into the above analyses were based on correct, reflective responses only, but incorrect response are informative as well when those response indicate a reliance on intuition (e.g., an answer of “24 days” to the lily pad question). On average, participants produced reflective responses 57% of the time and intuitive responses 27% of the time. The remaining 16% were either irrelevant (e.g., “the square root of 48”) or incomplete (e.g., “not sure; math is hard”).

Intuitive responses were negatively correlated with composite science understanding scores ($r(182) = .37$, $p < .001$). Moreover, the difference between intuitive and reflective responding varied systematically by science understanding, as shown in Table 4. For each domain, we separated participants who earned a score of 0 to 2 (“low scorers”) from those who earned a score of 3 to 5 (“high scorers”) and compared the difference in reflective and intuitive responses provided. In all six domains, that difference was smaller for low scorers than for high scorers.

We confirmed the reliability of this effect with repeated-measures analyses of variance (ANOVA) in which CRT-5 response type (intuitive vs. reflective) was analyzed within participants and science understanding (low vs. high) was analyzed between participants. As expected, the interaction between CRT-5 response type and science understanding was significant in all six domains (all *F*’s > 4.0, all *p*’s < .05). In other words, the degree to which participants relied on reflection over intuition was positively associated with science understanding across a variety of domains.

Table 4: Mean differences in response type to the CRT-5 (reflective – intuitive) across domains.

Section	Low scorers	High scorers
Astronomy	0.6	1.5
Evolution	0.4	2.0
Geology	0.3	1.8
Mechanics	0.5	2.5
Perception	0.7	1.6
Thermodynamics	0.1	1.5

Discussion

Understanding complex scientific concepts often requires the restructuring of our earlier, intuitive conceptions – a process known as conceptual change. Here, we investigated domain-general correlates of one’s likelihood of having made conceptual change across six different domains of science: astronomy, evolution, geology, mechanics, perception, and thermodynamics. While the ability to analyze covariation-based data and an understanding of the nature of science (as a method of inquiry) were both significant predictors of science understanding, neither was as strong a predictor as one’s disposition towards cognitive reflection. In fact, cognitive reflection explained more variance in science understanding than that explained by gender, prior STEM coursework, covariation analysis ability, and NOS understanding combined.

These findings have important implications from both theoretical and pedagogical perspectives. Theoretically, they imply that cognitive reflection may be a prerequisite for changing certain cognitive structures, namely, concepts. All of us reason *through* our concepts, but we likely vary in how often we reason *about* our concepts, and it is this tendency – the tendency to reason about one’s concepts – that may underlie the shared variance between CRT scores and science understanding. Scoring highly on the CRT, after all, requires more than just inhibiting an intuitive response; it also requires the conceptual insight that one’s intuition is, in fact, wrong. Pedagogically, these findings imply that instructors could use the CRT as a diagnostic for determining who is likely to profit from instruction and who is not. Indeed, students who are low in cognitive reflection may actually benefit from different kinds of instruction than those who are high in cognitive reflection.

That said, we must acknowledge that these findings do not provide evidence of a *causal* relation between cognitive reflection and science understanding. While the results are consistent with the possibility that cognitive reflection facilitates science *learning*, they are also consistent with the possibility that cognitive reflection and science understanding are linked by some unmeasured variable, e.g., the quality of one’s prior education or the adequacy of one’s test-taking ability. It is also possible that learning complex scientific concepts increases one’s disposition towards cognitive reflection. Future research should therefore investigate the relation between cognitive reflection and science learning directly, either through prospective studies of the impact of cognitive reflection on achievement in science classes or through interventions designed to increase cognitive reflection. Such research could help determine not only the causal direction of the observed relations but also which aspects of cognitive reflection—e.g., inhibition, inconsistency detection, comprehension monitoring (see Zaitchik et al., 2014)—covary with science understanding.

Acknowledgments

We would like to thank the National Science Foundation for supporting this research via grant DRL-0953384 awarded to Andrew Shtulman. We would also like to thank Samuel Boland, Kelsey Harrington, Jai Levin, and Sharang Tickoo for their assistance with data collection and data analysis.

References

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*, 1-19.

Bishop, B. & Anderson, C.A. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27*, 415-427.

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

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try and see if it works: A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*, 514-529.

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, *24*, 467-563.

Eaton, J. F., Anderson, C. W., & Smith, E. L. (1984). Students' misconceptions interfere with science learning: Case studies of fifth-grade students. *The Elementary School Journal, 84*, 365-379.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25-42.

Fugelsang, J. A., & Thompson, V. A. (2003). A dual-process model of belief and evidence interactions in causal reasoning. *Memory & Cognition, 31*, 800-815.

Halloun, I. A., & Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53*, 1056-1065.

Kloos, H. (2007). Interlinking physical beliefs: Children's bias toward logical congruence. *Cognition*, *103*, 227-252.

Kuhn, T. S. (1957). *The Copernican Revolution*. Cambridge, MA: Harvard University Press.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*, 497-521.

Libarkin, J. C., Anderson, S. W., Dahl, J., Beilfuss, M., Boone, W., & Kurdziel, J. P. (2005). Qualitative analysis of college students' ideas about the Earth. *Journal of Geoscience Education*, *53*, 17-26.

Littmann, M. (2004). *Planets beyond: Discovering the outer solar system*. New York: Dover.

Mazens, K., & Lautrey, J. (2003). Conceptual change in physics: children's naive representations of sound. *Cognitive Development*, *18*, 159-176.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models*, Hillsdale, NJ: Erlbaum.

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

Shtulman, A. (2009). Rethinking the role of resubsumption in conceptual change. *Educational Psychologist*, *44*, 41-47.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution Challenges* (pp. 47-65). Cambridge, UK: Oxford University Press.

Shtulman, A., & Calabi, P. (2013). Tuition vs. intuition: Effects of instruction on naive theories of evolution. *Merrill-Palmer Quarterly, 59*, 141-167.

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24*, 261-289.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*, 165-173.

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories. *Cognition and Instruction, 25*, 337-398.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*, 1275-1289.

Trend, R. (2000). Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science. *International Journal of Science Education, 22*, 539-555.

Trundle, K. C., Atwood, R. K., & Christopher, J. E. (2007). A longitudinal study of conceptual change: Pre-service elementary teachers' conceptions of moon phases. *Journal of Research in Science Teaching, 44*, 303-326.

Tsai, C. C., & Chang, C. Y. (2005). Lasting effects of instruction guided by the conflict map: Experimental study of learning about the causes of the seasons. *Journal of Research in Science Teaching, 42*, 1089-1111.

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

Wiser, M., & Amin, T. (2001). Is heat hot? Inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction, 11*, 331-355.

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children. *Child Development, 85*, 160-175.