AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

https://doi.org/10.1037/dev0001670

Improbable but Possible: Training Children to Accept the Possibility of Unusual Events

Andrew Shtulman[1], Brandon Goulding[2], and Ori Friedman[3]

[1] Department of Psychology, Occidental College

[2] Department of Psychology, University of Winnipeg

[3] Department of Psychology, University of Waterloo

Young children tend to deny the possibility of events that violate their expectations, including events that are merely improbable, like making onion-flavored ice cream or owning a crocodile as a pet. Could this tendency be countered by teaching children more valid strategies for judging possibility? We explored this question by training children aged 4–12 ($n = 128$) to consider either the similarity between the target event and unusual events that have actually occurred or causal mechanisms that might bring the target event about. Both trainings increased children's acceptance of improbable events but only for the types of events addressed during training. Older children were more likely to accept improbable events, as were children who scored higher on a measure of cognitive reflection, but neither age nor cognitive reflection moderated the effects of training. These findings indicate that children can use both similarity and causality to assess possibility, but the use of this information is highly circumscribed, further demonstrating how robustly children conflate improbability with impossibility.

Public Significance Statement

Children tend to claim that unusual events are impossible, and this study explored the source of that tendency by training them to consider real events similar to the target events or mechanisms that cause the target events to occur. We find that both types of information increase children's acceptance of unusual events but only for events similar to those covered during training, which suggests that children appreciate the kinds of considerations relevant to possibility but do not seek out those considerations on their own.

Keywords: possibility judgment, causal reasoning, similarity, modal cognition

Could a person own a crocodile for a pet in real life? Could a person make onion-flavored ice cream in real life? These questions are about possibility, not factuality. No one in the history of the world may have owned a crocodile for a pet or made onion-flavored ice cream, but if we can identify circumstances for bringing these events about, then we can be reasonably confident that they are possible. For instance, we can be confident that owning a pet crocodile is possible by imagining how someone might find a baby crocodile in the wild and raise it in their backyard. We can be confident that making onion-flavored ice cream is possible by imagining how someone could chop up an onion and add the pieces to the ingredients for regular ice cream.

This pathway to affirming possibility draws on our knowledge of causal mechanisms—mechanisms that explain how the event might occur. An alternative approach is to draw on our knowledge of similar events—events that establish a precedent for the event under consideration. For instance, one might decide that owning a pet crocodile is possible because people own other exotic animals, such as tortoises and pythons. Crocodiles may seem similar enough to tortoises and pythons to accept that they too could be pets. Likewise, one might decide that making onion-flavored ice cream is possible because people have made other exotic flavors of ice cream, such as bean ice cream and spinach ice cream. Onions may seem similar enough to beans and spinach to accept that they too could be incorporated into ice cream.

These two types of considerations—causal mechanisms and empirical precedents—are each sufficient to overcome the knee-jerk reaction that owning a crocodile and making onion-flavored ice cream is impossible, at least sufficient for adults. Adults' intuitive sense of possibility is influenced by our perception of what is normal, typical, or useful (Phillips et al., 2019), but our final judgments

Andrew Shtulman https://orcid.org/0000-0002-4687-3099

This research was supported by an Understanding Human Cognition Scholar Award to Andrew Shtulman from the James S. McDonnell Foundation. All data and analyses can be found at the Open Science Framework (https://tinyurl.com/ynk25a74 and https://osf.io/t6bmf).

Andrew Shtulman served as lead for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing–original draft. Brandon Goulding served in a supporting role for conceptualization, formal analysis, and methodology. Ori Friedman served as lead for writing–review and editing and served in a supporting role for conceptualization, formal analysis, and methodology.

Correspondence concerning this article should be addressed to Andrew Shtulman, Department of Psychology, Occidental College, 1600 Campus Road, Los Angeles, CA 90041, United States. Email: shtulman@oxy.edu

frequently take mechanisms and precedents into account as well. That is, when judging the possibility of events we have neither experienced nor expect to, we consider the similarity between it and other unusual events that have actually occurred (Goulding & Friedman, 2023), as well as conditions or circumstances that might bring the event about (Shtulman & Tong, 2013).

Young children, on the other hand, do not rely much on either consideration; they often jump to the conclusion that an event cannot happen from their expectation that it would not happen or should not happen (Shtulman & Phillips, 2018). Children reliably claim that impossible events, like walking on water or walking through a wall, cannot happen in real life, but they also claim that improbable events cannot happen either, including events like owning a pet crocodile or eating onion-flavored ice cream (Goulding & Friedman, 2021; Nolan-Reyes et al., 2016; Shtulman & Carey, 2007). Indeed, 4-year-olds deny the possibility of improbable events nearly as often as they deny the possibility of genuinely impossible ones, and they do so regardless of whether the events are physical, biological, or psychological in nature (Cook & Sobel, 2011; Shtulman, 2009; Weisberg & Sobel, 2012).

Young children's rejection of improbable events is robust across contexts, including both the context in which they hear of the event and the context in which the event is purported to occur. They claim that improbable events found in a book or on the internet are impossible (Danovitch & Lane, 2020), and they are not particularly swayed by whether the events are conveyed as testimony ("I saw someone eat onion-flavored ice cream") or hearsay ("Someone told me that they ate onion-flavored ice cream"; Lane et al., 2018). They are skeptical of improbable events relayed by familiar informants (Williams & Danovitch, 2022) or several informants (Nissel et al., 2023), as well as those purported to occur in a faraway land (Bowman-Smith et al., 2019), in a dream (Goulding & Friedman, 2020), or by someone who really wants the events to occur (Chernyak et al., 2013). Children's judgments do vary from one context to another, but the variation is on the side of skepticism, not acceptance; they vacillate in how strongly or how consistently they judge improbable events impossible.

Prompting children to think more deeply about improbable events does not change their judgments either. Children who are prompted to explain how an improbable event might occur still deny that it could occur (Nancekivell & Friedman, 2017). They might explain, for instance, that a person who owns a peacock in a story bought that peacock at a pet store but then deny that someone could own a peacock in real life. Prompting children to envision improbable events in their mind also does not change their judgments (Lane et al., 2016). Children who envision themselves buying a peacock at a pet store will continue to deny that someone could own a peacock in real life—presumably because pet stores do not sell peacocks.

