CHILD DEVELOPMENT

Child Development, xxxx 2020, Volume 00, Number 0, Pages 1–17

The Role of Alternative Theories and Anomalous Evidence in Children's Scientific Belief Revision

Patricia A. Ganea, Nicole E. Larsen, and Vaunam P. Venkadasalam
University of Toronto

Children's naive theories include misconceptions which can interfere with science learning. This research examined the effect of pairing anomalies with alternative theories, and their order of presentation, on children's belief revision. Children believe that heavy objects sink and light ones float. In a pre-, mid-, and post-test design, 5-year-olds ($N = 96$) were assigned to one of two conditions, where they were either exposed to an alternative theory about buoyancy and then observed anomalies (Explanation-First), or the reverse (Anomalies-First). At mid-test, children were more likely to revise their beliefs after exposure to an alternative theory than anomalies alone. At post-test, children revised their naïve belief when they had access to an alternative theory before the anomalous evidence than in the opposite order.

Much of children's early science learning is informal, and the naïve theories about scientific concepts that they build through observation and cultural learning can interfere with the formal learning of accurate scientific knowledge (Brewer, Chinn, & Samarapungavan, 1998; Vosniadou, 2013). For example, children's everyday experience with objects makes it difficult for them to grasp that people can live on the bottom of the earth ("Why don't they fall off?"; Hannust & Kikas, 2010) or that lighter objects can sink faster than heavier ones (Penner & Klahr, 1996). Children's informal experiences with the natural world provide the basis for many of the robust misconceptions which can influence the acquisition of accurate scientific knowledge (Carey, 1999; Vosniadou, 2013). There is evidence that some misconceptions persist into adulthood and co-exist alongside correct theories, despite later formal training (Coley & Tanner, 2012; Pine, Messer, & St. John, 2001; Potvin & Cyr, 2017; Shtulman & Valcarcel, 2012). These misconceptions, whether held explicitly or implicitly, influence our thinking and behavior, and are robust in the face of counterevidence (Potvin & Cyr, 2017; Shtulman, 2017). The implication of these findings is that early intervention is critical, because it allows children to develop robust scientifically accurate concepts that in time can outcompete naïve misconceptions (Kelemen, 2019).

Although substantial debate has played out among researchers and educators about the relative merits of direct instruction and discovery-based instruction (i.e., Furtak, Shavelson, Shemwell, & Figueroa, 2012; Klahr & Nigam, 2004), Kuhn (2007) emphasizes that *what* is being taught is at least as important a question as *how* it is being taught, and the delivery method should be informed by the content being taught. In domains where science misconceptions are present, a nuanced approach that considers both children's naïve theories and scientific reasoning skills may be advantageous. Understanding what instructional strategies to use requires an understanding of how children's conceptions and scientific reasoning skills influence their reasoning when presented with a new theory. In this article we aim to examine the effect of anomalous evidence, alone and in combination with an alternative theory, on children's scientific belief revision ability.

In the pioneering model of conceptual change proposed by Posner, Strike, Hewson, and Gertzog (1982), learning is viewed as a rational process of conceptual change in which the learner gradually comes to accept new ideas because the evidence is compelling and logical. Given that the learner's

We thank the families who participated in this research and the Language and Learning Lab members at the University of Toronto for their help with recruitment and coding. We also thank Myrto Grigoroglou, Angela Nyhout and Begum Ozdemir for feedback on an earlier version of this manuscript. This research was supported by an Insight grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) to Patricia A. Ganea.

Correspondence concerning this article should be addressed to Patricia A. Ganea, University of Toronto, 252 Bloor St. West, Toronto, Ontario M5S 1V6. Electronic mail may be sent to patricia.ganea@utoronto.ca.

 0009-3920/2020/xxxx-xxxx
DOI: 10.1111/cdev.13481

current conceptions influence how they view the world, it is important to determine how learners' current conceptions change when presented with new ideas or evidence. According to Posner et al.'s classical model, a new conception is unlikely to replace an old one unless the learner experiences some dissatisfaction or cognitive conflict. Presenting learners with anomalous evidence is advocated as an essential instructional goal, in that it would lead the learner to experience dissatisfaction with their existing conception, which would then make them more receptive to learning the new theory that explains the anomaly. Posner and colleagues argue that for complex scientific problems it is unlikely that students can be simply taught the fundamental principles of a scientific theory and that teachers must use anomalies to get students receptive to new ideas. This view of inducing conceptual change through anomalies has been very influential in the educational literature (Hardy, Jonen, Möller, & Stern, 2006; Potvin & Cyr, 2017) and fits well with the constructivist principles of learning, whereby the child is making sense of the world through activities that allow for exploration and experimentation (Hardy et al., 2006).

The proposal by Posner et al. (1982) is also consistent with the influential developmental account of the child as a naive scientist (Gopnik & Meltzoff, 1997). According to the "theory theory" account, just like in formal scientific theory change, the child actively attempts to understand the world around and does so by spontaneously formulating, testing, and revising hypotheses in light of new evidence (Gopnik & Meltzoff, 1997). This view postulates that children have coherent and abstract theories about various domains of knowledge, such as biology, psychology, or physics. Just as formal scientists, when encountering evidence disconfirming an existing theory, children engage in exploratory, hypothesis testing behaviors in an attempt to examine the fit between the data and their naïve theories (Gopnik, 2012; Gopnik & Wellman, 2012). Thus, the observation of anomalous evidence, data that are in conflict with a naïve theory, is considered a key step in the process of theory change; the anomalous evidence has the potential to highlight an explanatory gap in one's theory and lead the learner to hypothesize about alternative explanations.

Contrary to these prominent theories, a large body of research indicates limitations in individuals' ability to learn on the basis of anomalies. The question of whether and how anomalous evidence can be used effectively as an instructional technique to drive belief revision and conceptual development deserves further attention. There is evidence that children's naïve beliefs influence the type of evidence they attend to, in that they are more likely to explore evidence that contradicts their naïve beliefs (i.e., anomalies) than evidence that is consistent with them (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Karmiloff-Smith & Inhelder, 1974; van Schijndel, Visser, van Bers, & Raijmakers, 2015). For example, after observing identical evidence (i.e., a block balancing at its geometric center on a balance beam) 6- to 7-year-old children showed different object preferences, depending on their naïve beliefs. Children who were Mass theorists (i.e., believing that objects balance at their center of mass) explored the block more than did Center theorists (i.e., believing that objects balance at their geometric center). The Center Theory children, for whom the block's behavior was consistent with their naïve theory, explored a novel object more. This pattern was reversed when children were presented with a block that was balancing at the center of mass. Thus, children's exploration of evidence is influenced by whether the evidence is consistent or not with their naïve theory. But what effect does anomalous evidence have on inducing belief revision?

Developmental findings indicate that in the face of disconfirming evidence, children are conservative and tend to maintain their beliefs. When evidence is inconsistent with their existing knowledge, children tend to infer hidden causal variables, while maintaining their current beliefs (Bonawitz et al., 2012; Saxe, Tenenbaum, & Carey, 2005; Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Schulz & Sommerville, 2006). For example, when children were presented with evidence contradicting their theory about balance relations they revised their belief only when an auxiliary variable was not present (Bonawitz et al., 2012). In other words, if a hidden variable (a magnet) could be used to explain away the counterevidence, children maintained their naïve belief. When children could not explain away the evidence, by invoking another variable, they were more likely to revise their belief and make better predictions about balance relations. In the case of robust and long-lasting scientific misconceptions, children's tendency to hold onto their naïve beliefs in the face of anomalies is pronounced (Koslowski, 1996; Kuhn, 1989; Penner & Klahr, 1996; Zimmerman, 2007). In one study, Penner and Klahr (1996) found that, when given the opportunity to explore and test their naïve belief that heavy objects sink faster than lighter ones, most 10-, 12-, and 14-year-olds designed experiments to confirm their naïve belief. Of the small number of participants (8 out of 30) who designed informative experiments and produced anomalous evidence, all but one participant found a way to fit the anomalous observation (i.e., light object sinking faster) within their existing knowledge. These findings demonstrate that children's naïve beliefs have a pervasive effect not only on children's interpretation of surprising, anomalous evidence, but also on their ability to generate evidence to test their current theories. Most of the younger children in this study generated experiments to confirm their naïve theory (i.e., that heavy objects sink) rather than to test different factors that contribute to whether an object sinks or not when in the water (Penner & Klahr, 1996; see also Rappolt-Schlichtmann, Tenenbaum, Koepke & Fischer, 2007; Schauble, 1996).

