Journal of Experimental Child Psychology 203 (2021) 105047

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Investigating the effectiveness of fantasy stories for teaching scientific principles

Emily J. Hopkins [a,*], Deena Skolnick Weisberg [b]

[a] Department of Psychology, University of Scranton, Scranton, PA 18510, USA
[b] Department of Psychological and Brain Sciences, Villanova University, Villanova, PA 19085, USA

ARTICLE INFO

Keywords:
Learning
Transfer
Fantasy
Stories
Scientific reasoning

ABSTRACT

Educational media often contain fantastical information. Although some prior research suggests that this information interferes with children's learning, other work shows that fantasy benefits learning under certain circumstances. To investigate this issue and to clarify how different types of fantastical events might affect children's learning, we presented preschoolers (N = 99 in Study 1; N = 101 in Study 2) with stories that contained events that violated real-world physical laws, violated real-world biological laws, or did not violate any real-world laws. Within each story, we embedded two pieces of educational information, one each from the domains of biology and physics, to test (a) whether there are benefits of fantastical information on learning and (b) whether such benefits are domain specific. Across both studies, we found that children generally learned both types of information best from the story with physical violations, suggesting that such events can bolster children's learning.

Introduction

Much of children's educational media embeds their educational content inside fictional narratives. Television shows and books present entertaining stories about fictional characters while simultaneously attempting to teach real-world information such as science content, new vocabulary, and moral

* Corresponding author.
E-mail address: emily.hopkins@scranton.edu (E.J. Hopkins).

https://doi.org/10.1016/j.jecp.2020.105047
0022-0965/

lessons. A challenge for children is to separate the real information, which should be added to their knowledge of the world, from the fictional information, which is true only inside the world of the story. This challenge has been referred to as the "reader's dilemma" (Gerrig & Prentice, 1991), and it is especially acute for young children because their knowledge of the real world is still developing (see Hopkins & Weisberg, 2017, for a review). This dilemma is further complicated by the fact that much of children's media is not only fictional but also fantastical. That is, children's media contain not only story elements that are not true but also story elements that *cannot* be true. Fantasy content in children's media is extremely common (Goldstein & Alperson, 2020) and includes elements such as anthropomorphic animals, magical objects, and portals to other worlds. How does the presence of such fantastical content affect children's ability to learn from educational media?

Unfortunately, prior research aimed at answering this question is highly contradictory. On the one hand, there are many reasons to think that fantasy content in stories interferes with children's ability to learn from them. The fantasy may be distracting; cognitive resources that might otherwise be available to process the new educational material must instead be devoted to understanding the fantastical events (Fisch, 2000; Lillard, Drell, Richey, Boguszewski, & Smith, 2015). In addition, it is well known that similarity between the teaching context and the context to which children must apply new information facilitates transfer (e.g., Holyoak, Junn, & Billman, 1984). Fantastical stories are highly dissimilar from reality, and this dissimilarity may impede learning.

Consistent with this view, a number of studies have found that children do not learn as well from fantastical stories as they do from realistic stories (Ganea, Canfield, Simons-Ghafari, & Chou, 2014; Richert, Shawber, Hoffman, & Taylor, 2009; Richert & Smith, 2011; Walker, Gopnik, & Ganea, 2015). In these studies, children were taught some novel information such as new vocabulary words, properties of novel animals, and a new way to solve a problem. The information was embedded in a story that either was completely realistic or contained events that could not happen in reality (e.g., stickers raining from the sky). When children were tested afterward for their learning of this information and/or their ability to transfer it to an analogous context, they performed worse if they had heard the fantastical version of the story. Similarly, some elements of humor, such as exaggerated distortions, can interfere with children's perceptions of the educational elements in a story and make it harder for children to learn (Weaver, Zillmann, & Bryant, 1988; Zillmann et al., 1984).

However, there is a growing body of research suggesting that, in certain cases, there may be a *fantasy advantage* (Weisberg, 2016): Children may reason and learn better when fantasy is involved. For example, children are better able to solve logical syllogisms when asked to imagine that the events described in the syllogisms take place on a distant planet (Dias & Harris, 1988, 1990). They are better at reasoning about theory of mind when the characters are fictional (e.g., the Lion King) than when they are realistic (e.g., a puppy) (Lillard & Sobel, 1999). Children also learn the properties of novel objects better after those objects appear to violate physical properties of solidity and support (Stahl & Feigenson, 2015, 2017).

Most relevant to the current work, and in direct opposition to the work reviewed above, a few studies have found that children learned as well or better from fantastical stories as compared with realistic ones. For example, children can learn new facts about reality at least as well from a fantastical story as from a realistic story, if not better than the latter (Weisberg & Hopkins, 2020). In addition, a study of preschool children's learning of new vocabulary found better performance on a test of language production when these children had heard the words within fantastical stories (Weisberg et al., 2015). Finally, a study of children's learning of new problem solutions found that children transferred these solutions better from stories that contained events that violated real-world laws compared with stories that contained no such violations (Hopkins & Lillard, 2020). This was the case only when the law-violating events occurred earlier in the story than the target educational material, implying that the fantastical events played an important role in helping children to learn. One possible explanation of this effect is that fantastical events are surprising and violate children's expectations about how the world works. Because of this, the presence of such events may prompt children to increase their attention to the story, seek explanations for the violation, and look at subsequent story events with a more critical eye. In other words, this type of fantasy may create a positive *mise en place* for learning—putting children in the mindset that something unusual is happening that must be carefully attended to and processed (Stahl & Feigenson, 2015, 2019; Weisberg, Hirsh-Pasek, Golinkoff, & McCandliss, 2014).

Given these sets of inconsistent findings, do children learn better from realistic stories or from fantastical ones? This question is difficult to answer because the stories used in all of the prior work tend to include many different types of fantasy events (e.g., taking place on a faraway planet, having anthropomorphic animals). Thus, the goal of the current study was to investigate more deeply whether realistic or fantastical stories are more beneficial to children's learning and to begin to probe the circumstances under which one type of story might be better. In particular, unlike previous work, we focused on two distinct categories of impossible events: those that violate physical laws (e.g., gravity, solidity, continuity) and those that violate biological laws (e.g., aging, animacy). Simple concepts of physics are understood from a very young age; even infants know that one solid object cannot pass through another one, that an object will fall unless supported by another one, and that an object cannot disappear in one place and reappear in another place (see Baillargeon, 2004; Spelke, Breinlinger, Macomber, & Jacobson, 1992). By the time children reach preschool, they are confident that events such as a person walking through a wall and an object shrinking to a microscopic size are impossible.

But physical violations are not the only type of impossible events. There are also laws of biology that cannot be violated in real life. For example, aging progresses in only one direction and cannot be stopped or reversed. By preschool, children also understand these concepts; most children can correctly distinguish possible biological events from impossible ones (Cook & Sobel, 2011). However, this knowledge develops later than children's knowledge of physics, and preschoolers may be somewhat less confident in their predictions of what can and cannot happen in the biological domain (Carey, 1985; Inagaki & Hatano, 2006; Sobel & Weisberg, 2014). Therefore, violations of biological laws might not have the same impact on learning as violations of physical laws. On the other hand, Sobel (2006) found that both stories containing physical violations (e.g., walking through walls) and stories containing social/functional violations (e.g., wearing your shoes on your hands) led to better performance on a task assessing understanding of pretense; based on this, we may expect any type of violation to positively affect learning.

Although most previous studies have included only one type of violation or the other, or have used stories that mixed these types, the current studies presented a story containing only physically impossible events, a story containing only biologically impossible events, and a realistic story containing no impossible events. This more careful focus on separating these two categories of impossible events could help us to determine both whether fantasy content can benefit children's learning and which type of fantasy content might be most beneficial.

A second innovation of the current work was to include educational material that aligns with the type of impossible events presented in the stories. Specifically, we aimed to teach children about two different scientific principles drawn from the same domains as our impossible story events: biology and physics. This design can allow us to determine whether a match (or mismatch) between the domain of the violation events in the stories and the domain of the educational material might be beneficial to children's learning.