Similarity and Causality

A prompt that has proven effective at changing children's judgments about improbable events is informing them of similar events that have actually taken place (Goulding & Friedman, 2021). Children who would normally deny that someone could own a zebra for a pet will often accept this possibility if given examples of other unusual pets that people actually own, such as a pet elephant. In contrast, children given examples of typical pets, such as a pet dog, continue to deny the possibility of owning a zebra, indicating that they change their mind only when given a similar example—an example that bridges the gap between the target event and ordinary events known from everyday experience. The similar event establishes a precedent against which the target event appears less extraordinary.

Considering real-world precedents is one of two methods that adults use to assess possibility, as noted earlier. The other method is considering causal mechanisms. Prompting children to use this method appears to be less effective (Goulding et al., 2022). Children are unlikely to affirm the possibility of making onion-flavored ice cream if prompted to consider the mechanisms involved in making ordinary ice cream, such as blending and freezing the appropriate ingredients. Although blending and freezing are familiar mechanisms, children are disinclined to use them to model unfamiliar outcomes.

Prompting children to think of relevant mechanisms appears to be helpful only if tied to a precedent (Goulding et al., 2022), such as telling children that some people make pickle-flavored ice cream in real life and then explaining how. Providing children with both a precedent and a mechanism leads them to accept the possibility of improbable events 15%–20% more often than providing them with precedents alone. Still, it is unclear from this finding whether children's success is boosted by the inclusion of a precedent or the reframing of the mechanism to address how an improbable event might occur by ordinary means. That is, it remains an open question whether children would accept the possibility of pickle-flavored ice cream if they were told how a person could make pickle-flavored ice cream without stipulating that such ice cream exists. Without a precedent, children might treat mechanism information as purely hypothetical, which is not unreasonable. Magic spells provide a "mechanism" for bringing about impossible events, but it would be a mistake to accept a story about a spell as evidence of possibility.

How Malleable Are Children's Judgments?

The discovery that children's reasoning about improbable events can be improved by introducing precedents, either alone or in combination with relevant mechanisms, raises several questions about the malleability of children's possibility judgments. First, are precedents generally superior to mechanisms? Or might the latter be equally useful if framed in terms of improbable events? Precedents appear to be helpful because they bridge the gap between improbable events and ordinary ones, and mechanisms might be similarly helpful if used to model improbable outcomes rather than ordinary ones. On the other hand, precedents may be needed to convince children that ordinary mechanisms can, in fact, bring about improbable outcomes. Goulding et al. (2022) found that telling children that pickle-flavored ice cream exists and then explaining how it is made led them to accept the possibility of other unusual ice creams to a greater extent than simply telling them that pickle-flavored ice cream exists, but this finding leaves open the question of whether teaching children how to make pickle-flavored ice cream would be helpful on its own. Teaching children how to make strawberry-flavored ice cream was not helpful, which suggests that establishing the existence of an unusual precedent may have been necessary for children to learn anything from the causal information that followed.

Second, what is the scope of children's improvement? Does providing children with precedents and mechanisms improve their reasoning about improbable events in general or just those relevant to

the input provided? Such input helps children bridge the gap between ordinary events and closely related events, but it may also help them recognize that precedents and mechanisms are useful considerations for reasoning about possibility in general. Accordingly, children trained to consider precedents or mechanisms may increase their acceptance of improbable events beyond those used to illustrate their utility.

Third, how is the improvement in children's reasoning influenced by age and cognitive ability? Do older children benefit from possibility-judgment training more than younger children? The earlier training studies that involved precedents and mechanisms (Goulding et al., 2022; Goulding & Friedman, 2021) tested children within a narrow age range—five to seven—which may have underestimated the effectiveness of training. Older children know more precedents and mechanisms than younger children, so older children may be more successful at implementing a strategy that depends on such considerations.

In a similar vein, we sought to determine whether training is more beneficial for children who tend to privilege analysis over intuition—a disposition known as cognitive reflection (Frederick, 2005). Tests of cognitive reflection assess the propensity to monitor and correct an intuitive response. For instance, the Cognitive Reflection Test, Developmental Version (CRT-D; Young & Shtulman, 2020) consists of several brainteasers designed to elicit an intuitive, yet incorrect, response that children can correct upon further reflection, such as "Which weighs more: a pound of rocks or a pound of feathers?" Children who are better at answering such brainteasers demonstrate higher levels of rational thought and conceptual understanding (Shtulman & Young, 2023). They may also be better at using precedents and mechanisms to scrutinize their intuitions about possibility, especially after training.

In the present study, we investigated these questions by training children to use precedents or mechanisms to assess the possibility of unusual events. We investigated the first question by comparing the effectiveness of mechanisms to that of precedents, with the expectation that both would be effective if explicitly framed in terms of improbable events. We investigated the second question by asking children to judge the possibility of several categories of improbable events, some addressed during training and some not, with the expectation that training would increase children's acceptance of the former but may or may not increase their acceptance of the latter. And we investigated the third question by looking for interactions between training, age, and cognitive reflection, with the expectation that training would have a greater impact on older children and more cognitively reflective children.

On the whole, we sought to clarify the roles of causality and similarity in children's reasoning about possibility, as well as how their reasoning is shaped by age and reflectiveness. Training children to use information that contributes to adults' judgments can shed light on the cognitive underpinnings of children's judgments insofar that children are able to use that information. If children can use precedents or mechanisms to affirm the possibility of improbable events, then their understanding of possibility would appear to be consistent with adults, albeit less informed. And if children can apply the possibility-judgment strategies modeled during training to events beyond the training context, then children's reluctance to affirm unusual possibilities may stem from not explicitly seeking out information relevant to those possibilities.

Method

Participants

The participants were 128 children between the ages of four and twelve ($M_{age} = 7.6$, $SD = 2.0$; 5% were four, 16% five, 24% six, 17% seven, 15% eight, 8% nine, 5% ten, 6% eleven, and 4% twelve). We chose a wide age range to maximize variability in possibility judgments, as well as variability in receptiveness to instruction, on the assumption that older children would be more receptive to instruction than younger ones. We chose four as our lower bound to ensure that all participants understood the language used to elicit possibility judgments, namely, "Could a person ___ in real life?" By age 4, children use modal verbs like could, can, may, and might in their speech and understand that these verbs express possibility rather than factuality (see Byrnes & Duff, 1989; Ozturk & Papafragou, 2015).