Evidence from the educational and adult cognitive literature has similarly pointed out the minimal impacts of anomalies-based instruction on conceptual understanding (Chinn & Brewer, 1993; Duit, Treagust, & Widodo, 2008; Greenhoot, Semb, Colombo, & Schreiber, 2004; Klahr & Nigam, 2004; Penner & Klahr, 1996). Studies that compared different teaching conditions, via direct hands-on experience with anomalous evidence or hands-off approaches, where the learner is introduced to an alternative theory, show that learners are better at belief revision and knowledge generalization when they have access to an alternative theory rather than when they interact with belief-inconsistent evidence only (Masson, Bub, & Lalonde, 2011; Renken & Nunez, 2010). For example, research by Renken and Nunez (2010) has shown that adolescents are unlikely to revise the belief that the mass of an object affects its travel speed on a ramp when they conduct a belief-inconsistent experiment. In contrast, when they can read about a belief-inconsistent experiment, they are more likely to learn that objects fall at the same rate regardless of their mass and their ability to generalize that knowledge after a 3-month delay is also improved. Even when participants conducted valid experiments their prior beliefs did not change, suggesting that hands-on experience with anomalous evidence is not sufficient to lead to belief revision (Renken & Nunez, 2010). Another study by Masson et al. (2011) that examined revision of naïve beliefs about object motion showed that middle school children improved in learning about the parabolic shape of object motion only when they had received a tutorial on Newtonian principles of motion. Their beliefs did not show improvement in conditions where they had an interactive experience with realistic trajectories and with manipulating such trajectories.

To summarize, findings from both the developmental and educational fields indicate that direct experience with anomalous evidence has limited impact on the reduction of scientific misconceptions in both children and adults (Chinn & Brewer, 1998; Chinn & Malhotra, 2002; Hardy et al., 2006; Koslowski, 1996; Kuhn, 1989; Masson et al., 2011; Penner & Klahr, 1996; Renken & Nunez, 2010; Zimmerman, 2007). How can we increase learners' ability to benefit from the observation of anomalous evidence? Based on work with older children and adults, it has been proposed that providing the learner with an alternative theory can influence the interpretation of anomalous evidence and thus affect learning and belief revision (Chinn & Brewer, 1998; Masson et al., 2011). This effect is due to the alternative theory giving the learner access to information that draws attention to crucial aspects of a phenomena (Masson et al., 2011) and through which to interpret the surprising, anomalous evidence (Chinn & Brewer, 1998). As a result, the learner is less likely to discount anomalous evidence or to rely on their naïve theory to search for hidden variables as possible explanations for it.

In this research we examine the role of alternative theories, in the form of rich conceptual explanations, in promoting young children's ability to benefit from the observation of anomalous evidence. We extend on prior research using a tightly controlled design that contrasts anomalies-based instruction with explanation-based instruction, aiming to identify their individual and additive effects, and the effect of their sequence in the teaching process. Our design will parse out the effect of hands-on versus hands-off learning on a knowledge-rich task, where the learner has existing misconceptions. The majority of prior research on the effect of anomalies-based instruction has been conducted with school children and adults (i.e., Renken & Nunez, 2010, 2013; Hardy et al., 2006; Masson et al., 2011; Penner & Klahr, 1996). The current research extends this prior work by focusing on a sample of young children, whose theories may be more flexible and thus more responsive to anomalies compared to the samples used in prior research. We also build on recent research that has demonstrated the beneficial effect of rich conceptual explanations for young children's learning of new scientific knowledge, as shown by their improved scientific understanding when exposed to viable theories about concepts in the domain of biology (Ganea, Ma, & Deloache, 2011; Kelemen, Emmons, Seston Schillaci, & Ganea, 2014; Strouse & Ganea, 2016; Larsen, Venkadasalam, & Ganea, 2020; Venkadasalam & Ganea, 2018). We extend on this prior work by examining the effect of conceptual explanations (which provide an alternative viable theory to children's naive theory) on children's ability to capitalize on the observation of anomalous evidence, and their combined effect on children's scientific belief revision. As reviewed earlier, there is a lack of evidence that anomalies only can lead to belief revision, and there is a need to better understand how to best structure the learning environment to facilitate conceptual learning.

We focused on children's understanding of buoyancy, because this is a concept with common misconceptions (Hardy et al., 2006; Kuhn, 1993; Potvin & Cyr, 2017; Smith, Carey, & Wiser, 1985; Yue, Tomita, & Shavelson, 2008) and floating and sinking are activities widely used in science education (Kallery, 2015; Selley, 1993). Both children and adults have a strong bias to consider weight as the determining factor in what is causing an object to sink, and hold a strong, naïve belief that heavy objects sink faster than lighter ones (Penner & Klahr, 1996). Current results indicate that in the presence of belief-violating evidence, children find ways to fit the evidence (i.e., heavy object floating, light object sinking) to their naïve theory (Penner & Klahr, 1996). Here we examined the effect of pairing an alternative theory, which provides conceptual information about buoyancy, with belief-violating evidence, to determine both their individual and combined effect on children's understanding of what causes an object to sink or float. We also examined whether the timing of when children have access to an alternative theory, in relation to the observation of anomalous evidence, matters for their belief revision ability. Specifically, we asked whether children benefit from the provision of an alternative theory equally if the conceptual explanation is delivered before or after the observation of anomalous evidence.

Children's difficulty with the concept of buoyancy (the upward force on an object in liquid) stems from having to compare the relative densities of objects and water (Lehrer, Schauble, Strom, & Pligge, 2001; Smith et al., 1985). Children have great difficulty distinguishing density from weight (Wilkening & Cacchione, 2010), and as a result they tend to think that heavy objects sink and light objects float. When 5-year-olds notice anomalies to their intuitive theory, they sometimes spontaneously hypothesize about the material (e.g., wooden objects float; Selley, 1993). This is a promising step toward understanding density (i.e., mass volume ratio) because some materials are less dense than others and therefore sink at different rates. However, children have to move beyond the type of material and consider how mass is distributed. The goal in this study was to promote their ability to dissociate an object's behavior in water from its weight and to consider the role of shape (air-filled cavities and surface area) in explaining why objects sink or float. The alternative theory drew children's attention to the role of shape, by exposing them to content about how increasing the volume of an object, increases the buoyant force of the water and as a result the object is more likely to float (e.g., a ball of clay is more likely to sink than a flat, spread out piece of clay of similar weight).

In a pre-, mid-, and post-test design we assigned 5-year-old children to one of two conditions. In the *Explanation-First condition*, children received a rich conceptual explanation about buoyancy and then observed anomalous evidence in a guided activity. In the *Anomalies-First condition*, children first had the opportunity to observe the anomalous evidence and then they heard the conceptually rich explanation. Both of these learning strategies were delivered in developmentally appropriate ways. We used a picture book format to deliver the conceptually rich explanations, because picture book reading is an enjoyable activity for many children, and moreover, there is evidence that children learn new scientific information from picture books (Ganea et al., 2011; Kelemen et al., 2014; Venkadasalam & Ganea, 2018). For the anomalous evidence, we used a guided activity, to allow children to actively engage with real, physical objects (Nayfeld, Brenneman, & Gelman, 2011; Peterson & French, 2008) and to ensure that children could notice and produce the evidence with guidance.

Children's beliefs about sinking and floating were assessed through both predictions about whether pairs of objects would float or not when placed in the water and the justifications they gave. The pre-test measures indicated whether children held the misconception that heavy objects sink and light ones float. We expected that the majority of 5-year-olds would hold this misconception, consistent with evidence that this misconception is robust and can be found even among adults (Hardy et al., 2006; Penner & Klahr, 1996; Yue et al., 2008). The mid-test indicated whether children revised their belief after experiencing each type of learning opportunity (i.e., explanations or anomalous evidence). Consistent with prior evidence that children learn from conceptually rich explanations (Ganea et al., 2011; Kelemen et al., 2014; Strouse & Ganea, 2016; Venkadasalam & Ganea, 2018), we expected that children would be more likely to revise their beliefs when exposed to conceptually rich explanations than when observing belief-violating evidence alone. Children would be more likely to hold onto their misconceptions when experiencing anomalies only, given evidence that individuals often find ways to discount anomalous evidence (Bonawitz et al., 2012; Chinn & Brewer, 1998; Kuhn, 1989; Penner & Klahr, 1996; van Schijndel et al., 2015; Zimmerman, 2007).

