Title
Inconsistencies Among Beliefs as a Basis for Learning via Thought Experiments

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Author
Bascandziev, Igor

Publication Date
2020

Peer reviewed

Inconsistencies Among Beliefs as a Basis for Learning via Thought Experiments

Igor Bascandziev

Rutgers University, Newark
Department of Psychology
101 Warren Street, Newark, NJ 07103

Abstract

Although many studies have shown that being exposed to empirical data that contradict one's beliefs can lead to learning, it is not clear whether calling attention to inconsistencies among beliefs without the provision of new data, leads to learning. The present study asked whether calling attention to inconsistent beliefs via thought experiments leads to belief revision. Five-hundred-seventy-five participants were assigned to three different conditions in a pre-training, training, post-training design. The results showed that participants generated inconsistent beliefs between pre-training and training, but they did not spontaneously revise them at post-training (Baseline Condition). They did revise them, however, when they were asked to reason about the implications of the training thought experiments (Warning Condition) and when they saw an explicit inference drawn from the training thought experiments (Explicit Inference Condition). These results show that, with prompting, scientifically naïve adults can learn from thought experiments.

Keywords: thought experiments; learning; belief revision; naïve physics

Introduction

The richness of the human conceptual repertoire is unique in the natural world. However, our theories about the world and our concepts, which are atoms of beliefs, do not always agree with nature. Often, our theories are wrong, just like the theory about phlogiston in the history of science was wrong. The process of theory change, conceptual revision, and conceptual change is hard and prolonged, and it relies on specialized learning mechanisms (Carey, 2009). One important factor that motivates theory revisions are noticed contradictions between the predictions made by the learner on the basis of her conceptual system and data that come either from other people (Harris, 2012) or from direct empirical observation. Indeed, being exposed to data that contradict currently held beliefs leads to more exploratory behavior, more explanations, and more learning (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Legare, Gelman, & Wellman, 2008; Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Stahl & Feigenson, 2015). However, it is not clear whether exposing inconsistencies in the learner's system of beliefs, without the provision of new data, would lead to learning.

Thomas Kuhn (1977) proposed that conceptual change could be motivated by thought experiments (TEs). Thought experiments are experiments conducted in the head. They are typically presented as a narrative, which invites the learner to imagine a scenario where the learner applies her concepts as she usually does in the real world. The learner runs the experiment and sees the outcome (Brown & Fehige, 2014; Nersessian, 1992). Sometimes, the outcome of a thought experiment are two solutions that contradict each other, suggesting that there must be something wrong with the learner's concepts. It is thought experiments like this that according to Kuhn (1977) can reveal the exact way in which nature does not agree with the learner's conceptual system and can thus motivate conceptual revision. Given the lack of empirical studies, it is not clear whether scientifically naïve learners can benefit from such thought experiments. The present study begins to address this question.

The Machinery of TEs

There are several, not mutually exclusive ways in which thought experiments could generate outcomes. First, thought experiments could work just like any other inductive or deductive argument where the learner begins from some known premises and reaches a novel conclusion (Norton, 2004). Alternatively, thought experiments could be more like real experiments in the sense that they rely on quasi-observational or imagistic simulations (Gendler, 2004). In this case, conducting a thought experiment would consist of running a simulation where the process of simulating is similar to data collection. Finally, thought experiments could rely on model-based reasoning where the learner develops a mental model and runs simulations of the model, which is the basis for generating new outcomes (Nersessian, 1992). Any of the processes listed above could generate new outcomes, and in this paper, I am not trying to differentiate between them. Here, the focus is on a special case where the outcome of a thought experiment is inconsistent with other beliefs held by the learner, and therefore inconsistent with outcomes that could be drawn from other TEs. Would the realization that one's beliefs are inconsistent with each other lead to conceptual work and belief revision, and if yes, then how does the learner decide which beliefs are true?

Learning by thinking

That noticing inconsistencies in one's beliefs should benefit learning is supported by findings showing that providing self-explanations or conducting mental simulations is epistemically beneficial (Lombrozo, 2019). Some possible reasons for why this might be the case is that simulations allow learners to extract isolated representations from different modules and use them in downstream reasoning, which would not happen otherwise (Aronowitz & Lombrozo, in press). Furthermore, producing explanations allows learners to directly compare the explananda with prior beliefs and prior knowledge (Williams & Lombrozo, 2013), and they allow the learners to have better metacognitive awareness of inconsistencies or lack of explanatory depth (Chi, De Leeuw, Chiu, & Lavancher, 1994; Rozenblit & Keil, 2002).