Children were recruited from public playgrounds in the Los Angeles area and tested onsite. Six additional children were recruited but did not complete the session. The final sample was balanced for gender (51% female, 49% male) and was racially diverse (34% White, 23% Hispanic, 20% Asian, 10% Black, and 13% mixed).

Our sample size was determined by how many participants we could recruit and test in a 6-month period, with the goal of running at least 50 participants in each training condition. A retrospective power analysis (performed in G*Power 3.1.9.6) indicated that our final sample had an 89% probability of detecting a medium-sized, between-participant effect of training type (similarity vs. causality) and a 99% probability of detecting a medium-sized, within-participants effect of assessment period (pretest vs. posttest) in a repeated-measures analysis of variance (ANOVA).

Materials

Participant judged the possibility of 42 events across two assessment periods: pretest and posttest. The events were drawn from three categories: pets, ice creams, and houses. And each category contained seven events: two ordinary, two impossible, and three improbable.

The ordinary events were events that children had likely observed firsthand, such as owning a pet cat, making strawberry-flavored ice cream, and building a house out of wood. The improbable events were events that children would not have observed but could under the right circumstances, such as owning a pet peacock, making pickle-flavored ice cream, and building a house out of ice. These events violated empirical regularities but no physical laws. The impossible events, in contrast, violated physical laws and could never be observed; they included owning a pet unicorn, making lava-flavored ice cream, and building a house out of clouds.

The full list of events is presented in Table 1. We included more improbable events than ordinary events or impossible events because they were the only type expected to elicit variability, as well as the only type expected to change with training. Each event was accompanied by an image depicting the target animal (in the case of pets), the target ingredient (in the case of ice creams), or the target material (in the case of houses). Sample illustrations are presented in Table 2.

Children judged the possibility of each event in two stages. First, they were asked whether the event could happen in real life. If they said yes, the experimenter moved onto the next event. If they said no, they were asked whether the event was "kinda impossible" or "very impossible." This way of eliciting possibility judgments allowed

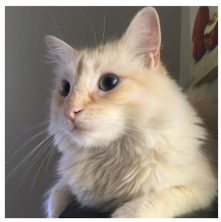
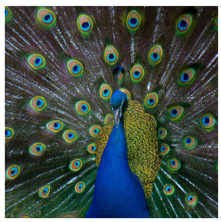
Table 1
Items Used to Create Ordinary, Improbable, and Impossible Events in Each Event Category (Pets, Ice Creams, and Houses)

Could a person…	Ordinary	Improbable	Impossible
Own a pet ___ in real life?	Cat	Buffalo	Dragon
	Dog	Chimpanzee	Fairy
	Hamster	Dolphin	Gnome
	Rabbit	Peacock	Unicorn
		Skunk	
		Zebra	
Make ___-flavored ice cream in real life?	Caramel	Cauliflower	Lava
	Chocolate	Chili pepper	Lightning
	Oreo	Garlic	Moon
	Strawberry	Mushroom	Thunder
		Oyster	
		Pickle	
Build a house out of ___ in real life?	Bricks	Ice	Clouds
	Logs	Marshmallows	Fire
	Stone	Paper	Rainbows
	Wood	Seashells	Wind
		Tires	
		Toothpicks	

children to qualify their judgments, thus increasing the sensitivity of our measure and, accordingly, its responsiveness to training. We scored children's judgments on a scale from 0 to 2, where 2 corresponded to "possible" (i.e., "could happen in real life"), 1 to "kinda impossible," and 0 to "very impossible."

This rating scale has been used successfully in previous studies (Shtulman et al., 2023; Shtulman & Phillips, 2018), as children readily distinguish between "kinda impossible" and "very impossible" without instruction or prompting. These response options provide children with an opportunity to distinguish improbable events from impossible ones even if they are inclined to judge both as impossible. Indeed, children in the present study selected "kinda impossible" most often for improbable events and "very impossible" most often for impossible events, as shown in Table 3. This correspondence validates

Table 2
Sample Images Used to Illustrate Ordinary, Improbable, and Impossible Events

Event category	Ordinary	Improbable	Impossible
Pets			
Ice creams			
Houses			

Note. With the exception of the cat photograph, the images were generated with the assistance of the DALL-E AI system. See the online article for the color version of this table.

Table 3
Mean Judgment Scores (and Standard Deviations) at Each Assessment Period, Along With the Proportion of Judgments That Were "Very Impossible" (0), "Kinda Impossible" (1), or "Possible" (2)

					Judgment		
Assessment	Event category	Event type	*M*	*SD*	0	1	2
Pretest	Pets	Ordinary	2.0	0.2	.01	.00	.99
		Improbable	0.9	0.7	.38	.30	.32
		Impossible	0.3	0.5	.75	.15	.10
	Ice creams	Ordinary	1.9	0.3	.02	.02	.96
		Improbable	1.2	0.7	.30	.25	.45
		Impossible	0.2	0.3	.85	.14	.01
	Houses	Ordinary	1.8	0.4	.05	.05	.90
		Improbable	0.8	0.6	.46	.30	.24
		Impossible	0.1	0.3	.88	.09	.03
Posttest	Pets	Ordinary	2.0	0.2	.01	.01	.98
		Improbable	1.1	0.7	.32	.26	.42
		Impossible	0.2	0.4	.83	.09	.08
	Ice creams	Ordinary	2.0	0.2	.01	.01	.98
		Improbable	1.4	0.7	.19	.21	.60
		Impossible	0.2	0.4	.87	.10	.03
	Houses	Ordinary	1.8	0.4	.05	.05	.90
		Improbable	0.7	0.6	.52	.25	.23
		Impossible	0.2	0.4	.87	.09	.04

children's understanding of the scale, as well as its sensitivity to nascent modal distinctions.

The study was reviewed by Occidental College's Institutional Review Board and approved as SP22-17-SHTU. The parents of all participants provided informed consent, and the participants provide verbal assent. The study was not preregistered.