The post-test scores enabled us to assess the combined effect of conceptually rich explanations and anomalous evidence, and their order of presentation, on children's belief revision ability. Overall, we expected that children's scores at post-test will be significantly higher than at mid-test—when they have experienced both learning opportunities than each learning opportunity (anomalous evidence or conceptual explanations) on its own. Regarding their order of presentation, we expected that children would have more gains in learning when they had access to an alternative theory while they observed the counterevidence. This was because the alternative theory has the potential to influence children's responses to anomalous evidence by providing a possible causal explanation for it (Chinn & Brewer, 1998) and possibly because the knowledge gained when observing anomalies from a concomitant theory may be more easily abstracted (Masson et al., 2011). As such, the children in the *Explanation-First condition* will maintain their knowledge gains from mid-test and score as well, if not higher, at post-test. For children in the *Anomalies-First condition*, we considered two possibilities. One possibility was that, given evidence that children can learn from conceptually rich explanations, once children receive the conceptual information, they would show similar gains to what the children in the *Explanation-First condition* showed at mid-test. The other possibility was that, children would not show gains when the alternative explanation is given after their experience with the anomalous evidence. There could be several reasons why this may be the case. One possibility may be that because at mid-test children were expected to discount the anomalies and preserve their misconception (i.e., rely on it for their predictions and justifications), in this process their naïve misconception would not only be maintained but also reinforced. Another possibility may be that when exposed to the new theory, children have more difficulty retrieving their memory of the anomalous examples they had seen to reinforce this new theory. This study would not differentiate between these two possible reasons. However, in both cases, at post-test children may be expected to learn from the explanation and improve their predictions relative to mid-test, but not to the same degree as children who received the explanation first (from pre- to mid-test).

Method

Participants

Ninety-six 5-year-old children ($M = 5.49$; range: 5.03–5.99, 48 males) participated in this study. Fourteen additional children were excluded because they had a perfect score on the pre-test ($n = 2$), failed comprehension questions about the story content ($n = 6$), had a receptive language score 2 *SD*s below the mean ($n = 1$), were inattentive and unable to complete the tasks ($n = 4$), or due to parental interference ($n = 1$).

Equal numbers of children were randomly assigned to one of two conditions: Explanation-First ($n = 48$, $M_{age} = 5.50$, 24 males), and Anomalies-First ($n = 48$, $M_{age} = 5.49$, 24 males). We developed two books and two guided activities to teach children about buoyancy. Within each condition, children read one of the books and completed one of the activities.

Participants were recruited from a database of families that have indicated they were interested in participating in research. Children were individually tested by a female experimenter in a quiet room at the laboratory location. The largest group of children were White (48%), but the sample also included Asian (19%), Latin American (3%), Black (2%), and Mixed Race (23%) children. An additional 5% of families declined to disclose ethnicity information. All children spoke English fluently. The majority of children came from middle-class families and the mode parental education level was a Bachelor's Degree.

Materials

Two different books and two different activities were developed to ensure that any differences in learning were not a function of the type of book or activity. All the materials and the activity scripts used in this research can be found at https://osf.io/cx9yu/?view_only=777e69a3cd1040e491c90c815d9537b2. All 16 possible combinations of the books and activities were included and were counterbalanced, such that six children received each possible combination. A nonfiction information book, and a narrative, fiction book, which contained the same conceptual information about buoyancy and shape were created and illustrated in the lab. The pictures used in each book were the same, but the text varied to fit the story type (see Figure 1 for samples of book pages). Based on previous research that showed that children can learn from both informational and fictional books, we expected no differences between the types of books used in this research (Venkadasalam & Ganea, 2018). Due to the similar design of the activities, no differences were expected between the activity types either.

For the first guided activity (Object Activity), children were presented with objects of varying weight and material and were asked to predict and then test whether they would sink or float in the water. Twelve objects were used in this activity, and were presented to children in three groups of four objects each. In this activity, children had a worksheet where they could record their predictions, and the results observed. In the second guided activity (Clay Activity), children made pieces of clay into different shapes and then tested whether they would sink or float in the water. Materials for the second activity were limited to two pieces of clay. As a measure of learning pre-, mid- and post-tests were administered. The materials for each test phase included four pairs of objects, for a total of 12 object pairs, divided into three object sets. Within each set of four pairs, two pairs of objects were the same weight, and two were different weights. Children had an opportunity to weigh these pairs of objects, so they knew which was heavier and which was lighter. Two of the pairs of objects were made of the same material and two were made of different materials. Materials included: metal, plastic, rubber, and glass. For each test phase, children received a different object set, but the order in which children received these sets were counterbalanced.

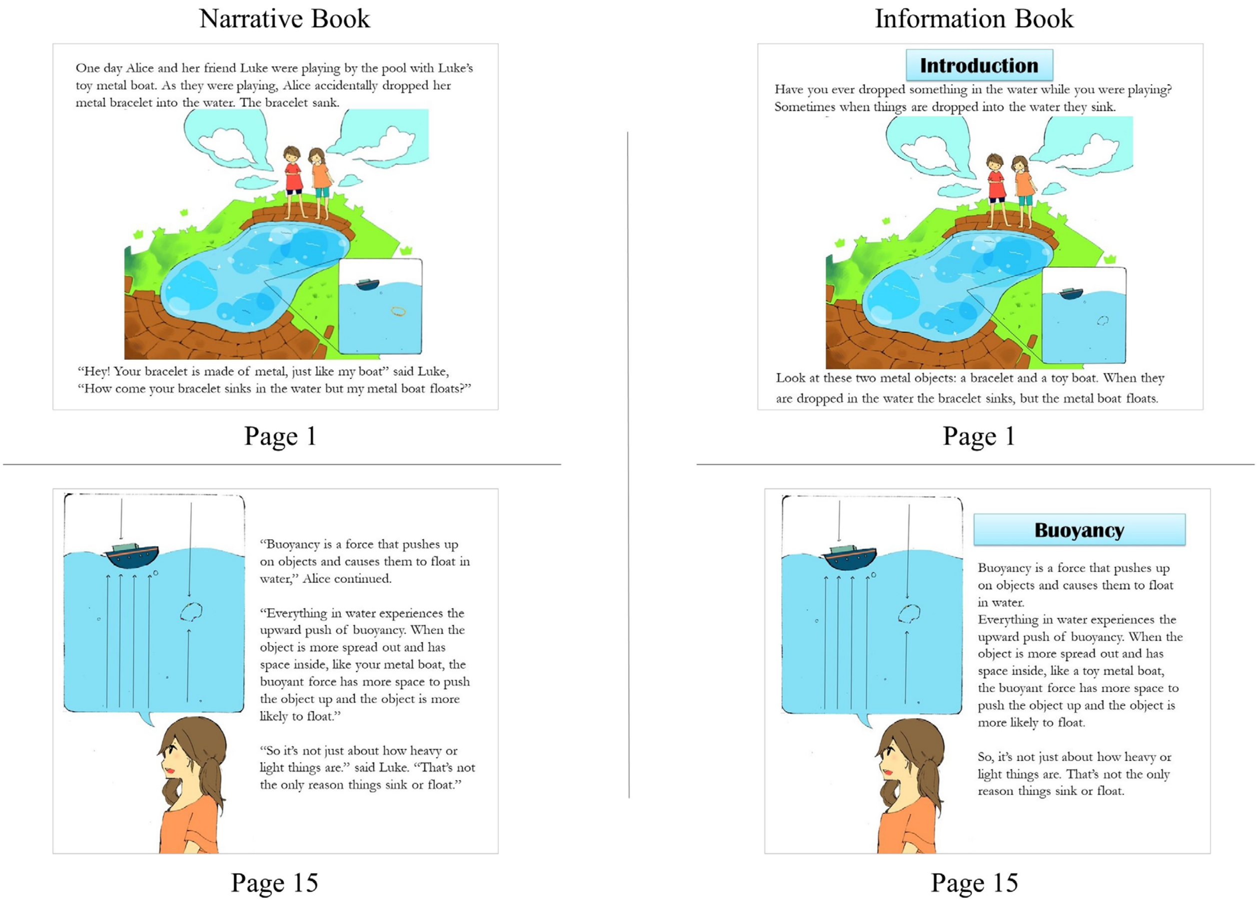

Figure 1. Sample pages from the narrative and information books (see Materials for full books at https://osf.io/cx9yu/?view_only=777e69a3cd1040e491c90c815d9537b2)

Procedure

There were six phases in this study: a language assessment, pre-test, learning phase one (depending on the condition, Explanation-First or Anomalies-First), mid-test, learning phase two (Anomalies/ Explanation, depending on the condition), and the post-test (see Figure 2). The entire session was video-recorded and lasted approximately 40 min–1-hr. Children were randomly assigned to the Explanation-First or Anomalies-First condition, and within each condition they were randomly assigned to read one book and complete one activity.