For the domain of biology, we aimed to teach children about biological inheritance, which is a principle in the biological domain that children have some knowledge of by the preschool period. They understand that animals have certain inborn traits characteristic of their species (Gelman & Wellman, 1991) and that offspring tend to resemble their biological parents (Springer & Keil, 1991; Williams & Smith, 2010). However, preschoolers do not yet understand that acquired traits cannot be inherited from a parent (Springer & Keil, 1989); this understanding begins to emerge a few years later. Our stories aimed to address this misconception. For the domain of physics, we aimed to teach children about how balance scales operate because children's understanding of balance is also developing around this age. Children as young as 2 or 3 years understand the role of weight in balance—that the heavier end of a balance scale will sink (Halford, Andrews, Dalton, Boag, & Zielinski, 2002). However, preschoolers were correct only half of the time on questions about how distance from the fulcrum would affect the balance of a scale. In contrast, 5- and 6-year-olds answered these questions correctly close to 90% of the time. Thus, preschoolers have incomplete knowledge about both of these topics, so they may be able use educational material in our storybooks to improve their understanding.

Study 1 employed a 3 (Story Type: realistic, physical violations, or biological violations) $\times$ 2 (Principle Type: inheritance or balance) design, where story type was a between-participants variable and principle type was a within-participants variable. That is, all children saw educational material about both inheritance and balance in their stories, but their stories contained only one type of violation (or no violations at all).

Based on prior work using stories with impossible events of this type, we predicted that children would learn better from stories containing impossible events than from a realistic story. However, we formed four competing hypotheses (H1–H4) about how the type of violation and the type of educational content might interact:

H1: Any type of impossible event will heighten overall attention to the story and lead to better learning of any type of information. That is, for both the inheritance principle and balance principle, learning will be better in both the physical violations and biological violations stories than in the no violations story.

H2: Violations will have an impact only within domain. For example, observing physically impossible events will improve how children learn subsequent physics information but not biology information. Thus, learning of the inheritance principle will be best when the story contains biologically impossible events, and learning of the balance principle will be best when the story contains physically impossible events.

H3: Only physically impossible events will affect learning; learning of both principles will be better when the story contains physically impossible events compared with either of the other two stories.

H4: Only biologically impossible events will affect learning; learning of both principles will be better when the story contains biologically impossible events compared with either of the other two stories.

Study 1

Method

Participants

This study was conducted in a large city in the eastern United States. The final sample contained 99 children (52 girls): 34 in the *no violations* condition (mean age = 57.6 months), 32 in the *physical violations* condition (mean age = 58.3 months), and 33 in the *biological violations* condition (mean age = 57.9 months). The mean age of participants in the whole sample was 57.9 months (range = 42–71 months), and this did not differ significantly by condition (all p values > .61). Race/ethnicity was identified via parent report; the majority of participants (64%) were White, 5% were Asian, 6% were Black or African American, 8% identified as more than one race, and 17% did not report race/ethnicity. An additional 24 children were run but not included in the final sample. Of these children, 5 were uncooperative or appeared to have difficulty in understanding the questions, 2 were younger than our target age group, and 17 were displayed atypical response patterns (described in Results).

Materials

There were three main components to this study. The *story* presented children with information about two scientific principles, and the *inheritance problem* and *balance problem* assessed children's learning of these two principles.

Story. In all three conditions, children were read an illustrated storybook about a day in the life of a character named Jamie (pronouns were gender-matched to the participant). The book had 15 pages that depicted the same essential sequence of events: Jamie woke up, let his/her dog outside, washed dishes, watered plants, fed his/her hamster, went to the park to play in the sandbox, played on a seesaw, found a bird's nest, went home for dinner, and went to bed.

The seesaw and bird's nest events were used to teach the target scientific information. The two-page seesaw event (Fig. 1A) presented children with information about balance. The first page showed Jamie on one end of the seesaw and two children on the other end; the text explained that the side with more children on it would go down to the ground because it was heavier. The second page showed Jamie at the very end of one side of the seesaw and one child closer to the fulcrum on the other side. The text explained that Jamie's side went down to the ground because he/she was closer

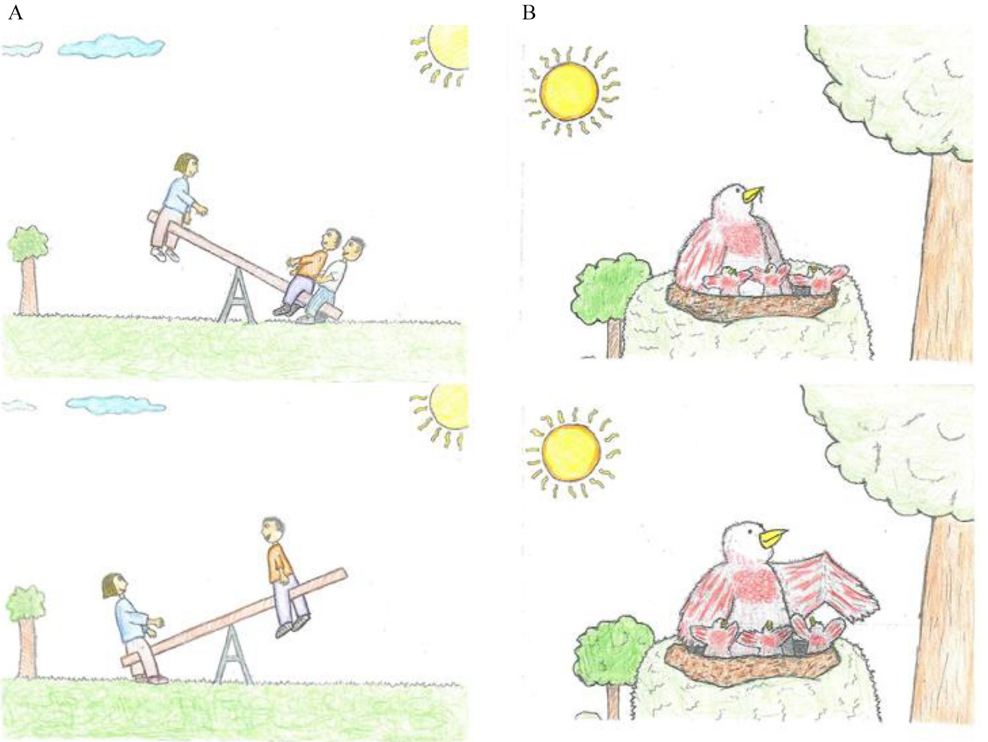


Fig. 1. Educational events in the storybook: (A) the seesaw event; (B) the bird's nest event.

to the edge. This was intended to teach children that both weight and distance are important for balance.

The two-page bird's nest event (Fig. 1B) presented children with information about inheritance. The first page showed a mother bird and three baby birds in a nest. The text explained that the mother bird was born with red feathers, and so the babies were also born with red feathers. On the second page, the mother bird was shown extending a bent wing. The text explained that she hurt her wing in an accident, but the baby birds have straight wings because they were not in an accident like the mother. This was intended to teach children that inborn traits can be passed on to offspring, but acquired traits cannot. These two teaching events were exactly the same in all three conditions. The order of these two events within the story was counterbalanced across participants.

Three other events in the story varied across conditions: letting the dog outside, watering plants, and feeding the hamster. In the no violations condition, the dog came back inside covered in mud, Jamie smelled the flowers after watering them, and the hamster ran on its wheel. In the physical violations condition, children heard versions of these events that violated physical laws: The dog shrunk down to a very small size, the flowers floated in the air, and the hamster walked through the solid wall of its tank. In the biological violations condition, children heard versions of these events that violated biological laws: The dog turned into a puppy, the flowers had faces and said "Good morning," and the hamster spontaneously changed the color of its own fur.[1] These sets of events were designed to be as similar to each other as possible while still conforming to their condition.

[1] Note that although the color of hamsters' fur can change in real life (e.g., as a response to stress or to a change in seasons), the depiction of this change in the stories was spontaneous and complete and hence entirely unrealistic.