Procedure

Participants judged the possibility of 21 events—seven from each of three categories—before and after an experimenter-led training. The training highlighted a general strategy for judging possibility. Approximately half of the participants ($n = 69$) were taught the strategy of identifying known events similar to the target event (the similarity training), and the other half ($n = 59$) were taught the strategy of identifying causal mechanisms that might bring the event about (the causality training). Our rationale for using a pre–post design, as opposed to a between-subject design, was to increase the sensitivity of our intervention, as each child provided their own baseline against which the effectiveness of training could be measured. The trainings took 5–7 min to complete, and both the similarity training and the causality training used the same improbable events as illustrations.

The first illustration was owning a pet crocodile. Participants who received the similarity training were asked to think of other unusual pets. They were then told that some people own pet elephants, pet tortoises, and pet pythons in real life and were encouraged to see the similarity between these pets and a pet crocodile. Participants who received the causality training were asked to think of how a person might own a pet crocodile in real life. They were then told a person could find a baby crocodile in the wild, raise it in their backyard, and feed it fish or porkchops. In both conditions, owning a pet crocodile was treated as a hypothetical event and never stipulated as real, so as not to introduce a precedent into the causality training.

Next, participants were asked to consider the possibility of making onion-flavored ice cream. Participants in the similarity training were told about other unusual ice creams that people have made for real (broccoli ice cream, spinach ice cream, bean ice cream), and participants in the causality training were told the steps involved in making onion-flavored ice cream (chopping onions and stirring them into a mixture of milk and sugar). This event, like the previous one, was framed as hypothetical rather than actual. The precedents provided in the similarity training were illustrated with photos, as were the causal steps described in the causality training. We should also note that, in the similarity training, we used the word "unusual" to prompt children to think of precedents beyond those encountered in daily life but did not insinuate that such precedents were inappropriate or bad. Below, we use "unusual" in this same (statistical) sense.

The trainings not only illustrated normative strategies for judging possibility but also began and ended with a general description of that strategy. The similarity training stressed that "a good way to decide whether something is possible is to try to think of something similar that you know is real," and the causality training stressed that "a good way to decide whether something is possible is to try to think of how it could happen." Critically, both trainings mentioned unusual pets and unusual ice creams but not unusual houses. Unusual houses were left unmentioned to test whether children would generalize the strategies illustrated with the first two categories to the third. The scripts for both trainings are included in the Appendix.

Following the training, participants judged the possibility of 21 new events. The events listed in Table 1 were assigned to one of two assessments counterbalanced across participants; pretest items for half the participants in each condition appeared as posttest items for the other half and vice versa. Preliminary analyses revealed no effect of counterbalancing on participants' judgments and no interaction between counterbalancing and training condition, so we dropped this variable from subsequent analyses.

Participants made their judgments in blocks, such that all seven events from an event category were presented together. Within a block, participants saw one ordinary event and one impossible

event before seeing any improbable events, as a way of acclimating them to the task. The remaining events were presented in a random order, as were the blocks themselves. For each event, children were asked whether it could happen "in real life." The latter phrase was intended to clarify that scope of the judgment and block the interpretation that we were asking whether the event could happen "under normal circumstances." Findings from previous studies (Shtulman & Carey, 2007; Shtulman & Phillips, 2018) suggest that "in real life" is sufficient to specify the relevant meaning of could, and older children in the current study clearly understood this meaning, as they judged the majority of improbable events in each event category possible.

In addition to judging possibility, participants also completed a nine-item assessment of their cognitive reflection: the CRT-D (Young & Shtulman, 2020). The CRT-D measures children's disposition to privilege analysis over intuition and consists of nine brainteasers. One such brainteaser is "What do cows drink?" The semantic association between cows and milk prompts the intuitive response "milk," but a moment's reflection reveals that, although cows produce milk, they drink water. Children completed the CRT-D at the beginning of each experimental session, and their responses were scored for accuracy, yielding a composite score that could range from 0 to 9. In actuality, these scores ranged from 0 to 7 and averaged 2.6 ($SD = 1.6$). Children's CRT-D scores were strongly correlated with age ($r = .52$, $p < .001$), consistent with previous studies (see Shtulman & Young, 2023).

The CRT-D was administered on iPads using the Qualtrics app, as were the pretest, posttest, and training. All materials, data, and analyses can be found on the Open Science Framework: https://osf.io/t6bmf.

Results

Participants' possibility judgments are displayed as a function of assessment period (pretest, posttest), event category (pets, ice creams, houses), and event type (ordinary, improbable, impossible) in Table 3. These judgments are displayed both as means and as response distributions—namely, the proportion of events judged "very impossible" (scored 0), "kinda impossible" (scored 1), and "possible" (scored 2). Participants generally accepted the possibility of ordinary events, affirming they could happen in real life at both pretest and posttest, and rejected the possibility of impossible events, judging them "very impossible" at both pretest and posttest. Participants' judgments for improbable events were more variable. At both assessment periods, participants' mean judgments hovered around 1, corresponding to "kinda impossible." However, the distribution of judgments shifted toward "possible" (and away from "very impossible") between pretest and posttest but only for the trained categories.

We analyzed the reliability of these effects with a repeated-measures ANOVA. Event category, event type, and assessment period were within-participants factors, and training type (similarity, causality) was a between-participant factor. This analysis revealed main effects of event category, $F(2, 252) = 52.00$, $p < .001$, $\eta_p^2 = .292$, and event type, $F(2, 252) = 909.59$, $p < .001$, $\eta_p^2 = .878$, but no main effects of assessment period, $F(1, 126) = 3.17$, $p = .078$, $\eta_p^2 = .025$, or training type, $F(1, 126) = 0.05$, $p = .83$, $\eta_p^2 < .001$. Training type did not interact with any of the other variables, but assessment period interacted with both event category, $F(2, 252) = 7.27$, $p < .001$, $\eta_p^2 = .055$, and event type, $F(2, 252) = 9.43$, $p < .001$, $\eta_p^2 = .070$. The three-way interaction between assessment period, event category, and event type was also significant, $F(4, 504) = 8.27$, $p < .001$, $\eta_p^2 = .062$.