Language Assessment

Children's general language was assessed with the Toolbox Picture Vocabulary Test (TPVT). The TPVT is a receptive vocabulary measure administered in a computerized adaptive format (National Institutes of Health, 2015). Children were excluded if they scored 2 *SDs* below the age-standardized mean.

Learning Phase

In this phase, we used picture books to deliver the conceptually rich explanations and guided activities to present children with the anomalous evidence. Children in the Explanation-First condition were read the book prior to the activity, whereas those in the Anomalies-First were read the book following the activity. The goal of each phase (i.e., to learn how objects sink and float) was explicitly identified for children. During the book reading, the experimenter read either the nonfiction book or the fiction book to each child. To ensure that children paid attention during the book reading, they were asked four open-ended comprehension questions directly after the book reading (Book Questions can be found in Supplementary Materials). Children who answered more than two comprehension questions incorrectly were excluded. Additionally, children were asked an open-ended question about the book content (i.e., Why do you think that some objects in the book floated and some of them sank?). The learning prompt was asked in an attempt to get children to reflect on the content they were exposed to in the book. A similar learning prompt was asked after the children completed the guided activity as well.

During the activity, the experimenter guided children through either the Object Activity or the Clay Activity. The Object Activity involved children making predictions about whether different types of objects will sink or float. Children were shown 12 objects in total, presented in three groups. For each object group, children were asked to make a prediction and then record it on a worksheet. Then children tested the objects in the water to see if their predictions were right and then recorded these observation results on the worksheet as well. This activity closely matched the tests of learning measure described below. The Clay Activity involved manipulating two pieces of clay matched in weight to make various shapes that either floated or sank. Children were asked to make the pieces of clay into shapes they thought would float and sink and then tested the shapes in the water. The purpose of this phase was for children to observe that something with a set weight can both sink and float, depending on its shape. Children were guided through both activities to ensure they produced and observed anomalous evidence that countered their naïve theories. After the activity children were asked an open-ended question about the content of the activity (i.e., Object Activity: Why do you think that these objects floated and these objects sank?; Clay Activity: Why do you think that sometimes the clay floated and sometimes the clay sank?). The learning prompt was asked to get children to reflect on the anomalous observation experience.

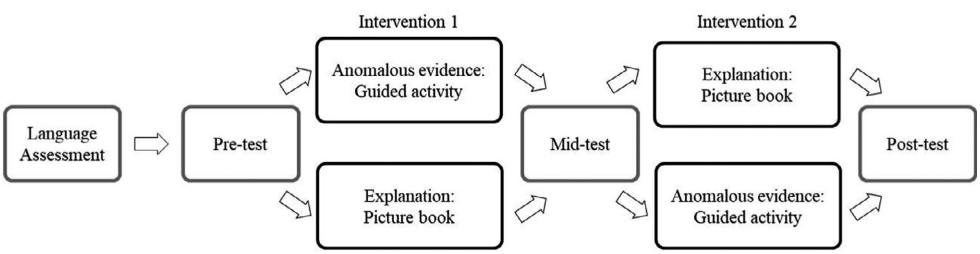

Figure 2. Schematic of the procedure.

Tests of Learning (Pre-, Mid-, Post-Test)

Each test phase (pre-, mid-, and post-test) followed the same procedure. Using one of the sets of objects, children were given each object pair, one pair at a time, to inspect. They were provided with a scale so they could weigh the objects in order to tell definitively which object was heavier. The experimenter also told the children what each object was made of so there was no ambiguity. To avoid differences in response patterns within the sample, within each object set the pairs of objects were presented in the same order to each participant (different weight/different material, same weight/same material, same weight/different material, and different weight/same material). We varied weight and material across object pairs as these are dimensions that children have misconceptions about. However, shape was the dimension that affected whether an object floated or sank in each pair of objects. Shape was salient both within and across object pairs, such that if children understood the effect of shape, they would be able to correctly identify the sinker and floater.

After children were given time to inspect and feel the objects, weigh them, and were told what they were made of, children were asked the prediction question: “If I took these two objects and put them into the water, which one would float on the top and which one would sink to the bottom?”. Following this question, children were asked to explain their predictions. The sequence of this questioning was counterbalanced, so half of the time children were asked “Why would this one float?” first, followed by “Why would this one sink?” and vice versa. While the question was phrased as an either-or statement, some children did indicate that they thought “both objects would float” or “both objects would sink”. Similarly, for these predictions, children were asked to explain their reasoning for each object separately. Children received neutral feedback (“Thank you”) after answering each question. Two children were excluded because they answered all four test questions correctly at pre-test and were able to explain their reasoning behind their answers with a high degree of accuracy (6/8 on a coding scale for explanations below).

Coding

Predictions

Children’s predictions were coded a score of 1 if children correctly identified which object in the pair would float and which object would sink. A score of 0 was assigned if children incorrectly identified the floater and sinker in each object pair, or if they said both objects would sink or both objects would float. Scores were summed across the four trials, such that the total score ranged from 0 to 4 for each test phase. Two research assistants coded 100% of the children’s responses from the video recordings. The coders were naive to the hypotheses of the study, the condition, and test phase. There was high interrater reliability determined by Cohen’s $\kappa = .91$, $p < .001$, a 95.66% agreement rate. The coders resolved disagreements through discussion.

Explanations

Children’s explanations were scored on a scale from 0 to 2. A score of 2 was assigned to explanations that focused on the distribution of mass, by mentioning an object’s shape, such as, for example describing floating objects as "spread out" or "having space inside" and sinking objects as "scrunched up" or "having no space inside". A score of 0 was assigned if children gave an explanation that incorporated a misconception (e.g., “Heavy objects sink” or “Plastic floats”), an explanation that had no discernable theory (e.g., “This will sink because it has polka dots.”) or answered “I don’t know”. A score of 1 was assigned to explanations that used a misconception in conjunction with an explanation that considered the distribution of an object’s mass, mentioning either the object’s shape and/or surface area (e.g., “It’s spread out and light so it floats”). Explanations also received a score of 1 if they described floating objects as "big" and sinking objects as "small", as this begins to show an understanding of the importance of surface area and shape. In these instances, it was clear that children used “big” or “small” to refer to the volume of the object and not as a substitute for heavy or light. Even when children were aware that the pairs had the same weight, they described the objects as big and small (e.g., “Object A floats because it is flat and big and object B sinks because it is smaller” or “Object A floats because it has space and it’s big. Object B sinks because it’s long and tiny”). These children’s explanations were distinctly different from children who said an object would sink because it is big and heavy, who received a score of 0. Scores were summed across the four trials, such that they ranged from 0 to 8 for each test phase. As with the predictions, 100% of the explanations were coded by naive research assistants. For the explanations, there was high interrater reliability between coders determined by Cohen's $\kappa = .86$, $p < .001$, a 93.49% agreement rate. The coders resolved disagreements through discussion.