Balance problem. This problem used illustrations of a simple balance scale with three pegs on each side of a central fulcrum (Fig. 2A). The scale was shown with wooden blocks placed under each side so that it was horizontal to the ground. One block had the outline of a star on it, and the other block had the outline of a crescent moon. The problem was introduced using a picture of the scale with no weights on the pegs. On test items, different numbers of green weights were shown on the scale's pegs. These test items varied on three different variables that were counterbalanced to create 16 possible trial orders (see Appendix A). The first variable was whether the problem focused on the dimension of weight or distance. On weight trials, there were more discs on one side than on the other side, but the discs on each side were equidistant from the center. On distance trials, there was the same number of discs on each side, but the discs on one side were closer to the center than the other discs. Weight and distance trials always occurred in an ABBA order (i.e., reverse counterbalancing). The second variable was the difference in the number of discs displayed (for weight trials) or the difference in the distance from the center (for distance trials). Weight trials depicted either 1 disc versus 2 discs or 1 disc versus 3 discs. Distance trials depicted weight on the first peg versus second peg or on the first peg versus third peg. The other variable concerned which side of the scale would go down—the side with the star or the side with the moon (i.e., the correct answer). Children always received equal numbers of trials where the star and moon were correct; across participants, this was distributed evenly across weight/distance trials and across 1 versus 2/1 versus 3 trials.

Inheritance problem. Each trial of this problem presented children with pictures of two animals, side by side, each with a different version of a trait (Fig. 2B). As with the balance problem, test items varied on three variables that were counterbalanced to create 16 possible trial orders (see Appendix A). The first variable was the type of trait. Trials involved either inborn traits (e.g., eye color, fur color) or acquired traits (e.g., a lost tail, an injured ear); these were always presented in an ABBA order. The second variable was whether the traits were illustrated using photographs of cats or of bunnies. The third variable concerned whether the correct response was depicted on the right side or left side of the card. Children always received equal numbers of trials where left and right were correct; across participants, this was distributed evenly across inborn/acquired and cat/bunny trials.

Procedure

Participants were tested in a quiet area of their preschool/day-care center ($n = 75$), in their home ($n = 2$), in the lab ($n = 2$), or at a local children's museum ($n = 20$). Regardless of testing location, the procedure consisted of three phases: a pretest, story exposure, and a posttest.

Experimenter 1 administered the pretest. Participants completed four trials each (blocked) of the balance and inheritance problems. The order of the two problem blocks matched the order that the inheritance and balance events were presented in the story and was counterbalanced across participants. On balance problem trials, the experimenter first showed the empty balance scale picture and explained that when weights were put on the pegs and the blocks underneath were removed, the side of the scale where the moon was would go down to the ground, the side where the star was would go down, or both sides would stay up. The experimenter then administered four test trials in 1 of the 16 predetermined orders (randomly assigned). On each trial, the experimenter pointed to the stack of discs on each side and asked children, "When we take the blocks away, what will happen? Will the side with the moon go down, the side with the star go down, or will both sides stay up?" Children were given a score of 0 to 4 indicating how many trials they answered correctly.

For the inheritance problem, the experimenter administered four test trials in 1 of the 16 predetermined orders (randomly assigned). On each trial, she showed the children a card containing two animal pictures. She explained that the trait depicted can vary (e.g., "Some bunnies have black and white fur, and some have all black fur"; "Some cats have ears that stick up, and some have ears that flop down"). She then pointed to one of the animals and explained that the animal's trait was inborn ("This bunny was born with all black fur") or that the trait was acquired ("This cat hurt her ear in an accident and now it flops down"). The target question was which version of the trait the babies of that animal would display ("When she has babies, will they have black and white fur or all black fur?"; "When she has babies, will they have ears that stick up or ears that flop down?"). Children were given a score of 0 to 4 indicating how many trials they answered correctly.

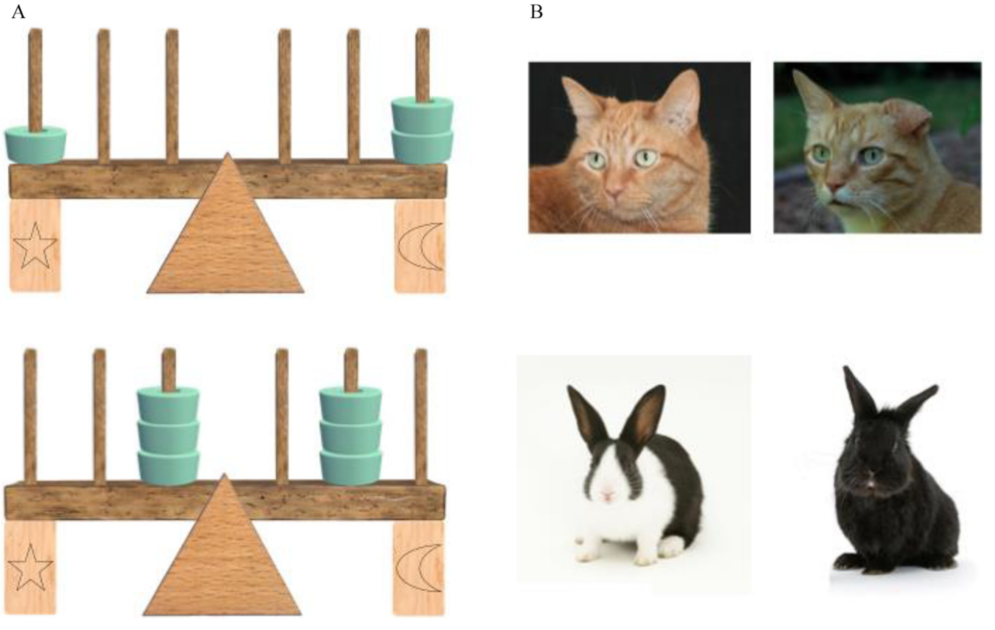


Fig. 2. Example stimuli for the balance problem (A) and the inheritance problem (B).

After the pretest, Experimenter 1 told children that she needed to get the next part of the game ready and introduced Experimenter 2. Experimenter 1 left the room while the story was being read so that she was blind to the condition that children were in. Experimenter 2 then read children their assigned story. During the story reading, Experimenter 2 did not introduce any extratextual information or direct children's attention to any particular aspect of the story. This experimenter was also instructed to respond positively but neutrally to children's questions or bids for attention in order to keep the presentation of the story as similar as possible across participants.

After the story, Experimenter 1 returned to administer the posttest. The procedure for the posttest was identical to that for the pretest except that different versions of the test items were used. The order of the balance and inheritance problems matched the order used at pretest. Children again received a score of 0 to 4 for each problem, indicating how many trials they answered correctly.

After the posttest, Experimenter 1 administered three additional tasks in a fixed order: free recall for the story, a set of directed memory questions about the story, and a set of reality status questions. For the free recall task, the experimenter asked children to tell her what happened in the story; the experimenter used generic prompts (e.g., "What else happened?") until children did not provide any more details. Responses were coded for the number of total story details correctly recalled by children, whether children included details of the seesaw or bird's nest event, and whether children included any details of the three violation events. Next, the experimenter asked eight questions about events from the story (e.g., "What did Jamie's mom make him/her for dinner?"); children's responses were scored as correct or incorrect, and children were given a total score from 0 to 8. Finally, the experimenter asked whether the nine events that varied by condition (dog, plant, and hamster events from each of the three conditions) could happen in real life. All children, regardless of condition, were asked about the same set of events, which included the entirely realistic version of the event from the no violations story (e.g., hamster running on a wheel) and the two impossible versions of the same event from the two fantastical stories (e.g., hamster walking through a wall; hamster changing the color of its fur). Again, children were given a score from 0 to 9, indicating the number of correct responses.

Results

Exclusions

As mentioned above, 17 children were excluded from the sample for atypical response patterns. Of these children, 14 appeared to misunderstand the balance problem. They often responded that "both sides will go up," a physical impossibility suggesting that they were confused about how the balance scale worked. When they did respond that one side or the other would go down, they always chose the incorrect side, suggesting that they thought the heavier/farther side would go up rather than down. That is, these children did not simply respond poorly; other children who responded at floor were maintained in our final sample. While these children may have simply been responding incorrectly, we believe it is more likely that they did not correctly interpret the balance scale stimuli. To conservatively include only children who understood the parameters of the balance problem, these children were removed from the sample. These children were on average about 2.5 months younger (55.1 months) than children who did not show this inverted pattern (57.7 months), $t(160) = 2.03$, $p = .044$.