To explore these interactions, we ran repeated-measures ANOVAs on each event category by itself. These analyses revealed a main effect of event type for all three categories—pets: $F(2, 252) = 468.09$, $p < .001$, $\eta_p^2 = .788$; ice creams: $F(2, 252) = 638.88$, $p < .001$, $\eta_p^2 = .835$; houses: $F(2, 252) = 640.75$, $p < .001$, $\eta_p^2 = .836$—as well as interactions between event type and assessment period for the trained categories of pets, $F(2, 252) = 12.37$, $p < .001$, $\eta_p^2 = .089$, and ice creams, $F(2, 252) = 12.45$, $p < .001$, $\eta_p^2 = .090$, but not the untrained category of houses, $F(2, 252) = 0.45$, $p = .64$, $\eta_p^2 = .004$.

We further analyzed the effects of assessment period by comparing pretest scores with posttest scores for each type of event. Paired t tests revealed that scores for improbable events increased in the trained categories of pets, $t(127) = 3.30$, $p = .001$, $d = 0.29$, and ice creams, $t(127) = 4.99$, $p < .001$, $d = 0.44$, but not the untrained category of houses, $t(127) = -1.54$, $p = .126$, $d = 0.04$. These tests also revealed that scores for impossible events decreased in the pets category, $t(127) = -2.64$, $p = .009$, $d = -0.23$—an unexpected effect, possibly due to due to baseline differences in the acceptance of impossible events at pretest (children accepted impossible pets slightly more often than impossible ice creams or impossible houses). All other scores remained comparable from pretest to posttest.

Training thus improved children's judgments in the trained categories, but to what degree? Did children become more likely to judge improbable events as possible or less likely to judge them as "very impossible," choosing "kinda impossible" instead? To address this question, we summed the number of events in each event category judged possible and compared those sums across assessment periods with paired t tests (see Figure 1). There was a significant increase in the number of improbable pets judged possible, $M_{\text{difference}} = 0.31$, $t(127) = 2.96$, $p = .004$, and the number of improbable ice creams judged possible, $M_{\text{difference}} = 0.45$, $t(127) = 5.33$, $p < .001$, but no increase in the number of improbable houses judged possible, $M_{\text{difference}} = 0.04$, $t(127) = 0.46$, $p = .65$. Training thus increased children's acceptance that improbable events can happen, moving them across the line from "impossible" to "possible" for one or more events.

Training was effective on the whole (for the trained categories), but was one type of training more effective than the other? The omnibus analysis presented at the beginning of this section revealed no effect of training type, and neither did the category-specific analyses conducted as follow-ups. There was no main effect of training type for any category, nor did training type interact with event type or assessment period for any category (all $ps > .20$). The absence of such effects indicates that the similarity training and the causality training were equally effective at increasing participants' acceptance of improbable events, at least for the trained categories (see Figure 2). Indeed, the Bayes factor for the effect of training type on posttest judgments for improbable events was around .2 for all categories (.193 for pets, .195 for ice creams, .203 for houses; prior = .707), indicating it is five times likelier that the trainings exerted comparable effects than that one training was more effective than the other.

Finally, we explored how possibility judgments for improbable events varied by participants' age and CRT-D scores. We entered

Figure 1

The Number of Participants Who Judged 0, 1, 2, or 3 Improbable Events Possible in Each Event Category at Each Assessment Period

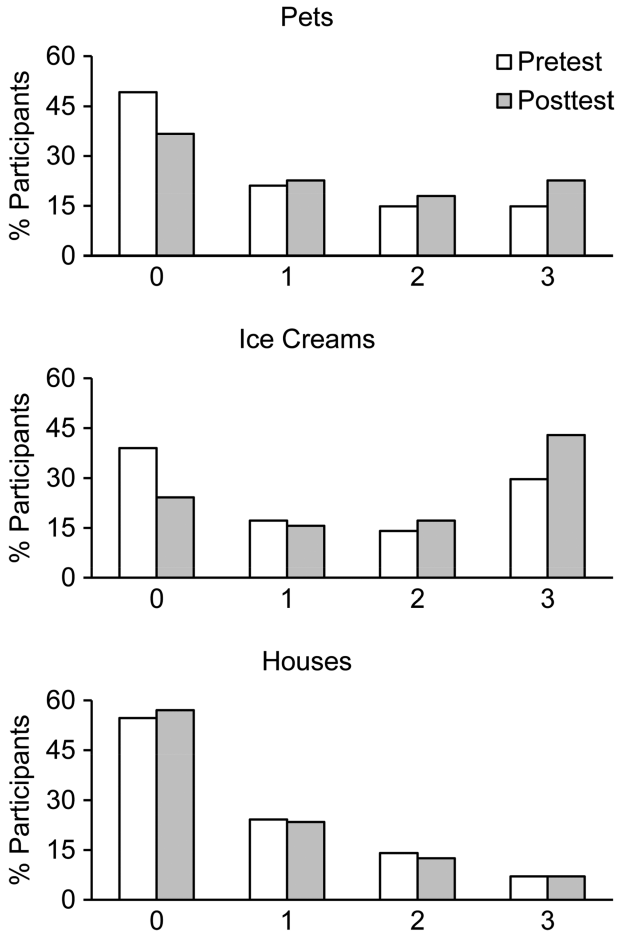
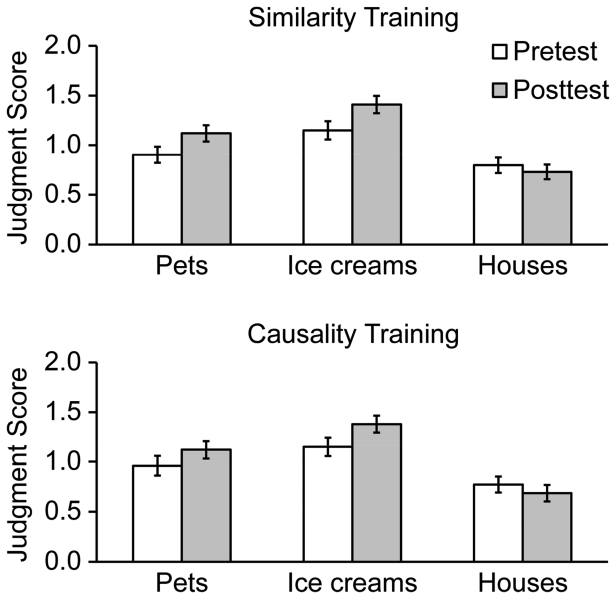