There are, however, many reasons why pointing out inconsistencies in one's beliefs might not benefit learning. First, it is not clear that learners would even notice inconsistencies in their beliefs. One possible obstacle might be the inability to see deep structural relationships between various beliefs. Learners may hold inconsistent beliefs without ever realizing that they are structurally related to each other (Gentner & Markman, 1997; Gick & Holyoak, 1980; 1983). Another obstacle, as recent evidence suggests, is that both scientifically naïve individuals and scientists seem to be blissfully unaware of harboring what appear to be naïve theories and scientific theories that are in conflict with each other (e.g., see Shtulman & Harrington, 2016; Shtulman & Legare, in press; Shtulman & Lombrozo, 2016; Shtulman & Valcarcel, 2012). Although these studies do not offer examples of conflicting beliefs that exist at a conscious level, they still show that there are underlying competing representations and computations that produce outputs that are in conflict with each other and can therefore sometimes, e.g., under speeded conditions (Goldberg & Thompson-Schill, 2009) or because of weakened EFs (Tardiff, Bascandziev, Sandor, Carey, & Zaitchik, 2017), produce conflicting beliefs that reach consciousness. Finally, even if the learner notices some explicit inconsistencies, there are other potential pitfalls. For example, even though humans are generally motivated to maintain internal consistency among beliefs, values, and actions (Festinger, 1957), as is well known, the ways by which cognitive dissonance is avoided are not necessarily epistemically prudent (Hart, Allbarracin, Eagly, Brechan, Lindberg, & Merril, 2009), and they may include processes such as ignoring of conflicting information.

In the present study, I asked two general questions. First, could thought experiments be used to show that individuals are harboring inconsistent beliefs? Second, would individuals who generate inconsistent beliefs move to revise them, and if yes, then in which direction, and under the influence of which factors? I asked these questions in the context of a scientific understanding of forces and motion. According to Newtonian mechanics, no force is required for an object at rest to remain at rest and no force is required for an object in motion to remain in motion. Although many people may know how to recite this law, it does not mean that their underlying concepts of force and motion are Newtonian. Indeed, the mastering of Newtonian concepts is very difficult (Halloun & Hestenes, 1985). The naïve understanding of forces and motion resembles the medieval impetus theory, according to which objects set in motion receive an impetus (a force that is inside the object while it is moving), which slowly dissipates and becomes weaker, leading to the slowing down and eventual stopping of the object (Clement, 1982; McCloskey, 1983; McCloskey, Caramazza, & Green, 1980). Although thought experiments have been traditionally used to elicit beliefs that are consistent with the impetus theory (e.g., see McCloskey, 1983), other representations derivable from perceptual memory or implicit in perceptual-motor procedures are consistent with Newtonian mechanics, and could be elicited by thought experiments that highlight relevant contexts. In the present study, I contrasted thought experiments designed to elicit beliefs consistent with the impetus theory to thought experiments inspired by Galileo Galilei and designed to elicit beliefs consistent with Newtonian mechanics in a Pre-Training (Impetus TEs), Training (Newtonian TEs), Post-Training (Impetus TEs) design. I asked whether participants would spontaneously notice inconsistencies between their Pre-Training and Training beliefs, which would lead them to change their initial beliefs (Baseline Condition) and whether this would be augmented by receiving a warning to monitor for such inconsistencies (Warning Condition) or by seeing an explicit inference that follows from the Training TEs and directly contradicts the impetus outcomes from the Pre and Post-Training TEs (Explicit Inference Condition).

Method

Participants

I recruited 600 participants (200 per condition) from the US on Amazon's Mechanical Turk platform. A total of 25 participants were excluded (10 from the Baseline Condition; 7 from the Warning Condition; and 8 from the Explicit Inference Condition) because they failed the control questions embedded in the survey. The final sample consisted of 575 participants with an average age of 36 (range = 18 – 74; *SD* = 11.25). Two-hundred-seventy-nine identified as females, two-hundred-ninety-two identifies as males, and 4 identified as non-binary. A total of 313 participants had taken a high-school or college-level physics class (130 took both; 154 took high-school physics only; and 29 took college physics only) and 262 had not taken any physics classes.