Because we were interested in learning (i.e., improvement from pretest to posttest), we also excluded 3 children who showed ceiling performance (3 or 4 items correct) on both problems at pretest. There were also 9 children who showed ceiling performance for the balance problem (but not the inheritance problem) at pretest and 31 children who showed ceiling performance for the inheritance problem (but not the balance problem) at pretest. For these children, we dropped their data from both pretest and posttest only for the problem where they showed ceiling performance. The number of children whose data were included for each problem in each condition is shown in Table 1. Children who were at ceiling on one problem ($n = 40$, $M_{age} = 57.8$ months) were not significantly different in age from children who were not at ceiling on either problem ($n = 59$, $M_{age} = 58.0$ months), $t(97) = 0.18$, $p = .859$.

We first checked whether the two problems were of comparable difficulty for participants. We found that the inheritance problem was in fact easier than the balance problem (Table 2). Even after dropping those at ceiling at pretest, scores on the inheritance problem were higher than scores on the balance problem at both pretest [paired $t(58) = -6.99$, $p < .001$] and posttest [paired $t(58) = -4.52$, $p < .001$]. However, when collapsing across conditions, there was no significant difference in learning (as measured by difference scores subtracting pretest score from posttest score) between the physical and biological problems [paired $t(58) = 0.45$, $p = .653$].

Effects of story type

To investigate learning in each condition, we first conducted a series of paired *t* tests comparing pretest and posttest scores for each problem (i.e., inheritance or balance) in each condition (i.e., no violations story, biological violations story, or physical violations story) (see Fig. 3). For the inheritance problem, scores were significantly higher at posttest than at pretest in the physical violations condition [paired $t(23) = 4.08$, $p < .001$, $d = 0.83$] and in the biological violations condition [paired $t(20) = 2.83$, $p = .010$, $d = 0.62$]. For the balance problem, scores were significantly higher at posttest than at pretest only in the no violations condition [paired $t(31) = 2.83$, $p = .008$, $d = 0.50$].

To determine whether the problems or conditions differed significantly from each other, we conducted a mixed-effects regression (including random intercepts by participant) predicting posttest scores (converted to *z* scores for ease of interpretation) from children's age, pretest scores, problem type (inheritance or balance), and condition (no violations, biological violations, or physical violations). Mixed-effects regression was used because the majority of the participants were included in the model twice—once for the biological problem and once for the physical problem. The random effects structure adjusts for this nonindependence. Condition was simple-effects coded to compare each level with the reference level (the no violations condition). The results of this regression are shown in Table 3.

There were significant main effects of pretest scores and of problem type (scores on the balance problem were lower than those on the inheritance problem) as well as a significant interaction between problem type and condition. To understand the interaction, we conducted separate regressions for each problem type, predicting posttest scores from age, pretest scores, and condition

Table 1
Number of data points included for each problem by condition in Study 1.

Problem	No violations condition (n = 34)	Physical violations condition (n = 32)	Biological violations condition (n = 33)
Inheritance	23	24	21
Balance	32	26	32

Table 2
Mean pretest and posttest scores by problem in Study 1.

Problem	Pretest	Posttest
Inheritance	1.65 (0.57; 0–2)	2.19 (0.95; 0–4)
Balance	1.01 (0.84; 0–2)	1.29 (1.21; 0–4)

Note. Standard deviations and ranges of observed scores are shown in parentheses.

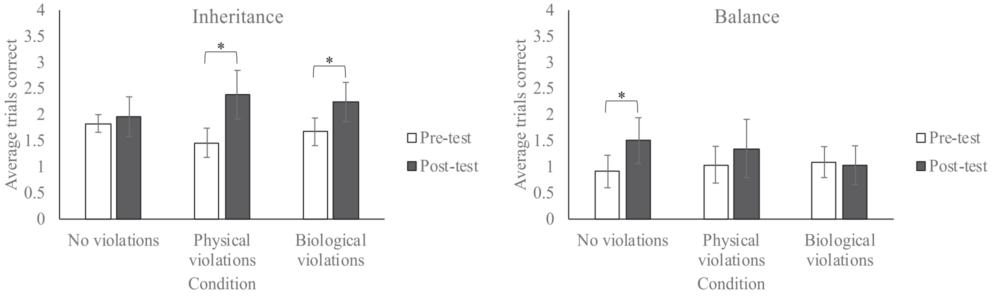


Fig. 3. Pretest and posttest scores by problem type and condition in Study 1. Asterisks indicate where posttest scores were significantly higher than pretest scores (paired t test, p < .05). Error bars represent 95% confidence intervals.

(Table 4). Mixed-effects modeling was not necessary here because the problems were analyzed separately, and thus each child was included only once in each regression.

For the inheritance problem, posttest scores (controlling for pretest scores) were significantly higher in the physical violations condition than in the no violations condition. Children in the physical violations condition answered 2.38 questions correct on average at posttest; this was 0.49 standard deviations above children in the no violations condition, who answered 1.96 questions correct (a medium-sized effect). The biological violations (M = 2.24) and no violations conditions were not significantly different from each other.

For the balance problem, posttest scores were marginally lower in the biological violations condition than in the no violations condition. Children in the biological violations condition answered 1.03 questions correct on average at posttest; this was 0.47 standard deviations below children in the no violations condition, who answered 1.50 questions correct (a medium-sized effect). The physical violations (M = 1.35) and no violations conditions were not significantly different from each other.

In addition, we explored whether pretest scores moderated the effect of condition on posttest scores. Unlike the rest of our analyses, these tests used the entire sample of children to examine how all levels of prior knowledge might differentially affect learning from the different stories. For the balance problem, we found no effect; children's learning did not differ by condition depending on their level of prior knowledge. For the inheritance problem, there was a marginal interaction; children who scored low at pretest benefitted slightly more from hearing the physical violations story (as compared with the no violations condition) than children who scored higher at pretest ($\beta = -.35$, p = .068).

Table 3
Mixed-effects regression for Study 1.

Predictor	*β*	*SE*	*t*	*p*	
Intercept	−1.75	0.69	−2.52	.013	*
Age (months)	.02	.01	1.87	.064	+
Pretest	.36	.10	3.64	<.001	***
Problem type	−.26	.08	−3.33	.001	**
Condition					
Physical	.16	.17	0.92	.359	
Biological	−.08	.17	−0.45	.657	
Problem Type × Condition					
Physical	−.35	.17	−2.03	.044	*
Biological	−.39	.17	−2.26	.025	*

+ $p < .10$.
* $p < .05$.
** $p < .01$.

Table 4
Analysis by problem in Study 1.

A. Inheritance					
Predictor	Coefficient	*SE*	*t*	*p*	
Intercept	−1.33	0.97	−1.38	.174	
Age (months)	.02	.02	1.36	.179	
Pretest	.30	.17	1.70	.094	+
Condition					
Physical	.49	.24	2.06	.044	*
Biological	.30	.24	1.28	.206	
B. Balance					
Predictor	Coefficient	*SE*	*t*	*p*	
Intercept	−2.03	1.02	−1.99	.049	*
Age (months)	.02	.02	1.26	.210	
Pretest	.38	.13	3.00	.004	**
Condition					
Physical	−.20	.25	−0.78	.435	
Biological	−.47	.24	−1.97	.052	+

+ $p < .10$.
* $p < .05$.
** $p < .01$.

Effects of memory and fantasy/reality understanding

We anticipated that children's learning might be related to their memory of the story (particularly for the educational material) or to their understanding of the reality status of the events in the story (based on prior work showing that children with better understanding of the realism of stories' educational material tend to learn better; e.g., Mares & Sivakumar, 2014). Investigating these variables can also help us to determine under what circumstances children are more likely to learn the target scientific information in the story. We used six different variables to investigate these issues: total events recalled during free recall (out of 20), seesaw details included in free recall (out of 2), bird's nest details included in free recall (out of 2), details about the violation events (or the parallel events in the no violation story) included in free recall (out of 3), number of memory questions answered correctly (out of 8), and number of reality status questions answered correctly (out of 9). Descriptive statistics for these variables are shown in Table 5. None of these variables differed significantly by condition (all p values between .270 and .920), and none was significantly correlated with difference scores on either the bal-

Table 5
Mean scores and percentages for memory and reality status measures in Study 1.