Figure 2

Mean Possibility Judgment Score (and Standard Error) for Improbable Events in Each Event Category Before and After Each Type of Training (Which Covered Only Pets and Ice Creams)



both variables as covariates in repeated-measures ANOVAs analyzing the effect of assessment period (pretest, posttest) on possibility-judgment scores for just the improbable events. Age was a significant predictor for the trained categories of pets, $F(1, 125) = 25.13$, $p < .001$, $\eta_p^2 = .167$ and ice creams, $F(1, 125) = 26.28$, $p < .001$, $\eta_p^2 = .174$, but not the untrained category of houses, $F(1, 125) = 3.06$, $p = .083$, $\eta_p^2 = .174$. Cognitive reflection was a significant predictor for all three—pets: $F(1, 125) = 4.79$, $p = .03$, $\eta_p^2 = .024$; ice creams: $F(1, 125) = 5.09$, $p = .026$, $\eta_p^2 = .039$; houses: $F(1, 125) = 15.64$, $p < .001$, $\eta_p^2 = .011$.

Critically, however, main effects were the only effects observed; neither age nor cognitive reflection interacted with assessment period for any of the event categories, indicating that participants of all ages and CRT-D scores benefited equally from training. The absence of interaction effects for the untrained category of houses suggests that neither age nor cognitive reflection facilitated transfer either. That said, age and CRT-D scores were both significant predictors of possibility judgment in general, as shown in Table 4. Age predicted possibility judgment scores for improbable events in all event categories at both assessment periods, as did CRT-D scores. Moreover, four of the six correlations between CRT-D scores and possibility judgment scores remained significant after controlling for age, indicating that cognitive reflection was a unique predictor of children's judgments.

Age and CRT-D scores explained much of the variance in judgments, but another source of variance was participants' overall disposition to affirm the possibility of improbable events. This disposition can be seen from the within-participant correlations displayed in Table 5. Across event categories and assessment periods, some participants were more inclined to affirm the possibility of improbable events than were others. All 15 correlations were significant, averaging .47 in magnitude, and 13 of those correlations remained significant even when controlling for age and CRT-D scores, indicating that individual differences in possibility judgment are not wholly reducible to age or reflectiveness.

Discussion

Young children deny the possibility of improbable events, though recent research suggests that they can be induced to accept such events if told of a similar event that has actually occurred or a causal pathway by which the event might occur (Goulding et al., 2022; Goulding & Friedman, 2021). Here, we investigated whether one of these considerations—empirical precedents or causal mechanisms—is more useful in boosting children's acceptance of improbable events, whether these considerations help children reason only about closely related events or about possibility in general, and whether their ability to use such considerations increases with age and cognitive reflection. We found that precedents and mechanisms were equally useful, when framed to address the same improbable events, but only for closely related events. Children did not generalize the use of such considerations to a novel category of improbable events, regardless of their age or cognitive reflectiveness.

Table 4
Correlations Between Possibility Judgment Scores for Improbable Events and Age, CRT-D Scores, and CRT-D Scores Controlling for Age

Assessment	Event category	Age	CRT-D	CRT-D controlling for age
Pretest	Pets	.43***	.37***	.19*
	Ice creams	.49***	.35***	.13
	Houses	.29***	.37***	.27**
Posttest	Pets	.54***	.38***	.14
	Ice creams	.51***	.43***	.22*
	Houses	.34***	.43***	.32***

Note. CRT-D = Cognitive Reflection Test, developmental version.
* $p < .05$. ** $p < .01$. *** $p < .001$.

These findings highlight the value of precedents and mechanisms in expanding children's notion of what is possible while also underscoring the robustness of their skepticism toward expectation-violating events. On one hand, both types of considerations improved children's ability to affirm the possibility of improbable events without also leading them to affirm the possibility of truly impossible ones, like owning a pet unicorn or making lava-flavored ice cream. If anything, these considerations made children more cautious when reasoning about impossible events, as revealed by the pre–post decrease in possibility-judgment scores for owning impossible pets. Furthermore, both trainings proved equally effective at increasing children's acceptance of improbable events, indicating that the kinds of considerations that shape adults' sense of possibility (Goulding & Friedman, 2023; Shtulman & Tong, 2013) also shape children's, provided they are aware of them.

On the other hand, neither type of consideration increased children's acceptance of improbable events that were not explicitly addressed during training. Children appeared to use the precedents and mechanisms provided to them, but they were not compelled to identify such considerations on their own when reasoning about an untrained category. And they used those considerations only modestly when reasoning about events for which they are directly relevant, boosting their acceptance of one or two improbable events in the trained categories rather than all such events.

Our study clarifies the role of similarity and causality in early modal reasoning. We discovered that an intervention that teaches children how to reason about possibility as a domain-general inference will have limited practical value if not accompanied by domain-specific information. Children can use similarity to affirm unexpected possibilities but only when connected to specific precedents, and they can use causality but only when connected to specific mechanisms. While we expected that older children and cognitively reflective children might be able to recruit this information on their own from prior knowledge, we found that they too exhibited little to no transfer. These results indicate that reasoning about possibility is a knowledge-demanding process and that expanding children's sense of possibility requires expanding their knowledge. Modeling possibility-judgment strategies in the abstract will not work.

One concern, though, is that using houses as a transfer category may have hampered children's ability to demonstrate abstract, domain-general learning. Perhaps children could have learned the strategies modeled during training but were unable to apply them to the untrained category because they did not know of any relevant precedents or mechanisms, despite having learned to search for them. This may be a concern for our youngest participants but is less plausible for our older ones, in their last years of elementary school. These children would have heard about (or seen) many atypical houses, including igloos, teepees, mud huts, and glass towers. They would also have had ample experience building houses of their own, including Lego houses, gingerbread houses, houses of cards, and pillow forts. In fact, 45% of children accepted the possibility of at least one improbable house at pretest, implying they knew some precedent or mechanism relevant to this category, and pretest performance for houses was statistically similar to that of pets, $M_{\text{difference}} = 0.2$, $t(127) = 1.92$, $p > .05$ (see Figure 1). Still, future research could directly disentangle the transferability of training from the content of the transfer category by systematically varying which categories are used in training and which are used to assess transfer.