Design and Stimuli

The experiment consisted of 4 Pre-Training TEs designed to elicit impetus responses, 4 Training TEs designed to elicit Newtonian responses, and 4 Post-Training, which were the same 4 TEs from Pre-Training. After running and answering each TE, participants reported the basis for their response (e.g., theoretical knowledge, specific memories, simulation of the event), and how confident they were in the basis of their response.

The 4 Pre- and Post-Training TEs were inspired by previous work (e.g., Clement, 1982; McCloskey, 1983; McCloskey et al., 1980) that used thought experiments (i.e., narratives that invited participants to imagine a scenario and give a qualitative prediction about what would happen) to uncover misconceptions that people have about forces and motion. All 4 thought experiments in the present study were variations on a theme: a vehicle is moving at a constant velocity in a straight line; an object that was attached to the vehicle is either dropped downwards or launched upwards;

participants are asked to describe either the trajectory of the object or where it would land. Here is an example of one such thought experiment: "A train is moving at a constant velocity of 100 MPH in a straight line. Inside the train, there is a mechanical claw that is holding a ball. The mechanical claw is fixed and rigid and so it does NOT move as a result of vibrations. Furthermore, the claw is located halfway along the ceiling between the front and the rear ends of the car. At one point, the ball is released. Please ignore air resistance. There is no wind inside the car. The ball will fall: a) Behind the halfway point of the car floor b) Exactly on the halfway point of the car floor." A typical, incorrect answer that is consistent with the impetus theory of motion would be that the ball would fall behind the halfway point of the car floor because once the ball is detached from the train, it will start losing the force that kept it moving, and it will slow down relative to the train.

The 4 Training TEs were designed to abstract the Newtonian law that no force is required for bodies at rest to remain at rest and no force is required for bodies in motion to remain in motion. All 4 thought experiments asked participants to reason about the forces that act on a body while the body is on a vehicle that travels at a constant speed (e.g., a train or the spinning Earth), and thus realize that no such forces act on the body as a result of the motion of the vehicle. Here is an example of a Training TE: "Earth spins from west to east at a constant velocity of ~ 1000 MPH at the equator. Imagine yourself at the equator, standing on a flat surface, facing east. Would you feel that you are being pushed backward and you need to use your own force to remain in one place? a) Yes, I would feel forces pushing my body in a direction opposite from the direction in which the earth is moving b) No, I would not feel any forces pushing my body in a direction opposite from the direction in which the earth is moving."

In the Warning Condition, before receiving the Training TEs, participants received a note that they will next see problems designed to improve their understanding of the phenomena they reasoned about at Pre-Training. It asked them to "think about the implications of their answers, and about how they relate to their previous answers." It also asked them to "think about whether they would have answered the previous questions differently, had they been aware of the answers to the new questions."

In the Explicit Inference Condition, after receiving the Training TEs, participants received an explicit inference that followed from the Training TEs, saying that "… no forces are pushing you in a direction opposite from the direction of the motion. That means that if you jump upward, the earth's ground (or the vehicle's floor) will not escape beneath your feet. It also means that if you are dropped from a cliff, you will fall in a straight line, and while you are falling, you will not lose any speed relative to the ground (floor) beneath you." Participants were asked if they agreed with this statement.

Procedure

The surveys were prepared and administered via Qualtrics. After giving their consent, participants proceeded to the TEs. The demographic questions (age, gender) and whether they had taken any physics classes were asked at the end of the survey.

Results

A preliminary analysis showed that the relationship between previous physics education (high-school and college-level physics) and performance at Pre-Training was weak (*r*s < .1) and not statistically significant (*p*s>.05), although it was trending toward being significant. Unlike many other studies in which participants are college-age students who have recently been in contact with physics classes, the present sample was comprised of participants with a much wider age range, many of whom had not had any contact with formal education in years or even decades. It is likely that this is why the magnitude of the effect of education was very small. The remaining analyses were conducted by collapsing over this variable.