Variable	No violations	Physical violations	Biological violations
Free recall–total	3.94 (3.58), 19.7%	4.16 (3.32), 20.8%	3.82 (3.18), 19.1%
Free recall–seesaw	0.06 (0.24), 3.0%	0.09 (0.30), 4.5%	0.03 (0.17), 1.5%
Free recall–bird's nest	0.32 (0.68), 16.0%	0.19 (0.47), 9.5%	0.12 (0.33), 6.0%
Free recall–violations	0.59 (0.70), 19.7%	0.84 (0.88), 28.0%	0.79 (0.78), 26.3%
Memory questions	3.26 (1.35), 40.8%	3.50 (1.26), 43.8%	3.39 (1.43), 42.4%
Reality status questions	7.41 (1.60), 82.3%	7.34 (1.26), 81.6%	7.21 (1.43), 80.1%

Note. Standard deviations are given in parentheses.

ance problem (all *p* values between .287 and .796) or inheritance problem (all *p* values between .176 and .850).

Discussion

This study was designed to investigate whether and under what circumstances impossible events in stories can bolster children's learning of new scientific information from those stories. The results on the inheritance problem supported our overall hypothesis that impossible events positively affect learning; children who heard stories containing physical violations and children who heard stories containing biological violations answered significantly more questions correct at posttest than at pretest. However, performance only in the physical violations condition was significantly better than performance in the no violations condition, and there was a marginal indication that this was especially true for children with lower levels of prior knowledge about this problem. Thus, these results partially support either H1 or H3; both physical and biological violations led to significant improvement in response to the inheritance problem, but the effect was stronger for physically impossible events. These data run counter to H2 because the physical violations positively affected learning of a biological principle; thus, we did not observe only within-domain effects.

In contrast, the results on the balance problem did not support any of our hypotheses. Contrary to what we expected, the only condition where significant learning on the balance problem was observed was the no violations condition. This is more in line with previous research showing that children learn best from realistic stories (Ganea et al., 2014; Richert et al., 2009; Richert & Smith, 2011; Walker et al., 2015).

These conflicting results may be due to the difference in difficulty of the two problems. Many children struggled to understand the balance problem, and scores were significantly lower at both pretest and posttest compared with the inheritance problem. This presents an interesting possibility about the effect of fantasy on learning: The effect of fantasy may differ based on children's prior knowledge or familiarity with a topic. When a topic is difficult or unfamiliar (such as our balance problem), children may default to quarantining information from fantastical stories so as to avoid acquiring incorrect knowledge. But when they are dealing with a domain with which they feel more comfortable (such as our inheritance problem), fantasy may have a positive effect, cueing them to process the new information more deeply (Hopkins & Lillard, 2020; Stahl & Feigenson, 2019; Weisberg et al., 2014; Weisberg & Hopkins, 2020).

To further investigate this possibility and test our proposed hypotheses, Study 2 made slight modifications to the procedure. These modifications were primarily designed to make the two problems of more comparable difficulty. A brief introductory video was shown to participants before the pretest to help them understand how the balance scale operates, and a parallel video was also created for the inheritance problem to keep the two tasks as similar as possible. The study had two aims: first, to see whether the positive impact of fantasy on the inheritance problem would replicate and, second, to see whether children would show the same pattern on both problems if the balance problem was made less difficult for them.

Study 2

Method

Participants

This study was conducted in a large city in the eastern United States. The final sample consisted of 101 children (53 girls): 31 in the *no violations* condition (mean age = 55.9 months), 34 in the *physical violations* condition (mean age = 59.9 months), and 36 in the *biological violations* condition (mean age = 56.5 months). The mean age of participants in the whole sample was 57.5 months (range = 45–73). Children in the physical violations condition were significantly older on average than children in the biological violations condition [$t(68) = 2.25$, $p = .028$] and no violations condition [$t(62) = 2.56$, $p = .013$]; therefore, we controlled for age in analyses that compare conditions to account for this difference. Race/ethnicity was identified via parent report; the majority of participants (57%) were White, 2% were Asian, 33% were Black or African American, 2% identified as more than one race, and 6% did not report race/ethnicity. Participants were recruited from the same children's museum as in Study 1 ($n = 33$) and from a different set of local preschools ($n = 68$).

We also tested 17 children who were not included in the final sample. Of these children, 2 were uncooperative or appeared to have difficulty in understanding the questions. An additional 5 were dropped due to responding to the balance problem in the "inverted" pattern described in Study 1, and 9 were dropped for ceiling performance at pretest on both the inheritance and balance problems.

Materials

The same materials from Study 1 were used here. In addition, two short videos were shown prior to the pretest. These videos were designed to familiarize participants with the balance and inheritance test stimuli (the balance scale and animals, respectively), but no information about the processes of balance or inheritance was provided in these videos. In the balance video (22 s), the empty balance scale used in the pretest and posttest stimuli was shown with no supporting blocks under the two ends. The narrator explained that the scale could sometimes tip one way, sometimes tip the other way, and sometimes stay balanced. During this narration, the image of the scale was animated to show how it moves. Importantly, no weights were stacked on any of the pegs on the balance scale in this video, so children could not use it to learn about how weights affect the direction that the scale tips. In the inheritance video (24 s), the narrator explained that sometimes baby animals looked like their mothers while a small orange cat and large orange cat appeared on the screen. The narrator then explained that sometimes baby animals look different from their mothers while a small bunny and large bunny with different types of ears appeared on the screen. As in the balance video, this video only directed children's attention to an important characteristic of the stimuli (similarities and differences between baby animals and their mothers) without explaining anything about which traits might be expected to be similar or different or why.

Procedure

The procedure was identical to that from Study 1 except that the corresponding introductory video was shown before the first trial of each problem block in the pretest.

Results

Exclusions

We used the same criteria as in Study 1 to remove children who showed the inverted response pattern on the balance problem ($n = 5$) or who performed at ceiling at pretest ($n = 9$ were at ceiling on both problems). The number of children whose data were included for each problem in each condition is shown in Table 6. Children who showed the inverted pattern (56.4 months) were not significantly different in age from children who did not show the inverted pattern (57.5 months), $t(105) = 0.37$, $p = .706$. Children who were at ceiling on either problem ($n = 45$, $M = 58.0$ months) were not signif-

Table 6
Number of data points included for each problem by condition in Study 2.

Problem	No violations condition (n = 31)	Physical violations condition (n = 34)	Biological violations condition (n = 37)
Inheritance	26	24	27
Balance	21	30	29

icantly different in age from children who were not at ceiling on either problem (n = 57, M = 57.6 months), $t(99) = -0.30$, p = .768.

Unlike in Study 1, there was no difference in the difficulty of the two problems. After dropping those at ceiling on the pretest, scores on the inheritance problem were not significantly higher than scores on the balance problem at either pretest [paired $t(56) = -1.18$, p = .245] or posttest [paired $t(56) = -0.52$, p = .603] (Table 7). When collapsing across condition, there was no significant difference in learning (as measured by difference scores subtracting pretest score from posttest score) between the physical and biological problems [paired $t(56) = 0.25$, p = .803].

Effects of story type

To investigate learning in each condition, we first conducted a series of paired t tests comparing performance at pretest and posttest on each problem (Fig. 4). For both problems, the largest change from pretest to posttest was observed in the physical violations condition. However, this was significantly different only for the balance problem [paired $t(29) = 2.28$, p = .030, d = 0.42], not for the inheritance problem [paired $t(23) = 1.66$, p = .110, d = 0.34].

To determine whether the problems (inheritance and balance) or conditions (no violations story, biological violations story, and physical violations story) differed significantly from each other, we conducted a mixed-effects regression (including random intercepts by participant) predicting posttest scores from age, pretest scores, problem type, and condition. Condition was simple-effects coded to compare each level with the reference level (the no violations condition). The results of this regression are shown in Table 8.

There was a significant main effect of age and of pretest scores. Unlike Study 1, there was no significant main effect of condition and no significant interaction between condition and problem type.