Further reason to doubt that children learned general strategies but had trouble applying them is that their acceptance of improbable events in the trained categories was far from ceiling. Prior to training, children claimed that events in the trained categories were possible 38% of the time; following training, they did so 51% of the time—an increase of 13%. Training changed children's mind about a few improbable pets and a few improbable ice creams but not many. Their insistence that most of the events in the trained categories are impossible (kinda or very) implies that training improved their judgments quantitatively but not qualitatively. Children appeared to change their mind when they saw connections between the target events and the precedents or mechanisms provided to them but did not heed the general advice of identifying such

Table 5
Correlations Among Possibility Judgments Scores for Improbable Events in Each Event Category at Each Assessment Period

No	Assessment	Event category	1	2	3	4	5	6
1	Pretest	Pets	—	.54***	.39***	.64***	.46***	.43***
2	Pretest	Ice creams	.41***	—	.37***	.54***	.69***	.31***
3	Pretest	Houses	.27**	.25**	—	.34***	.29**	.60***
4	Posttest	Pets	.52***	.36***	.19*	—	.58***	.44***
5	Posttest	Ice creams	.28**	.57***	.11	.39***	—	.38***
6	Posttest	Houses	.29**	.14	.52***	.30***	.20*	—

Note. Zero-order correlations are displayed above the diagonal and partial correlations controlling for age and CRT-D scores are displayed below. CRT-D = Cognitive Reflection Test, developmental version.
* $p < .05$. ** $p < .01$. *** $p < .001$.

considerations, as a test of their intuitions. It is possible that asking children to evaluate improbable events twice in the same experiment limited their learning insofar as their judgments at pretest established a pattern they felt compelled to maintain at posttest, but the specificity of the training effects speaks against this possibility, as children's judgments improved in some categories but not others.

To our surprise, age and cognitive reflection did not predict learning or transfer. They did, however, predict baseline acceptance of improbable events, as shown in Table 4. Older children were more likely to accept improbable events than younger children, presumably because older children know more precedents and mechanisms relevant to these events. And children who scored high on the CRT-D were more accepting of improbable events than those who scored low, presumably because reflection leads children to question their immediate intuitions. Indeed, cognitive reflection predicted possibility judgments above and beyond age, possibly because the knowledge that comes with age is most useful when actively reflected upon. Although we did not measure reflection in the moment (with response times or judgment justifications), the correspondence between CRT-D scores and the acceptance of improbable events suggests that reasoning about possibility involves the same tradeoff between intuition and analysis that is required for solving brainteasers like "What do cows drink?" (see Shtulman et al., 2023, for similar findings).

One reason age and cognitive reflection did not facilitate learning is that older children and cognitively reflective children may have already been using the strategies modeled for them. In this same vein, we observed robust individual differences that held regardless of training (see Table 5). These differences were not reducible to age or cognitive reflection, suggesting they represent baseline differences in knowledge independent of age, baseline differences in strategy use independent of reflection, or baseline differences in some other skill or disposition, such as openness to the unexpected or optimism about what is achievable in the future.

The individual differences we observed among children may be precursors to individual differences observed among adults. Research on adults' possibility judgments suggests that only some adults seek out precedents and mechanisms when reasoning about extraordinary events, such as traveling to another galaxy or bringing an extinct species back to life; others focus on facts or principles that may preclude the events from occurring (Shtulman & Tong, 2013). Both approaches could, in principle, converge on the same judgments, but adults who focus on facts and principles judge fewer extraordinary events possible than those who focus on precedents and mechanisms. The development of modal cognition appears to have more than one endpoint, even if adults, on the whole, are more accepting of possibilities that violate their expectations (Shtulman & Carey, 2007; Shtulman & Phillips, 2018).

Conclusion

Reasoning about possibility is critical to problem solving and decision making, but this skill develops slowly. Young children's judgments of what could occur are unduly influenced by their expectations of what would occur or should occur. These judgments can be swayed by prompting children to consider empirical precedents and causal mechanisms, which help them bridge the gap between unexpected events and ordinary events, but the impact of such prompting is limited in scope. Children can use precedents and mechanisms to affirm the possibility of closely related events but do not appear to identify precedents and mechanisms when contemplating novel improbabilities. While adults affirm many more improbable events than children, it remains an open question whether this developmental change requires explicit reflection on possibility or is driven by increased knowledge of the precedents and mechanisms relevant to specific events.

References

Bowman-Smith, C. K., Shtulman, A., & Friedman, O. (2019). Distant lands make for distant possibilities: Children view improbable events as more possible in far-away locations. *Developmental Psychology*, *55*(4), 722–728. https://doi.org/10.1037/dev0000661

Byrnes, J. P., & Duff, M. A. (1989). Young children's comprehension of modal expressions. *Cognitive Development*, *4*(4), 369–387. https://doi.org/10.1016/S0885-2014(89)90049-X

Chernyak, N., Kushnir, T., Sullivan, K. M., & Wang, Q. (2013). A comparison of American and Nepalese children's concepts of freedom of choice and social constraint. *Cognitive Science*, *37*(7), 1343–1355. https://doi.org/10.1111/cogs.2013.37.issue-7

Cook, C., & Sobel, D. M. (2011). Children's beliefs about the fantasy/reality status of hypothesized machines. *Developmental Science*, *14*(1), 1–8. https://doi.org/10.1111/j.1467-7687.2009.00949.x

Danovitch, J. H., & Lane, J. D. (2020). Children's belief in purported events: When claims reference hearsay, books, or the internet. *Journal of Experimental Child Psychology*, *193*, Article 104808. https://doi.org/10.1016/j.jecp.2020.104808

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Goulding, B. W., & Friedman, O. (2020). Children's beliefs about possibility differ across dreams, stories, and reality. *Child Development*, *91*(6), 1843–1853. https://doi.org/10.1111/cdev.13386

Goulding, B. W., & Friedman, O. (2021). A similarity heuristic in children's possibility judgments. *Child Development*, *92*(2), 662–671. https://doi.org/10.1111/cdev.13534

Goulding, B. W., & Friedman, O. (2023). Perceived similarity explains beliefs about possibility. *Journal of Experimental Psychology: General*, *152*(10), 2830–2841. https://doi.org/10.1037/xge0001422