Results

Two main analyses were conducted to examine the effect of condition on children's performance across the three test phases. First, we conducted confirmatory analyses to examine the predictions children made, followed by the explanations they provided, to test our predictions. We also explored the responses that children gave to the open-ended learning prompts. An exploratory analysis of the individual patterns of responses across both predictions and explanations is presented in Supplementary Materials (Table S1). In preliminary analyses, an independent *t*-test showed no differences in receptive language scores between the two conditions, $t(94) = -1.63$, $p = .11$, $d = .33$, 95% CI [−9.67, 0.96]. Preliminary analyses for predictions and explanations showed no effect of age in months or of gender for either condition, therefore these factors were not considered in the following analyses.

Predictions

Predictions were ordinal in nature (range: 0–4) and were analyzed using nonparametric tests. We ensured there were no differences between the scores at mid- and post-test as a result of the two types of books and activities used. A Mann–Whitney *U*-test indicated that prediction scores for mid- and post-test were similar for both book types (*p*s > .84), and activity types (*p*s > .12). As there were no differences found between the type of books and activities used in the intervention, these factors were collapsed in the following analysis. A Mann–-Whitney *U*-test showed that the pre-test scores were similar for the Explanation-First ($M = 1.31$, $SD = 1.32$) and Anomalies-First ($M = 1.56$, $SD = 1.25$) conditions at baseline, $U = 1017.00$, $z = -1.02$, $p = .31$. Table 1 displays the percentage of correct predictions children made in the three test phases for each condition.

A generalized estimating equation analysis with multinomial distributions and cumulative logit link functions was conducted to investigate whether children correctly predicted which object would sink and which object would float. This type of analysis was selected to accommodate the ordinal nature of the dependent variable and the presence of a within-subject factor (pre-, mid-, and post-test scores) in the data. The Anomalies-First condition and the pre-test were used as the reference groups for this test.

There was no main effect of condition ($p = .32$), nor a difference between pre- and mid-test ($p = .79$), nor pre- and post-test ($p = .10$). However, there was a significant interaction between condition and test phase. From pre- to mid-test children in the Explanation-First condition were more likely to make correct predictions, Wald $\chi^2(1) = 21.57$, $p < .001$, $b = 1.54$, $SE = .33$, compared to the Anomalies-First condition. Given their starting levels of understanding at the pre-test, the odds ratio for children in the Explanation-First condition exhibiting a higher level of understanding about sinking and floating at mid-test was 4.67, 95% CI [2.44, 8.94], relative to Anomalies-First condition. Children in the Explanation-First condition were also more likely to make more correct predictions from pre-to post-test, Wald $\chi^2(1) = 14.68$, $p < .001$, $b = 1.53$, $SE = .40$, compared to the Anomalies-First condition. Given their starting levels of

Table 1
The Proportion of Correct Responses, Classified Into Five Categories (0–4), as a Function of Test and Condition (The Actual Number of Responses Are in Parentheses With 48 Total Responses for Each Condition Per Test Phase)

	Test phase					
	Pre-test		Mid-test		Post-test	
Prediction score (out of 4)	Anomalies-First	Explanation-First	Anomalies-First	Explanation-First	Anomalies-First	Explanation-First
0/4	27% (13)	40% (19)	27% (13)	19% (9)	25% (12)	8% (4)
1/4	19% (9)	19% (9)	21% (10)	0% (0)	15% (7)	15% (7)
2/4	33% (16)	19% (9)	27% (13)	25% (12)	25% (12)	8% (4)
3/4	13% (6)	17% (8)	15% (7)	29% (14)	21% (10)	38% (18)
4/4	8% (4)	6% (3)	10% (5)	27% (13)	15% (7)	31% (15)

understanding at the pre-test, the odds ratio for children in the Explanation-First condition exhibiting a higher level of understanding about sinking and floating at post-test was 4.62, 95% CI [2.11, 10.11], relative to the Anomalies-First condition.

Post hoc Wilcoxon Signed Rank Tests were conducted using a Bonferroni correction to account for multiple comparisons (.05/6 = .0083), which set the alpha level at $p = .008$. There was a significant increase in children's prediction score in the Explanation-First condition between pre- and mid-test ($z = -4.64$, $p < .001$) and pre- and post-test ($z = -4.86$, $p < .001$), but not between mid- and post-test ($z = -1.50$, $p = .13$). In the Anomalies-First condition there was no significant increase in prediction scores for any of the test phases (*p*s > .13; see Figure 3).

Explanations

Explanation scores ranged from 0 to 8 which allowed us to treat them as continuous and analyze them using parametric tests. We also ensured there were no differences between the two types of books and activities for explanations. An independent *t*-test indicated that explanation scores for mid- and post-test were similar for both book types (*p*s > .13), and activity types (*p*s > .17), so they were collapsed in the following analysis. An independent *t*-test showed that the pre-test scores were similar across both conditions at baseline, $t(94) = -.17$, $p = .87$, $d = .03$, 95% CI [−0.81, 0.68], with a mean pre-test score of 1.10, $SD = 1.98$ for the Explanation-First condition and 1.04, $SD = 1.69$ for the Anomalies-First condition. The percentage of correct explanations children in each condition made across the three test phases can be found in Table S2 in Supplementary Materials. At pre-test, the majority of children in both conditions referenced weight in their explanations. For a breakdown of the type of misconceptions that children held about sinking and floating see Table S3 in Supplementary Materials.

A 2 (condition) × 3 (test phase) mixed-measures analysis of variance was run to determine the effect of condition and test phase on the accuracy of children's explanations. We found a significant main effect of test phase, $F(2, 94) = 59.83$, $p < .001$, $\eta_p^2 = .39$. Pairwise comparisons with Bonferroni corrections indicated that the mean scores at mid-test ($M = 2.83$, $SD = 0.26$) were significantly higher than scores at pre-test ($M = 1.07$, $SD = 0.18$), and that the post-test mean scores ($M = 3.96$, $SD = 0.32$) were significantly higher than scores at pre- and mid-test. There was a main effect of condition, $F(1, 94) = 19.29$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons with Bonferroni corrections revealed that the mean scores in the Explanation-First condition ($M = 3.55$, $SD = 0.30$) were significantly higher than those in the Anomalies-First condition ($M = 1.69$, $SD = 0.30$). Most importantly, there was also an interaction between test phase and condition, $F(2, 94) = 22.97$, $p < .001$, $\eta_p^2 = .20$. This interaction shows that children in both conditions performed differently across the three test phases.

We used estimated marginal means to determine the nature of this interaction, see Figure 4. This graph shows children's explanation scores in the

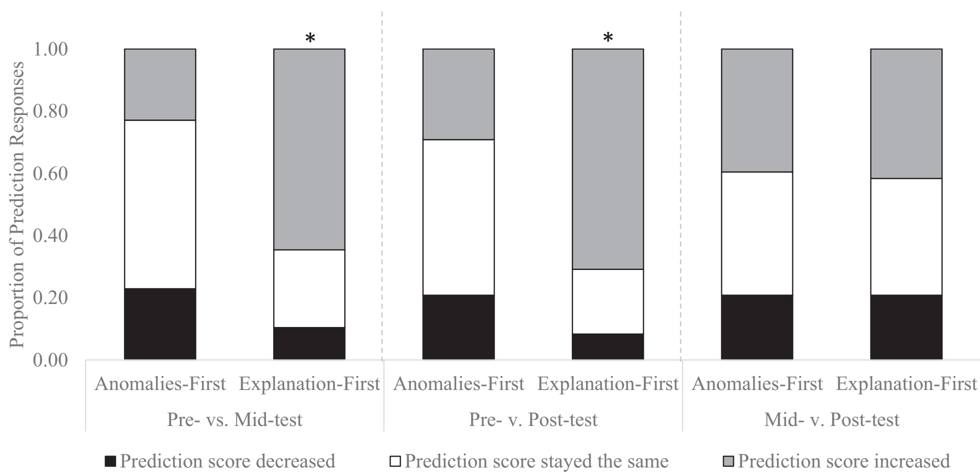

Figure 3. Proportion of prediction responses that decreased, stayed the same or increased across two test phases, as a function of condition.
*$p < .001$.