Table 7
Mean pretest and posttest scores by problem in Study 2.

Problem	Pretest	Posttest
Inheritance	1.50 (0.70; 0–2)	1.69 (1.10; 0–4)
Balance	1.35 (0.82; 0–2)	1.60 (1.10; 0–4)

Note. Standard deviations and range of observed scores are shown in parentheses.

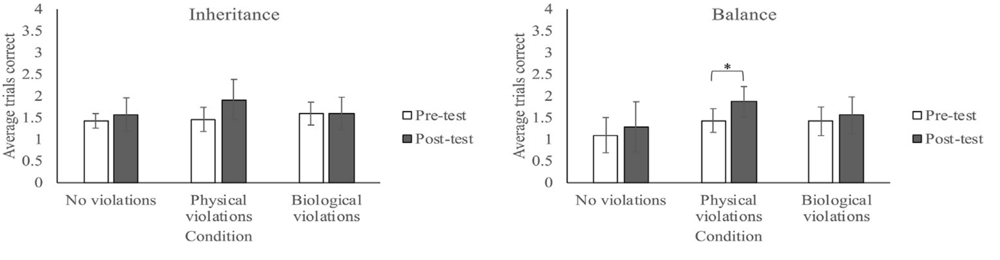


Fig. 4. Pretest and posttest scores by problem type and condition in Study 2. Asterisk indicates where posttest scores were significantly higher than pretest scores (paired t test, p < .05). Error bars represent 95% confidence intervals.

Table 8
Mixed-effects regression for Study 2.

Predictor	β	SE	t	p	
Intercept	−2.49	0.67	−3.70	<.001	***
Age (months)	.03	.01	2.66	.009	**
Pretest	.49	.10	5.04	<.001	***
Problem type	.01	.07	0.12	.907	
Condition					
Physical	.17	.19	0.89	.375	
Biological	−.06	.18	−0.35	.724	
Problem Type × Condition					
Physical	−.04	.18	−0.20	.841	
Biological	.02	.18	0.14	.892	

** $p < .01$.
*** $p < .001$.

As in Study 1, we also explored whether pretest scores moderated the effect of condition on posttest scores using the whole sample. Unlike in Study 1, there was no such effect for the inheritance problem. However, for the balance problem, children who scored higher at pretest benefitted more from the biological violations (as compared with the no violations condition) than children who scored lower at pretest ($\beta = .36$, $p = .042$).

Effects of memory and fantasy/reality understanding

We used the same six variables as in Study 1 to investigate whether memory and/or fantasy/reality understanding was related to children's learning: total events recalled during free recall (out of 20), seesaw details included in free recall (out of 2), bird's nest details included in free recall (out of 2), violation details included in free recall (out of 3), number of memory questions answered correctly (out of 8), and number of reality status questions answered correctly (out of 9). Descriptive statistics for these variables are shown in Table 9.

Only one variable differed significantly by condition. Of the three possible violations or corresponding events (the dog, plants, and hamster), children in the physical violations condition included an average of 0.77 in their free recall. This was significantly higher than the average in the whole sample collapsing across condition ($M = 0.57$, $p = .028$). In addition, two variables from the free recall measure were significantly correlated with difference scores on the balance problem when all three conditions were analyzed together ($\alpha = .008$ to adjust for multiple comparisons) (Fig. 5). Children who had larger learning scores included more total details from the story ($r = .39$, $p < .001$), and they included more details specifically about the violation events ($r = .58$, $p < .001$). When analyzed separately by condition, this latter result was significant in each of the two violations conditions (physical violations: $r = .37$, $p = .031$; biological violations: $r = .42$, $p = .011$), but not in the no violations condition ($r = .13$, $p = .481$). There were no significant correlations between difference scores on the inheritance problem and any of the memory or reality status variables.

Table 9
Mean scores for memory and reality status measures in Study 2.

Variable	No violations	Physical violations	Biological violations
Free recall–total	2.61 (2.85), 13.1%	3.29 (3.30), 16.5%	2.41 (2.54), 12.1%
Free recall–seesaw	0.03 (0.18), 1.5%	0.10 (0.30), 5.0%	0.08 (0.28), 4.0%
Free recall–bird's nest	0.06 (0.25), 3.0%	0.19 (0.54), 9.5%	0.11 (0.31), 5.5%
Free recall–violations	0.42 (0.56), 14.0%	0.77 (0.88)*, 25.7%	0.46 (0.61), 15.3%
Memory questions	2.77 (1.45), 34.6%	3.09 (1.36), 38.6%	2.61 (1.27), 32.6%
Reality status questions	7.00 (1.81), 77.8%	7.29 (1.61), 81.0%	6.67 (1.76), 74.1%

Note. Standard deviations are given in parentheses.
* Significantly different from the mean collapsed across conditions ($p < .05$).

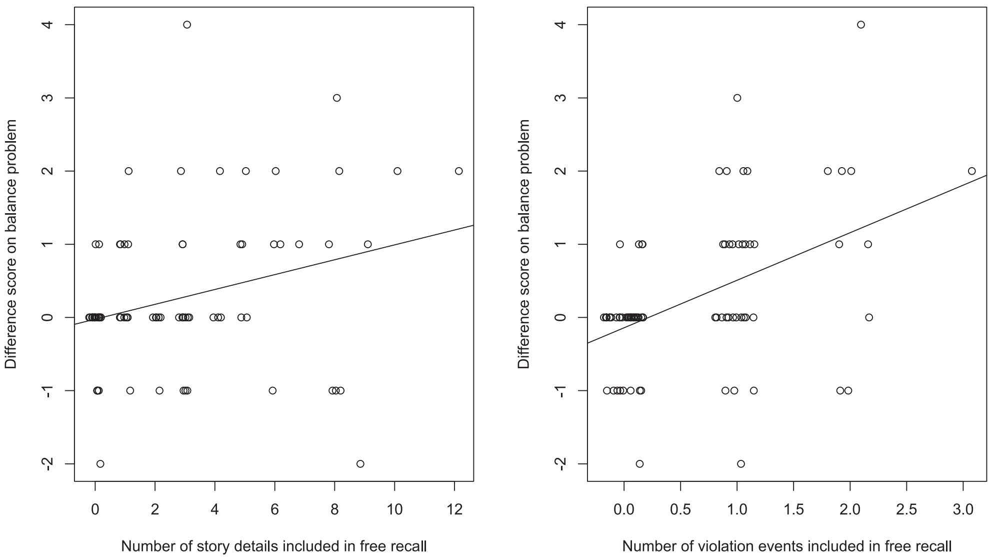


Fig. 5. Correlations between learning of the balance principle and free recall performance.

Discussion

The modification we made to the procedure in Study 2 was successful at reducing children's confusion with the balance problem; unlike Study 1, there was no significant difference in performance across the two tasks and hence no interaction effect. Thus, this replication of Study 1 allowed us to test our proposed hypotheses more directly. Although the effects in Study 2 were smaller than those in Study 1, they were in the same direction. For both problems, the most learning was observed in the physical violations condition, although the improvement from pretest to posttest was significant only for the balance problem and the main effect of condition was not significant in a regression analysis. If anything, these results are most consistent with H3; physically impossible events in particular may have a positive impact on learning, although there was some hint that the biological violations were particularly helpful for children who had more preexisting knowledge of the balance principle we were aiming to teach.

In addition, children in the physical violations condition were more likely to recall the violation events, and those who recalled more violation events were more likely to learn the balance principle. This result supports the idea that the physically impossible events are especially salient to children and promote deeper processing (and hence potentially more learning) of the story content.

General discussion

Our main goal in the current work was to examine the effect of fantastical events (i.e., events that violate real-world laws) on children's learning from storybooks that contain those events. Although the benefits of similarity for learning are well known, some prior work suggests that there is a fantasy advantage whereby fantastical contexts can be beneficial to children's learning and reasoning in certain circumstances (Hopkins & Lillard, 2020; Richert & Schlesinger, 2016; Stahl & Feigenson, 2015, 2017; Weisberg & Hopkins, 2020; Weisberg et al., 2015). Our findings support that conclusion. In both studies, children successfully learned new scientific information from a story that contained fantastical events; there was no consistent advantage to children's learning from a realistic story. These results imply that realistic stories are not automatically superior teaching tools and that fantastical events can sometimes be beneficial to children's learning.