Goulding, B. W., Stonehouse, E. E., & Friedman, O. (2022). Causal knowledge and children's possibility judgments. *Child Development*, *93*(3), 794–803. https://doi.org/10.1111/cdev.13718

Lane, J. D., Ronfard, S., & El-Sherif, D. (2018). The influence of first-hand testimony and hearsay on children's belief in the improbable. *Child Development*, *89*(4), 1133–1140. https://doi.org/10.1111/cdev.12815

Lane, J. D., Ronfard, S., Francioli, S. P., & Harris, P. L. (2016). Children's imagination and belief: Prone to flights of fancy or grounded in reality? *Cognition*, *152*, 127–140. https://doi.org/10.1016/j.cognition.2016.03.022

Nancekivell, S. E., & Friedman, O. (2017). She bought the unicorn from the pet store: Six-to seven-year-olds are strongly inclined to generate natural explanations. *Developmental Psychology*, *53*(6), 1079–1087. https://doi.org/10.1037/dev0000311

Nissel, J., Li, H., Cramer, A., & Woolley, J. D. (2023). Three men make a tiger: The effect of consensus testimony on Chinese and U. S. children's judgments about possibility. *Journal of Cognition and Culture*, *23*(1), 98–126. https://doi.org/10.1163/15685373-12340154

Nolan-Reyes, C., Callanan, M. A., & Haigh, K. A. (2016). Practicing possibilities: Parents' explanations of unusual events and children's possibility thinking. *Journal of Cognition and Development*, *17*(3), 378–395. https://doi.org/10.1080/15248372.2014.963224

Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, *11*(3), 191–214. https://doi.org/10.1080/15475441.2014.905169

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007

Shtulman, A. (2009). The development of possibility judgment within and across domains. *Cognitive Development*, *24*(3), 293–309. https://doi.org/10.1016/j.cogdev.2008.12.006

Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development*, *78*(3), 1015–1032. https://doi.org/10.1111/j.1467-8624.2007.01047.x

Shtulman, A., Harrington, C., Hetzel, C., Kim, J., Palumbo, C., & Rountree-Shtulman, T. (2023). Could it? Should it? Cognitive reflection facilitates children's reasoning about possibility and permissibility. *Journal of Experimental Child Psychology*, *235*, Article 105727. https://doi.org/10.1016/j.jecp.2023.105727

Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology*, *165*, 161–182. https://doi.org/10.1016/j.jecp.2017.05.012

Shtulman, A., & Tong, L. (2013). Cognitive parallels between moral judgment and modal judgment. *Psychonomic Bulletin & Review*, *20*(6) 1327–1335. https://doi.org/10.3758/s13423-013-0429-9

Shtulman, A., & Young, A. G. (2023). The development of cognitive reflection. *Child Development Perspectives*, *17*(1), 59–66. https://doi.org/10.1111/cdep.12476

Weisberg, D. S., & Sobel, D. M. (2012). Young children discriminate improbable from impossible events in fiction. *Cognitive Development*, *27*(1), 90–98. https://doi.org/10.1016/j.cogdev.2011.08.001

Williams, A. J., & Danovitch, J. H. (2022). Is what Mickey Mouse says impossible? Informant reality status and children's beliefs in extraordinary events. *Journal of Cognition and Development*, *23*(3), 323–339. https://doi.org/10.1080/15248372.2021.2022680

Young, A., & Shtulman, A. G. (2020). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychological Science*, *31*(11), 1396–1408. https://doi.org/10.1177/0956797620954449

(*Appendix follows*)

Appendix

Training Protocols

Similarity Training

Let's talk more about how to decide whether something is possible or impossible. A good way to decide is to think of something similar that you know is real.

Here's a crocodile. If you were trying to decide whether a person could own a pet crocodile in real life, you could think of other animals that people own for real. What are some animals that people own for real?

That's right! But a crocodile doesn't seem very similar to those animals, does it? Can you think of some pets that are more unusual? (Good examples! or It's hard, isn't it!)

You know what, some people own elephants as pets. Some people own tortoises as pets. And some people own pythons as pets. So now that we've thought of some pets that are more similar to a crocodile, it seems possible a person could own a crocodile for a pet in real life, right?

Let's think about whether something else is possible. Here are some onions. Let's think about whether a person could make onion-flavored ice cream in real life. What are some flavors of ice cream that people make in real life? Can you think of some flavors that are more unusual? (Good examples! or It's hard, isn't it!)

You know what, some people actually make bean-flavored ice cream. Some people make broccoli-flavored ice cream. And some people make spinach-flavored ice cream. So now that we've thought of some ice cream flavors that are similar to onion, it seems possible someone could make onion-flavored ice cream in real life, right?

So remember: A good way to decide whether something is possible is to try to think of something similar that you know is real.

Causality Training

Let's talk more about how to decide whether something is possible or impossible. A good way to decide is to think of how it could happen.

Here's a crocodile. If you were trying to decide whether a person could own a pet crocodile in real life, you could think of how they might do it. What are some things a person would need to keep a crocodile as a pet?

(Good ideas! or It's hard, isn't it!) You would need a place for the crocodile to live. You wouldn't want to keep it in your house, but you could keep it in a pond or a pool in the backyard. You would also need to feed the crocodile. You wouldn't want to feed it cereal, but you could feed it fish or porkchops. Where might a person get a crocodile?

And you'd have to get a crocodile that you could keep. You might be able to get one from the zoo or find one in the wild.

So now that we've figured out how someone could find and care for a crocodile, it seems possible a person could own a crocodile for a pet in real life, right?

Let's think about whether something else is possible. Here are some onions. Let's think about whether a person could make onion-flavored ice cream in real life. What would a person need to make onion-flavored ice cream?

(Good ideas! or I know, it's hard!) You would need to make regular ice cream first, so you'd need some milk. You'd also need some sugar. Then you would add onions. You could buy onions at the store, chop them into little pieces, and then stir the pieces into the ice cream mix. It might not taste very good, but you could make it.

So now that we've figured out how someone could add onions to ice cream, it seems possible a person could make onion-flavored ice cream in real life, right?

So remember: A good way to decide whether something is possible is to try to think of how it could happen.

Received April 18, 2023
Revision received August 14, 2023
Accepted October 6, 2023 ■