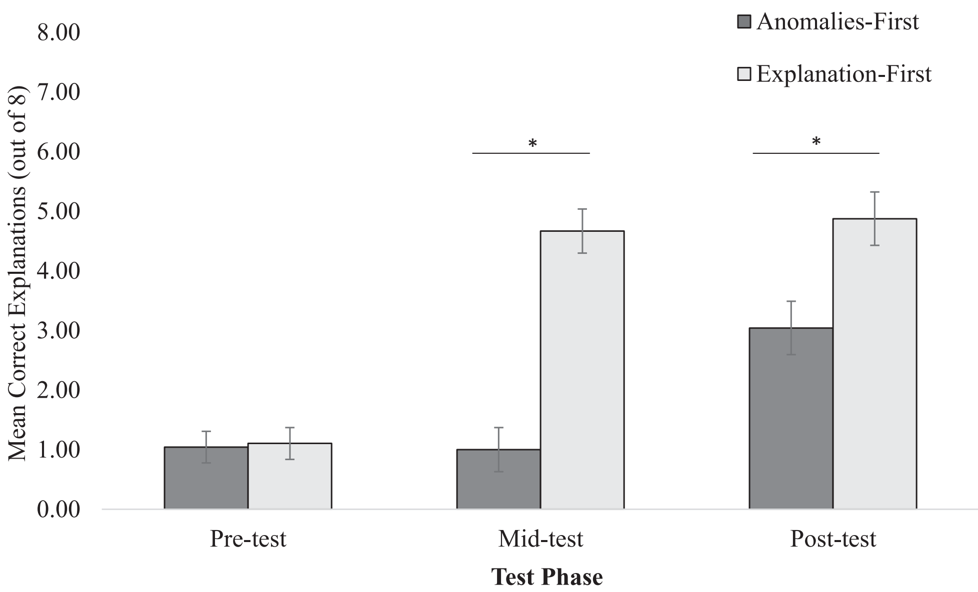

Figure 4. Mean correct explanation responses as a function of test phase and condition.
Note. Explanation scores ranged from 0 to 8 at each test phase, $*p < .001$.

Explanation-First significantly improved between pre-test to mid-test, after receiving the conceptual information about buoyancy, and that these gains were maintained from mid-test to post-test after observing the anomalous evidence. In contrast, children in the Anomalies-First condition did not make significant improvements from pre-test to mid-test, after observing the anomalous evidence. While there were improvements in the Anomalies-First condition between mid-test and post-test, after receiving the conceptual information, these improvements were still significantly lower than their counterparts in the Explanation-First condition at post-test.

Responses to Open-Ended Learning Prompts as a Function of Learning Opportunity

After each type of learning opportunity, children were asked an open-ended question that invited them to reflect on the activity/book and think of a reason why some objects floated and others sank. The responses to the open-ended questions after book reading indicate the extent to which children were receptive to the new explanation from the book, as shown by their reference of information from the book. The responses to the open-ended questions after the anomalous activity indicate whether children spontaneously mention alternative explanations for why the anomalies occurred.

The results mirrored the findings reported earlier for explanations. A Mann–Whitney *U*-test showed that, when asked an open-ended question at the end of the guided activity, children's explanations were significantly higher when the activity followed the book (Explanation-First condition, $M = 1.42$, $SD = 0.90$) than when the activity preceded the book (Anomalies-First condition, $M = 0.51$, $SD = 0.80$), $U = 552.50$, $z = -4.70$, $p < .001$. This indicates that overall children did not mention alternative explanations when presented with anomalies only and that their interpretation of the anomalous evidence benefited from the presence of an alternative theory. That is, children were more likely to hypothesize about other factors than weight in causing an object to sink or float, when they had access to an alternative theory while observing the anomalous evidence. Children's responses to the open-ended question after the book also mirrored the overall pattern of results earlier, in that their explanations were significantly higher in the Explanation-First ($M = 1.54$, $SD = 0.74$) condition than the Anomalies-First ($M = 1.00$, $SD = 0.90$) condition, $U = 775.00$, $z = -3.06$, $p = .002$. That is, children in the Explanation-First condition referenced the correct theory more than children in the Anomalies-First condition. This indicates that when children received the conceptual explanation after experiencing the anomalous evidence, their gains in learning from the book were dampened compared to when the conceptual information was given after the anomalous evidence.

Discussion

By the time they start formal schooling children have developed naïve theories about the world around them. In the physical domain, many of these theories include misconceptions which can interfere with later science learning. Consistent with prior research, the majority of children in this study held the naïve belief that heavy objects sink and light ones float. When asked to predict which of two objects would sink and which would float, children responded incorrectly on 64% of trials (246 out of 384) and the majority of them (88%, 84 of 96) used weight as the reference factor for at least one of the trials. In the current research we examined children's ability to revise this misconception when presented with anomalous evidence (e.g., heavy objects floating and light ones sinking) and conceptually rich information about buoyancy. The conceptually rich information provided an alternative theory to children's naïve beliefs. We also examined whether the order in which children are exposed to anomalous evidence and conceptually rich explanations matters for their belief revision ability.

The results from the mid-test relative to pre-test indicated whether children revised their belief after experiencing each type of learning opportunity (i.e., explanations or anomalous evidence). Consistent with evidence that individuals find ways to discount anomalous evidence (Chinn & Brewer, 1998; Klahr & Dunbar, 1988; Koslowski, 1996; Penner & Klahr, 1996), we found that children are resistant to setting aside their naïve theory when confronted with counter-evidence. In the Anomalies-first condition, children maintained their misconceptions—they continued to rely on their naïve belief about weight to make new predictions and justifications. This suggests that even if the observation of anomalous evidence may have led children to spontaneously generate new hypotheses about what may cause an object to sink or float (Selley, 1993), this did not manifest in their predictions or explanations. In contrast, in the Explanation-first condition, when children were exposed to rich conceptual explanations about what makes objects sink and float, children were more likely to rely on this information to make new predictions and justifications. This gives corroborating evidence that children can learn scientific information from picture books (Ganea et al., 2011; Kelemen, 2019; Kelemen et al., 2014; Strouse & Ganea, 2016; Venkadasalam & Ganea, 2018) and can apply this knowledge to reason and interpret new phenomena.

Our main goal was to examine the effect of conceptual explanations—which provided an alternative viable theory to children's naive theory about sinking and floating—on their ability to learn from the observation of anomalous evidence. Overall, children performed better at post-test when presented with rich conceptual explanations prior *to* observing anomalous evidence than the other way around. Compared to pre-test, children were more likely to give better predictions and explanations at post-test in the Explanation-First condition than in the Anomalies-First condition. However, when the alternative theory was given *after* the anomalous evidence, in the Anomalous-First condition, children's gains in performance were more reduced. This difference in belief revision across conditions, as a function of when the conceptual information was given, indicates that differences in the way children represent and explain anomalies (within their pre-existing theory or a new alternative theory) can lead to different learning outcomes. In the Explanation-first condition, children had access to an alternative mechanism to explain the anomalous evidence—the presence of air-filled cavities or the surface area of an object matters for whether an objects sink and float. In the Anomalies-first condition, an alternative causal explanation was not available to interpret the inconsistent evidence and children relied on their naïve theory in both their predictions and explanations. This pattern of responses fits with reports that individuals are not quick to abandon a theory when an alternative hypothesis is not readily available to explain the counterevidence (Chinn & Brewer, 1993; Klahr & Dunbar, 1988).

When we compare the change in performance from mid-test to post-test, two results are important to consider with respect to the effect of anomalous evidence in children's belief revision ability. First, the results indicate that observing the anomalous evidence after receiving the conceptual explanations did not lead to significant changes in children's predictions or explanations about sinking and floating. In other words, the anomalous evidence did not contribute over and above the positive effect of conceptual explanations on children's predictions and explanations at mid-test. Children retained their gains from mid-test and did not improve significantly after witnessing the counter-evidence. Intuitively this is surprising because one would expect that the hands-on experience with anomalous evidence would help children refine and deepen their understanding of the alternative theory they just learned and lead to greater learning outcomes. Future research may investigate whether integrating anomalous evidence with conceptual knowledge within the activity may lead to increased effects on learning, compared to when these two sources of knowledge are presented successively.