Pre-Training Performance. The average performance of all 575 participants at Pre-Training was $M = 2.32$ (out of 4); $SD = 1.46$; range = 0 – 4. A one-way ANOVA with Pre-Training performance as the dependent variable and Condition as the independent variable showed that participants' performance was comparable across conditions at Pre-Training ($F(2, 572) = .49$, $p = .61$). The average confidence across all 4 TEs that participants had in the bases of their answers at Pre-Training was $M = 3.43$ (on a 1 to 5 scale); $SD = 1.04$; range = 1 – 5. This initial result replicates previous findings showing that somewhere between 47% and 68% (depending on the problem) gave correct Newtonian answers, whereas the majority of incorrect answers (30% to 51%) were curvilinear impetus responses (McCloskey et al., 1980). In the present study, the Pre-Training items were answered correctly by 59%, 67%, 49%, and 56% of all participants. The similarity and persistence of the impetus like reasoning in the different samples (i.e., undergraduate students vs Amazon MTurk participants with a time gap of nearly 40 years between the two; i.e., 1980s in the case of McCloskey (1980) vs the present study), is striking.

Training Performance and Comparison with Pre-Training. In contrast to the Pre-Training performance, many more participants gave correct Newtonian responses to the training TEs. The average performance was $M = 3.29$ (out of 4); $SD = 1.02$; range = 0 – 4. Broken down by item, 67%, 85%, 86%, and 90% of all participants gave Newtonian responses on the Training TEs. These results show that many of the participants who had given impetus like responses at Pre-Training, gave Newtonian responses at Training. In other words, the very same people seem to espouse beliefs that contradict each other. To check that participants were comparable across the 3 different conditions, I conducted a one-way ANOVA with performance on Training TEs as a dependent variable and Condition as an independent variable,

Table 1. Average performance at Pre- and Post-Training by condition.

	Baseline C.	Warning C.	Explicit Inference C.
Pre-Training	2.37	2.35	2.23
Post-Training	2.43	2.75	2.61

which confirmed that there were no significant differences between the 3 conditions ($F(2, 572) = 1.61$, $p = .21$.

I expected the confidence that participants had in the basis of each answer on the Training TEs would be higher than their confidence at Pre-Training, because the Training TEs were based on direct perceptual-motor memories about the kinds of forces that we feel on our own bodies while on a vehicle that is moving at a constant speed. This prediction was confirmed. The average confidence was M = 3.88 (on a 1 to 5 scale); $SD = .96$; range = 1 – 5. This was significantly higher than the confidence at Pre-Training ($M = 3.43$) ($t(574) = 13.61$, $p < .001$). Importantly, there were no significant differences among participants in the three conditions in terms of their confidence levels at Training ($F(2, 572) = 1.54$, $p = .22$).

Change Between Pre- and Post-Training. Did participants spontaneously notice inconsistencies between their Pre-Training and Training beliefs, and did that lead to change between Pre- and Post? Table 1 presents that Pre- and Post-Training performance by Condition. A 3 (Condition) x 2 (Test) repeated measures ANOVA found a significant effect of Test ($F(1, 572) = 29.30$, $p < .001$, $\eta^2 = .05$) and a significant Test x Condition interaction ($F(2, 572) = 4.82$, $p = .008$, $\eta^2 = .02$). Next, I computed the simple effect of Test for each of the three conditions. As the inspection of Table 1 suggests, only participants in the Warning ($F(1, 572) = 20.90$, $p < .001$, $\eta^2 = .04$) and the Explicit Inference condition ($F(1, 572) = 17.90$, $p < .001$, $\eta^2 = .03$) changed their answers between Pre- and Post-Training and the change was in the direction of giving more answers consistent with Newtonian mechanics at Post-Training. There was no significant change in the Baseline condition ($F(1, 572) =.35$, $p = .56$, $\eta^2 = .001$). In summary, although participants across all three conditions were confident in the bases of their Training TE outcomes, only those in the Warning and Explicit Inference conditions changed their Post-Training responses in the direction consistent with the Training TEs. It is important to note that the change between Pre- and Post in the Warning and the Explicit Inference conditions was not a result of changes on a small number of thought experiments between Pre- and Post-Training. A series of McNemar Chi Square tests showed a significant change in the number of participants who switched from impetus like responses at Pre-Training to Newtonian responses at Post-Training on *all* 4 TEs in both the Warning and Explicit Inference conditions (ps < .05), whereas there was no significant change on *any* of the TEs in the Baseline condition (ps > .22).