Furthermore, we were able to examine the nature of this effect in more detail than in prior studies because we compared children's learning of the same educational material from stories that contained two different types of impossible events: biological and physical. We found that stories that contained biologically impossible events did not generally help children's learning. Although there appeared to be some benefit to learning of the inheritance problem from the biological violations story in Study 1, it was weaker than the effect of physical violations and was not replicated in Study 2. On the other hand, stories that contained physically impossible events appeared to provide a more general benefit to children's learning; for the inheritance problem in both studies (although statistically significant only in Study 1) and for the balance problem in Study 2, children showed greater learning from the story with physical violations than from the story with no violations. Because the benefit of these physically impossible events was seen across both problem types, there does not appear to be a need for the story's target educational information to match the type of fantasy that the story contains in order for children to learn successfully.

In turn, these results suggest a possible mechanism for why fantastical events can sometimes benefit children's learning: There may be an interaction between children's existing knowledge of the physical domain and the physically impossible events presented in stories. Children understand the domain of physics very well by preschool; even infants demonstrate knowledge of some physical principles (e.g., Baillargeon, 2004; Spelke et al., 1992). Their knowledge of biology develops more slowly, however, and children in our target age range still do not fully understand the nature of biological processes (see Carey, 1985; Inagaki & Hatano, 2006). Thus, it is possible that the physical violations signaled to children more strongly the need to pay careful attention to the content of the story and whether the story events could happen in reality, increasing their learning (see Hopkins & Lillard, 2020; Weisberg et al., 2014). If this is indeed the case, then biological violations could potentially benefit learning for older children who have had more time to internalize the principles of biology; this is supported by the finding in Study 2 that children who scored higher at pretest benefitted more from the biological violations. Future work should examine this possibility directly.

However, it is also possible that a different or additional mechanism is at play. Children may be more generally interested in biological impossibility than in physical impossibility (i.e., biophilia) and hence may have been more distracted from the learning task by the impossible biological events. Such violations may also be less common in children's media and hence may have required more cognitive effort to process than the more familiar violations of physical principles (see Goldstein & Alperson, 2020).

Together, the two current studies reinforce the idea that the effect of children's prior knowledge interacts with the effects of fantasy (see also Weisberg & Hopkins, 2020). When the fantasy came from a domain with which children were less familiar (i.e., biology), or when the information being taught was more difficult (i.e., the balance principle in Study 1), the positive effects of fantasy were smaller or nonexistent. In fact, in Study 1 children learned the balance principle best from the realistic story. When children are less sure of themselves, they may benefit more from a high degree of similarity between the story and reality. Alternatively, when children are unsure of themselves, they may prefer to quarantine information from fantastical stories to avoid accidentally acquiring incorrect information. Therefore, it is overly simplistic to ask whether fantasy or reality is better for learning. The answer is more likely that either can be effective, but it depends on what is being taught and who is doing the learning (Strouse, Nyhout, & Ganea, 2018).

Future work should investigate this more nuanced effect of fantasy in more detail, especially with respect to how it might vary for different types of educational material. Here, we investigated only children's learning of scientific principles, but children's stories are often used to teach moral and social lessons, new vocabulary words, or historical or cultural information. It may be the case that impossible events are beneficial to children's learning across all these domains, but it is also possible that such events benefit children's learning for only some of these topics.

In addition, the current studies were limited in that the reading environment did not necessarily resemble naturalistic story sharing between adults and children. For experimental control, the experimenter reading the story did not include any extratextual information or draw children's attention to any particular aspect of the story, and she also did not respond to children's questions or comments. Presenting these stories with this kind of additional context and responsiveness could serve to highlight the new scientific principles presented in the story and hence positively affect children's learning. It would also be important to measure differences in children's level of engagement or the types of questions children tend to ask during reading, which could link to their level of engagement and hence to their learning. Other limitations of the current work that future work could address are the issues that children heard the story only once and that the stories were created in a laboratory and were not as polished as commercially available books.

Finally, it is possible that the benefit of physically fantastical events may vary depending on other individual characteristics of the learners. For example, children differ on the degree to which they prefer engaging with fantastical media or include fantasy themes in their play (Gilpin, 2009; Sharon & Woolley, 2004). This variation has been shown to affect their learning from stories (Richert & Smith, 2011; Woolley, Boerger, & Markman, 2004). Similarly, although we did not find this effect in the current studies, children's ability to tell the difference between fantasy and reality may also affect their learning from fantastical stories (Mares & Sivakumar, 2014; Richert & Schlesinger, 2016). By taking into account not only story-level factors but also content-level and individual-level factors, future work in this area could continue to probe which types of stories are most beneficial for all learners.

Acknowledgments

Many thanks to all of the children who were involved in this research, as well as their parents and teachers and our partners at local schools and museums who made this work possible. Thanks also to the members of the University of Pennsylvania Cognition & Development Lab and the Scientific Thinking and Representation (STAR) Lab at Villanova University, especially Elysia Choi and Natasha Chlebuch, for their assistance in assembling stimuli and collecting data.

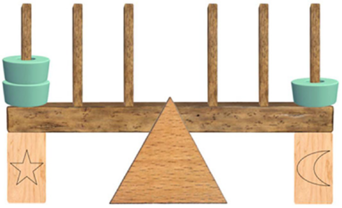
Appendix A

For the balance problem, trials varied on whether the two sides of the scale differed in weight (W) or distance (D) from the fulcrum, whether there were 1 versus 2 weights/pegs (1:2) or 1 versus 3 weights/pegs (1:3), and whether the correct response was presented on the left (L) or on the right (R). Each child received four different trials in 1 of the 16 predetermined orders to which they were randomly assigned.

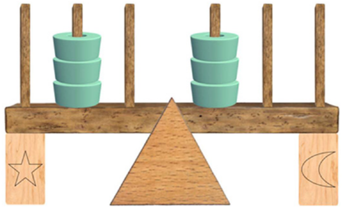
Trial order	Trial 1	Trial 2	Trial 3	Trial 4
1	W 1:2 L	D 1:2 L	D 1:3 R	W 1:3 R
2	W 1:2 R	D 1:2 L	D 1:3 R	W 1:3 L
3	W 1:3 L	D 1:3 L	D 1:2 R	W 1:2 R
4	W 1:3 R	D 1:3 L	D 1:2 R	W 1:2 L
5	D 1:2 L	W 1:2 L	W 1:3 R	D 1:3 R
6	D 1:2 R	W 1:2 L	W 1:3 R	D 1:3 L
7	D 1:3 L	W 1:3 L	W 1:2 R	D 1:2 R
8	D 1:3 R	W 1:3 L	W 1:2 R	D 1:2 L
9	W 1:2 L	D 1:2 R	D 1:3 L	W 1:3 R
10	W 1:2 R	D 1:2 R	D 1:3 L	W 1:3 L
11	W 1:3 L	D 1:3 R	D 1:2 L	W 1:2 R
12	W 1:3 R	D 1:3 R	D 1:2 L	W 1:2 L
13	D 1:2 L	W 1:2 R	W 1:3 L	D 1:3 R
14	D 1:2 R	W 1:2 R	W 1:3 L	D 1:3 L
15	D 1:3 L	W 1:3 R	W 1:2 L	D 1:2 R
16	D 1:3 R	W 1:3 R	W 1:2 L	D 1:2 L

For example, a child assigned to Trial Order 1 would be presented with the following four problems:

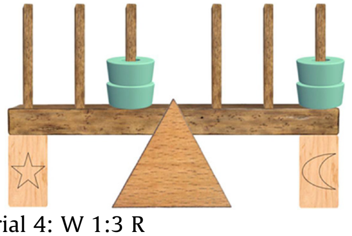
Trial 1: W 1:2 L



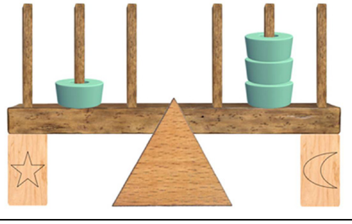
Trial 2: D 1:2 L



Trial 3: D 1:3 R



Trial 4: W 1:3 R



For the inheritance problem, trials varied on whether the child was asked about an inborn trait (I) or an acquired trait (A), whether the trait was applied to cats or bunnies, and whether the correct response was presented on the left (L) or on the right (R). Each child received four different trials in 1 of the 16 predetermined orders to which they were randomly assigned.