Second, the results also indicate that observing anomalies first interfered with children's ability to benefit from the subsequent explanations. For children in the Anomalies-First condition, we considered two possibilities. One possibility was that, once children received the conceptual information, they would show similar gains to what the children in the Explanation-First condition showed at mid-test. The other possibility was that, at post-test children would learn from the explanation and improve their predictions relative to mid-test, but not to the same degree as children who received the explanation first (from pre- to mid-test). Our findings show that children who observed anomalies first and then received conceptual information did not improve across test phases (from mid-test to post-test) in their prediction scores. At each test phase, children relied significantly on their naïve theory to predict which of two objects would sink or float. There was some improvement in children's explanations from mid-test to post-test, suggesting that some children incorporated the new conceptual information within their existing theoretical framework. Future research is needed to examine why the new conceptual information did not have the same impact in shaping children's predictions, compared to when the conceptual information was given first. We advanced two possible explanations, one that has to do with the reinforcing of the misconception as it was used to interpret the anomalous evidence and another that has to do with difficulties with integrating the new explanation with the prior experience of anomalous evidence. It is possible that the positive effect of a new theory on the observation of anomalies comes from the fact that one can interpret the evidence *as it occurs* in light of two concomitant perspectives. This process of comparison may lead to better abstraction and generalization of the knowledge gained through the interaction with the anomalous evidence. Exploring the mechanisms that underlie the effect of timing of an alternative theory on the interpretation of anomalous evidence is an exciting area for further research.

We have focused on how children respond to evidence in the context of a robust scientific misconception. The current findings have important implications for understanding the conditions under which children can engage in belief revision on the basis of counterevidence. Going back to the metaphor of the "child as a scientist," this research shows that, just like formal scientists, children are conservative and do not readily abandon their naïve theories when they witness anomalous evidence (Chinn & Brewer, 1993). Pioneering work by Karmiloff-Smith and Inhelder (1974) shows that even when they consider counterevidence, children first prefer to develop a new, independent theory, rather than change their current theory to come up with a unified account of the anomalous experience. The children in this research were more likely to accept a counterintuitive belief (e.g., that heavy objects float) when they had access to an explanatory causal mechanism that fit the anomalous evidence. In the absence of such a mechanism, children justified the anomalous evidence in terms of their naïve belief. The findings of this research fit with the prevalence teaching model proposed by Potvin (2013) to account for the coexistence of naïve misconceptions along valid scientific beliefs in development and their robustness in the face of counterevidence. According to this model, and in contrast to the classical model proposed by Posner et al., 1982, an appropriate way to sequence teaching is by presenting the desired conception first, so that it is available at the time when children experience the evidence that challenges their existing conception. This sequence could then be followed by discussion and analysis of the counterevidence, to ensure that the new theory becomes prevalent in relation to children's naïve theory.

There is a need for studies to further examine the relation between children's naïve theories and their ability to learn from different types of evidence. What other conditions might facilitate children's belief revision in response to counterevidence when robust misconceptions are present? It is likely that across different knowledge domains, children's belief revision will be influenced by how robust their naïve theory is, the strength of the evidence (i.e., how varied, reliable, and compelling it is), and as this research has shown, the availability of an alternative theory. Crucially, this research indicates that in cases where robust misconceptions are present, having access to an alternative conception as children observe and interact with anomalous evidence leads to increased reliance on counterintuitive beliefs in their predictions and explanations of new phenomena. Further research should explore the applicability of our current findings to different scientific concepts.

Finding ways to get children to consider alternative theories in developmentally appropriate ways and integrating them with exploratory activities is an important endeavor for future research on early science education. In the current research, the alternative theory was delivered through a picture book. Picture books are an excellent medium to present children with complex conceptual information in a naturalistic and developmentally appropriate manner. Both this study and related research have shown that children can acquire conceptually rich explanations from picture books, whether the format is narrative or expository (Emmons, Smith, & Kelemen, 2016; Ganea et al., 2011; Gripshover & Markman, 2013; Kelemen et al., 2014; Venkadasalam & Ganea, 2018). Few studies have used comparable texts to examine the effect of genre on children's learning of science concepts (Donovan & Smolkin, 2002; Duke & Billman, 2009), and proposals have been made for increasing the use of both narratives (Avraamidou & Osborne, 2009) and expository text (Donovan & Smolkin, 2002; Duke, 2000; Mantzicopoulos & Patrick, 2011) to teach science. Together with prior evidence (Ganea et al., 2011; Venkadasalam & Ganea, 2018), the current findings show that young children can learn equally well from both genres.

Future research could also work out whether the delivery method of counterintuitive beliefs and anomalous evidence has an impact on young children's naive beliefs. Previous research with older children that has examined the effectiveness of hands-on versus hands-off teaching techniques for belief-revision indicates an advantage of hands-off approaches. Reading about an experiment and its results and conclusions was more effective in promoting belief revision than doing the experiment itself (Renken & Nunez, 2010) and this effect lasted after a 3-month delay. It may be possible that presenting children with anomalies in a hands-off approach may lead to better learning outcomes. In the current research, the anomalies were introduced through a hands-on activity whereas the alternative theory was introduced through a picture book. It may be possible that both would be equally effective if delivered through hands-off approaches (but see Masson et al., 2011; Renken & Nunez, 2013 for negative results).

Another promising direction for future research is to enhance children's interpretation of the anomalies as they observe them through verbal explanations. In ongoing research, we investigate whether for young children, pairing anomalous evidence with verbal explanations has a positive impact on belief revision and whether children's learning through anomalies and explanations lasts over extended periods of time.

More broadly the current findings support the view advocated by many educators and academics that science education should start early in development (Bowman, Donovan, & Burns, 2001; Duschl et al., 2007; Eshach & Fried, 2005; Gelman & Brenneman, 2004; Morgan, Farkas, Hillemeier, & Maczuga, 2016). As shown in this research, a learner's prior knowledge constrains the interpretation of new evidence and influences the extent to which children learn and are willing to revise their real-world beliefs (see also Bonawitz et al., 2012; Penner & Klahr, 1996). Given evidence that children's scientific knowledge at preschool is a good predictor of children's later science learning (Morgan et al., 2016) and that science misconceptions endure into adulthood (Coley & Tanner, 2012; Pine et al.., 2001; Potvin & Cyr, 2017; Shtulman, 2017; Shtulman & Valcarcel, 2012), finding ways to address misconceptions early in development in essential.

References

Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education*, *31*, 1683–1707. https://doi.org/10.1080/09500690802380695

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*, 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2001). *Eager to learn: Educating our preschoolers*. National Academy Press. https://doi.org/10.1038/nrrheum.2011.77

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (1998). Explanation in Scientists and Children. *Minds and Machines*, *8*, 119–136. https://doi.org/10.1023/A:1008242619231

Carey, S. (1999). Knowledge acquisition: Enrichment or conceptual change. In E. Margolis & S. Laurence (Eds.), *Concepts: core readings* (2nd ed., pp. 459–487). Cambridge, MA: MIT Press.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*, 1–49. https://doi.org/10.3102/00346543063001001

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*, 623–654.

https://doi.org/10.1002/(SICI)1098-2736(199808)35:6<623:AID-TEA3>3.0.CO;2-O

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175–218. https://doi.org/10.1002/sce.10001

Coley, J. D., & Tanner, K. D. (2012). Feature approaches to biology teaching and learning common origins of diverse misconceptions: Cognitive principles and the development of biology thinking. *Life Sciences Education*, *11*, 209–215. https://doi.org/10.1187/cbe.12-06-0074

Donovan, C. A., & Smolkin, L. B. (2002). Children's genre knowledge: An examination of K-5 students' performance on multiple tasks providing differing levels of scaffolding. *Reading Research Quarterly*, *37*, 428–465. https://doi.org/10.1598/RRQ.37.4.5

Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629–646). Abingdon, UK: Routledge.

Duke, N. (2000). 3.6 Minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*, 202–224. https://doi.org/10.1598/RRQ.35.2.1

Duke, N. K., & Billman, A. K. (2009). Informational text difficulty for beginning readers. In E. H. Hiebert & M. Sailors (Eds.), *Finding the right texts: What works for beginning and struggling readers*. New York, NY: Guilford Press.