Did the participants who accepted the relevance of the Training TEs for the Post-Training TEs in the Warning and Explicit Inference Conditions also change their confidence between Pre- and Post-Training? As reported above, participants were more confident in the bases for Training TEs than for Pre-Training TEs. If they used the Training TEs as a basis for their Post-Training decisions, then we should expect that their confidence will increase between Pre- and Post-Training. A 2 (Pre and Post Confidence) x 3 (Condition) repeated measures ANOVA found a significant effect of Pre-Post Confidence (($F(1, 572) = 31.35$, $p < .001$, $\eta^2 = .05$) and a significant Test x Condition interaction ($F(2, 572) = 6.19$, $p = .002$, $\eta^2 = .02$). Next, I computed the simple effect of Pre-Post Confidence for each of the three conditions and found that the Confidence levels changed significantly only in the Warning ($F(1, 572) = 30.19$, $p < .001$, $\eta^2 = .05$) and the Explicit Inference ($F(1, 572) = 13.49$, $p < .001$, $\eta^2 = .02$) conditions, but not in the Baseline condition ($F(1, 572) =.30$, $p = .58$, $\eta^2 = .001$). This pattern suggests that participants in the Warning and Explicit Inference Conditions relied on the Training TEs when reasoning about the Post-Training TEs, which led to increased confidence.

Discussion

Although many previous studies have shown that exposing learners to empirical data that contradict their beliefs leads to more exploratory behavior and more learning, to my knowledge, there have been no empirical studies that have tested whether pointing out inconsistencies in learners' beliefs, via TEs, can lead to learning. The main goals of the present study were to: i) establish that thought experiments can be used to expose inconsistent beliefs within participants; and ii) establish whether scientifically naïve participants would notice those inconsistencies and change their beliefs. The first important and novel finding in the present study is that thought experiments can indeed be used to elicit inconsistent responses within individuals. I found that the very same people can give impetus like responses in one set of thought experiments and Newtonian responses in a different set of TEs. The second important finding is that participants did not spontaneously use the outcomes from the Training TEs in order to change their beliefs between Pre- and Post-Training, even though their confidence in the Training TEs was higher than that in the Pre-Training TEs. The third important finding is that pointing out the relevance of the Training TEs (Warning Condition) and providing the participants with an explicit inference that followed from the Training TEs (Explicit Inference Condition) resulted in belief revision between Pre- and Post-Training in a direction that was dictated by the higher confidence Training TEs.

These results are consistent with the view that thought experiments can be used to highlight contradictions among beliefs and that doing so may provide a context for belief revision in the service of conflict resolution, which may in turn motivate conceptual work and conceptual revisions (Kuhn, 1977). However, the evidence from the present study suggests that this happens only under certain circumstances. First, it is not sufficient to simply execute the TEs, as participants did in the Baseline Condition. That participants revised their beliefs in the other conditions suggests that in addition to executing the TEs, learners need to draw the relevant inferences (Explicit Inference Condition) and they need to compare those inferences side by side with the inferences following from the other TEs (Warning Condition). It is important to note, however, that even after going over these steps, not all participants who produced Newtonian responses at Training decided to change their beliefs at Post-Training. For example, even though 74% of all participants in the Explicit Inference condition agreed with the inference itself, at Post-Training they, on average, gave Newtonian responses only 65% of the time. This suggests that many of the participants who were giving impetus like responses on Pre- and Post-Training TEs and Newtonian responses on Training TEs either did not see any contradictions to be resolved or did not know how to resolve them. This is consistent with the literature showing how persistent naïve theories about physics are (Clement, 1982; Halloun & Hestenes, 1985; McCloskey, 1983).

Could these results be explained by appealing to experimental demands, namely that participants changed their Pre-Training answers not because they understood something they did not understand before, but because they thought that that was the goal of the experiment. This is unlikely for several reasons. First, participants in the Baseline Condition also received the same Pre- and Post-Training TEs, with the Training TEs in between. Receiving the same questions twice could have been interpreted as a signal that they should change their answers the second time around. They did not. Second, in the Warning Condition participants were asked to consider the outcomes of the Training TEs when they reason about the Post-Training TEs. If they did not understand what the relationship between the two was, the change between Pre- and Post-Training could have gone in any direction (i.e., from impetus to Newtonian and vice versa). Similarly, in the Explicit Inference Condition, participants had to decide on their own whether they agreed with the inference that connected the Pre/Post TEs with the Training TEs. As in the Warning Condition, if they did not understand the connection, the change between Pre- and Post-Training could have gone in either direction. Indeed, 25% did not agree with the inference.