Trial order	Trial 1	Trial 2	Trial 3	Trial 4
1	I Cat L	A Cat L	A Bunny R	I Bunny R
2	I Cat R	A Cat L	A Bunny R	I Bunny L
3	I Bunny L	A Bunny L	A Cat R	I Cat R
4	I Bunny R	A Bunny L	A Cat R	I Cat L
5	A Cat L	I Cat L	I Bunny R	A Bunny R
6	A Cat R	I Cat L	I Bunny R	A Bunny L
7	A Bunny L	I Bunny L	I Cat R	A Cat R

Appendix A (*continued*)

Trial order	Trial 1	Trial 2	Trial 3	Trial 4
8	A Bunny R	I Bunny L	I Cat R	A Cat L
9	I Cat L	A Cat R	A Bunny L	I Bunny R
10	I Cat R	A Cat R	A Bunny L	I Bunny L
11	I Bunny L	A Bunny R	A Cat L	I Cat R
12	I Bunny R	A Bunny R	A Cat L	I Cat L
13	A Cat L	I Cat R	I Bunny L	A Bunny R
14	A Cat R	I Cat R	I Bunny L	A Bunny L
15	A Bunny L	I Bunny R	I Cat L	A Cat R
16	A Bunny R	I Bunny R	I Cat L	A Cat L

For example, a child assigned to Trial Order 1 would be presented with the following four problems:

Trial 1: I Cat L

Some cats have all black fur, and some have black and white fur. This cat [left] was born with all black fur. When she has babies, will they have all black fur or black and white fur?

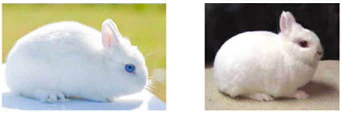
Trial 2: A Cat L

Some cats have ears that stick up, and some have ears that flop down. This cat [right] hurt her ear in an accident so now it flops down. When she has babies, will they have ears that stick up or ears that flop down?

Trial 3: A Bunny R

Some bunnies have no tails, and some have fluffy tails. This bunny [left] lost her tail in an accident. When she has babies, will they have no tails or fluffy tails?

Trial 4: I Bunny R



Some bunnies have brown eyes, and some have blue eyes. This bunny [right] was born with blue eyes. When she has babies, will they have brown eyes or blue eyes?

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2020.105047.

References

Baillargeon, R. (2004). Infants' reasoning about hidden objects: Evidence for event-general and event-specific expectations. *Developmental Science, 74*, 391–424.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
Cook, C., & Sobel, D. M. (2011). Children's beliefs about the fantasy/reality status of hypothesized machines. *Developmental Science, 14*, 1–8.
Dias, M. G., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology, 6*, 207–221.
Dias, M. G., & Harris, P. L. (1990). The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology, 8*, 305–318.
Fisch, S. M. (2000). A capacity model of children's comprehension of educational content on television. *Media Psychology, 2*, 63–91.
Ganea, P. A., Canfield, C. F., Simons-Ghafari, K., & Chou, T. (2014). Do cavies talk? The effect of anthropomorphic picture books on children's knowledge about animals. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00283.
Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition, 38*, 213–244.
Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science, 2*, 336–340.
Gilpin, A. T. (2009). Mechanisms for overcoming reality status biases. Unpublished doctoral dissertation, University of Texas at Austin.
Goldstein, T. R., & Alperson, K. (2020). Dancing bears and talking toasters: A content analysis of supernatural elements in children's media. *Psychology of Popular Media, 9*, 214–223.
Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002). Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology, 81*, 417–445.
Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Child Development, 55*, 2042–2055.
Hopkins, E. J., & Lillard, A. S. (2020). The Magic Schoolbus dilemma: How fantasy affects children's learning from stories. Manuscript submitted for publication.
Hopkins, E. J., & Weisberg, D. S. (2017). The youngest readers' dilemma: A review of children's learning from fictional sources. *Developmental Review, 43*, 48–70.
Inagaki, K., & Hatano, G. (2006). Young children's conception of the biological world. *Current Directions in Psychological Science, 15*, 177–181.
Lillard, A. S., Drell, M., Richey, E., Boguszewski, K., & Smith, E. D. (2015). Further examination of the immediate impact of television on children's executive function. *Developmental Psychology, 51*, 792–805.
Lillard, A. S., & Sobel, D. M. (1999). Lion Kings or puppies: The influence of fantasy on children's understanding of pretense. *Developmental Science, 2*, 75–80.
Mares, M.-L., & Sivakumar, G. (2014). "Vámonos means go, but that's made up for the show": Reality confusions and learning from educational TV. *Developmental Psychology, 50*, 2498–2511.
Richert, R. A., & Schlesinger, M. A. (2016). The role of fantasy–reality distinctions in preschoolers' learning from educational video. *Infant and Child Development, 26*, 1–17.
Richert, R. A., Shawber, A. B., Hoffman, R. E., & Taylor, M. (2009). Learning from fantasy and real characters in preschool and kindergarten. *Journal of Cognition and Development, 10*, 41–66.
Richert, R. A., & Smith, E. I. (2011). Preschoolers' quarantining of fantasy stories. *Child Development, 82*, 1106–1119.
Sharon, T., & Woolley, J. D. (2004). Do monsters dream? Young children's understanding of the fantasy/reality distinction. *British Journal of Developmental Psychology, 22*, 293–310.
Sobel, D. M. (2006). How fantasy benefits young children's understanding of pretense. *Developmental Science, 9*, 63–75.
Sobel, D. M., & Weisberg, D. S. (2014). Tell me a story: How children's developing domain knowledge affects their story construction. *Journal of Cognition and Development, 15*, 465–478.
Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*, 605–635.
Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development, 60*, 637–648.
Springer, K., & Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development, 62*, 767–781.
Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348*, 91–94.
Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition, 163*, 1–14.
Stahl, A. E., & Feigenson, L. (2019). Violations of core knowledge shape early learning. *Topics in Cognitive Science, 11*, 136–153.
Strouse, G. A., Nyhout, A., & Ganea, P. A. (2018). The role of book features in young children's transfer of information from picture books to real-world contexts. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00050.
Walker, C. M., Gopnik, A., & Ganea, P. A. (2015). Learning to learn from stories: Children's developing sensitivity to the causal structure of fictional worlds. *Child Development, 86*, 310–318.
Weaver, J., Zillmann, D., & Bryant, J. (1988). Effects of humorous distortions on children's learning from educational television: Further evidence. *Communication Education, 37*, 181–187.
Weisberg, D. S. (2016). The fantasy advantage. *Scientific American. Mind, 27*, 43–47.
Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., & McCandliss, B. D. (2014). Mise en place: Setting the stage for thought and action. *Trends in Cognitive Sciences, 18*, 276–278.
Weisberg, D. S., & Hopkins, E. J. (2020). Preschoolers' extension and export of information from realistic and fantastical stories. *Infant and Child Development, 29* e2182.
Weisberg, D. S., Ilgaz, H., Hirsh-Pasek, K., Golinkoff, R. M., Nicolopoulou, A., Dickinson, D. K., & Nicolopoulou, A. (2015). Shovels and swords: How realistic and fantastical themes affect children's word learning. *Cognitive Development, 35*, 1–14.
Williams, J. M., & Smith, L. A. (2010). Concepts of kinship relations and inheritance in childhood and adolescence. *British Journal of Developmental Psychology, 28*, 523–546.
Woolley, J. D., Boerger, E. A., & Markman, A. B. (2004). A visit from the Candy Witch: Factors influencing young children's belief in a novel fantastical being. *Developmental Science, 7*, 456–468.
Zillmann, D., Masland, J. L., Weaver, J. B., Lacey, L. A., Jacobs, N. E., Dow, J. H., ... Banker, S. R. (1984). Effects of humorous distortions on children's learning from educational television. *Journal of Educational Psychology, 76*, 802–812.