Duschl, R. A., Richard, A., Schweingruber, H. A., Shouse, A. W.; National Research Council (U.S.). Committee on Science Learning, K. T. E. G., National Research Council (U.S.). Board on Science Education, & National Research Council (U.S.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Emmons, N., Smith, H., & Kelemen, D. (2016). Changing minds with the story of adaptation: Strategies for teaching young children about natural selection. *Early Education and Development*, *27*, 1205–1221. https://doi.org/10.1080/10409289.2016.1169823

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology*, *14*, 315–336. https://doi.org/10.1007/s10956-005-7198-9

Furtak, E. M., Shavelson, R. J., Shemwell, J. T., & Figueroa, M. (2012). To teach or not to teach through inquiry: Is that the question? In S. M. Carver & J. Shrager (Eds.), *The journey from child to scientist: Integrating cognitive development and the education sciences* (pp. 227–244). Washington, DC: American Psychological Association.

Ganea, P. A., Ma, L., & DeLoache, J. S. (2011). Young children's learning and transfer of biological information from picture books to real animals. *Child Development*, *82*, 1421–1433. https://doi.org/10.1111/j.1467-8624.2011.01612.x

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*, 150–158. https://doi.org/10.1016/j.ecresq.2004.01.009

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*, 1623–1627. https://doi.org/10.1126/science.1223416

Gopnik, A., & Meltzoff, A. N. (1997). Words, thoughts, and theories. *Trends in Cognitive Sciences*, *1*, 122. Retrieved from https://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613(97)89059-6.pdf

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

Greenhoot, A. F., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology*, *18*, 203–221. https://doi.org/10.1002/acp.959

Gripshover, S., & Markman, E. (2013). Teaching young children a theory of nutrition: Conceptual change and the potential for increased vegetable consumption. *Psychological Science*, *24*, 1541–1553. https://doi.org/10.1177/0956797612474827

Hannust, T., & Kikas, E. (2010). Young children's acquisition of knowledge about the Earth: A longitudinal study. *Journal of Experimental Child Psychology*, *107*, 164–180. https://doi.org/10.1016/j.jecp.2010.04.002

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology*, *98*, 307–326. https://doi.org/10.1037/0022-0663.98.2.307

Kallery, M. (2015). Science in early years education: introducing floating and sinking as a property of matter. *International Journal of Early Years Education*, *23*, 31–53. https://doi.org/10.1080/09669760.2014.999646

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*, 195–212. https://doi.org/10.1016/0010-0277(74)90008-0

Kelemen, D. (2019). The magic of mechanism: Explanation-based instruction on counterintuitive concepts in early childhood. *Perspectives on Psychological Science*, *14*, 510–522. https://doi.org/10.1177/1745691619827011

Kelemen, D., Emmons, N. A., Seston Schillaci, R., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychological Science*, *25*, 893–902. https://doi.org/10.1177/0956797613516009

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–48. https://doi.org/10.1207/s15516709cog1201_1

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*,

15, 661–667. https://doi.org/10.1111/j.0956-7976.2004.00737.x
Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: The MIT Press. Retrieved from https://psycnet.apa.org/record/1996-98594-000
Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689. https://doi.org/10.1037/0033-295X.96.4.674
Kuhn D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77* (3), 319–337. https://doi.org/10.1002/sce.3730770306
Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, *42*, 109–113. https://doi.org/10.1080/00461520701263376
Larsen N. E., Venkadasalam V. P., Ganea P. A. (2020). Prompting Children's Belief Revision About Balance Through Primary and Secondary Sources of Evidence. *Frontiers in Psychology*, *11*, https://doi.org/10.3389/fpsyg.2020.01503
Lehrer, R., Schauble, L., Strom, D., & Pligge, M. (2001). Similarity of form and substance: Modeling material kind. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: 25 years of progress* (pp. 39–74). East Sussex, UK: Psychology Press.
Mantzicopoulos, P., & Patrick, H. (2011). Reading picture books and learning science: Engaging young children with informational text. *Theory Into Practice*, *50*, 269–276. https://doi.org/10.1080/00405841.2011.607372
Masson, M. E. J., Bub, D. N., & Lalonde, C. E. (2011). Video-game training and naïve reasoning about object motion. *Applied Cognitive Psychology*, *25*, 166–173. https://doi.org/10.1002/acp.1658
Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, *45*, 18–35. https://doi.org/10.3102/0013189X16633182
National Institutes of Health. (2015). *NIH Toolbox Picture Vocabulary Test*. Retrieved from http://www.nihtoolbox.org/
Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher-led instruction in preschool settings. *Early Education and Development*, *22*, 970–988. https://doi.org/10.1080/10409289.2010.507496
Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development*, *67*. https://doi.org/10.2307/1131748
Peterson, S. M., & French, L. (2008). Supporting young children's explanations through inquiry science in preschool. *Early Childhood Research Quarterly*, *23*, 395–408. https://doi.org/10.1016/j.ecresq.2008.01.003
Pine, K., Messer, D., & St. John, K. (2001). Children's misconceptions in primary science: A survey of teachers' views. *Research in Science & Technological Education*, *19*, 79–96. https://doi.org/10.1080/02635140120046240
Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 211–227. https://doi.org/10.1002/sce.3730660207
Potvin, P. (2013). Proposition for improving the classical models of conceptual change based on neuroeducational evidence: Conceptual prevalence. *Neuroeducation*, *1*, 16–43. https://doi.org/10.24046/neuroed.20130201.16
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. https://doi.org/10.1002/tea.21396
Rappolt-Schlichtmann, G., Tenenbaum, H. R., Koepke, M. F., & Fischer, K. W. (2007). Transient and robust knowledge: Contextual support and the dynamics of children's reasoning about density. *Mind, Brain, and Education*, *1*, 98–108.
Renken, M. D., & Nunez, N. (2010). Evidence for improved conclusion accuracy after reading about rather than conducting a belief-inconsistent simple physics experiment. *Applied Cognitive Psychology*, *24*, 792–811. https://doi.org/10.1002/acp.1587
Renken, M. D., & Nunez, N. (2013). Computer simulations and clear observations do not guarantee conceptual understanding. *Learning and Instruction*, *23*, 10–23. https://doi.org/10.1016/j.learninstruc.2012.08.006
Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents inferences about hidden causes by 10-and 12-month-old infants. *Psychological Science*, *16*, 995–1001. https://doi.org/10.1111/j.1467-9280.2005.01649.x
Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*, 102–119. https://doi.org/10.1037/0012-1649.32.1.102.
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*, 211–223. https://doi.org/10.1016/j.cognition.2008.07.017
Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*, 427–442. https://doi.org/10.1111/j.1467-8624.2006.00880.x
Selley, N. (1993). Why do things float? A study of the place for alternative models in school science. *The School Science Review*, *74*, 55–61.
Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York, NY: Basic Books.
Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215. https://doi.org/10.1016/j.cognition.2012.04.005

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, *21*, 177–237. https://doi.org/10.1016/0010-0277(85)90025-3

van Schijndel, T. J. P., Visser, I., van Bers, B. M. C. W., & Raijmakers, M. E. J. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, *131*, 104–119. https://doi.org/10.1016/j.jecp.2014.11.008

Strouse G. A., & Ganea P. A. (2016). Are Prompts Provided by Electronic Books as Effective for Teaching Preschoolers a Biological Concept as Those Provided by Adults? *Early Education and Development*, *27*, (8), 1190–1204. https://doi.org/10.1080/10409289.2016.1210457

Venkadasalam, V. P., & Ganea, P. A. (2018). Do objects of different weight fall at the same time? Updating naive beliefs about free-falling objects from fictional and informational books in young children. *Journal of Cognition and Development*, *19*, 165–181. https://doi.org/10.1080/15248372.2018.1436058

Vosniadou, S. (2013). Conceptual change in learning and instruction. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 23–42). New York, NY: Routledge. https://doi.org/10.4324/9780203154472.ch1

Wilkening, F., & Cacchione, T. (2010). Children's intuitive physics. *The Wiley-Blackwell Handbook of childhood cognitive development* (pp. 473–496). Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch18

Yue, Y., Tomita, M. K., & Shavelson, R. J. (2008). Diagnosing and dealing with student misconceptions: Floating and sinking. *Science Scope*, *34*, 34–39.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Supplementary materials.