The Training TEs in the present study were designed to draw from perceptual-motor memories about the kind of forces we feel on our own bodies when we are on a vehicle that is moving at a constant speed. This design was employed in order to generate a high number of accurate Newtonian responses about which participants would be highly confident. As reported in this study, executing such thought experiments was not sufficient. What is not clear is whether the high-confidence, bodily aspect of the thought experiments is necessary. Would participants still improve if the Training TEs asked them to reason about other physical objects? Although the present study cannot answer this question, preliminary results from a follow-up experiment suggest that the answer is yes.

In conclusion, the present study has shown that calling attention to inconsistencies in one's belief system via thought experiments can lead to learning. Furthermore, it has shown that just bringing the inconsistencies to the surface in a form of explicitly stated beliefs, is not sufficient for learning to take place. Additional work is required in the form of drawing inferences that follow from certain beliefs and actively comparing those inferences with other currently held beliefs. This finding is important because on the one hand it shows something fundamental about our naïve theories and the tolerance we have for inconsistent beliefs, and on the other hand it points toward potential mechanisms that are necessary for resolving inconsistencies, and therefore necessary for theory refinement and learning.

Acknowledgments

This research was supported by the Harvard University Mind Brain and Behavior research grant and by the Reed College Stillman Drake Award.

References

Aronowitz, S., & Lombrozo, T. (in press). Learning through simulation. *Philosophers' Imprint*

Bonawitz, E. B., van Schijndel, T., Friel, D. & Schulz, L. (2012). Balancing theories and evidence in children's exploration, explanations, and learning. *Cognitive Psychology, 64,* 215-234.

Brown, J. R., & Fehige, Y. (2014). Thought Experiments. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.*

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

Chi, M., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanation improves understanding. *Cognitive Science, 18,* 439-477.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50,* 66-71.

Festinger, L. (1957). *A theory of cognitive dissonance.* California: Stanford University Press.

Gendler, S. T. (2004). Thought experiments re-thought and reperceived. *Philosophy of Science, 71,* 1152-1163.

Gentner, D., & Markman, A. B. (1997). Structural alignment in analogy and similarity. *American Psychologist, 52,* 45-56.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12, 306-355.*

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Goldberg, R. F., & Thompson-Schill, S. L. (2009).

Developmental 'roots' in mature biological knowledge. *Psychological Science, 20,* 480-487.
Halloun, I., & Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53,* 1056-1065.
Harris, P. L. (2012). *Trusting what you're told: How children learn from others.* Cambridge, MA: Harvard University Press.
Hart, W., Albaracin, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135,* 555-588.
Kuhn, T. S. (1977). A function for thought experiments. In P. N. Johnson-Laird and P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 274-292). New York, NY: Cambridge University Press.
Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*, 929-944.
Lombrozo, T. (2019). Learning by thinking in science and in everyday life. In A. Levy & P. Godfrey-Smith (Ed.), *The scientific imagination* (pp. 230-249). Oxford University Press.
McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models* (229-324), Hillsdale, NJ: Erlbaum.
McCloskey, M., Caramazza, A., & Green, B. (1983). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science, 210,* 1139-1141.
Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science* (pp.3-44). Minneapolis: University of Minnesota Press.
Norton, J. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy, 26,* 333-366.
Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science, 26,* 521-562.
Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. C. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition, 109,* 211-233.
Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8,* 118-137.
Shtulman, A., & Legare, C. H. (in press). Competing explanations of competing: Accounting for conflict between scientific and folk explanations. *Topics in Cognitive Science*
Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner & A. S. Baron (Eds.), *Core Knowledge and Conceptual Change* (53-71). Oxford, UK: Oxford University Press
Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124,* 209-215.
Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348,* 91-94.
Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., & Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: A case study of vitalist biology. *Cognitive Psychology, 95,* 145-163.
Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology, 66,* 55-84.