AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 1, 184–199
https://doi.org/10.1037/xge0001494

The Development of Modal Intuitions: A Test of Two Accounts

Brandon W. Goulding[1], Farishteh Khan[2], Keisuke Fukuda[2], Jonathan D. Lane[3], and Samuel Ronfard[2]

[1] Department of Psychology, University of Winnipeg
[2] Department of Psychology, University of Toronto
[3] Department of Psychology & Human Development, Vanderbilt University

Young children, unlike adults, deny that improbable events can happen. We test two accounts explaining this developmental shift. The development = reflection account posits that this shift is driven by an emerging ability to reflect on modal intuitions. In contrast, the development = intuition account posits that this shift is driven by changes in modal intuitions themselves, due to age-related changes in what people know and how they sample their knowledge and memories. These accounts make competing predictions about how long children and adults should take to make possibility judgments. In Experiment 1, we asked 123 children (39 5-year-olds, 42 7-year-olds, 42 9-year-olds; 49.60% White) and 40 adults (50% White) to judge the possibility of 78 ordinary, improbable, and impossible events and recorded their response times. In Experiment 2, we tested an additional 52 adults (42.32% White) who were under speeded conditions and thus less able to reflect before responding. Our results favor the development = intuition account. At all ages, people judged improbable events more slowly than ordinary or impossible events, and slow responding did not consistently predict affirmation over denial. Further, adults' possibility judgments did not change under speeded conditions. We also fit a drift-diffusion model to our data, which suggested that adults and children may sample different kinds of knowledge when generating intuitions. Our findings suggest that possibility judgments are often driven by modal intuitions with little reflection, and that a developmental shift in what children know and how knowledge is retrieved can explain why these intuitions change over time.

Public Significance Statement

Judging possibility is a core cognitive ability that interacts with many, if not all, domains of thought. We use our beliefs about what is possible to understand and infer what will happen in the distant future, what has happened in the past, what we are seeing right before our eyes, and all things in between. This work explored the development of how quickly these judgments are made, which allowed for a test of whether judgments about possibility become slower and more deliberative with age. Children are often framed as unsophisticated thinkers, employing frugal reasoning strategies which lead them to incorrect answers. Conversely, adults are thought to be more careful and deliberative in their thinking. This work suggests that children and adults use the same strategy to infer what is possible, and further suggests that this strategy is fast and frugal at all ages. The findings suggest that people's beliefs about what can happen are typically driven by quick intuitions rather than reflective reasoning, and that these intuitions are shaped by knowledge and memory retrieval patterns that change with age. This research highlights the importance of studying whether cognitive development can often be explained by changes in knowledge, as opposed to a developmental shift in reasoning strategies. Further, this work suggests that people's explanations may not always capture their reasoning. People often justify their belief that an event is possible by explaining how it might happen, and children and adults often offer different kinds of explanations. This work suggests that people likely form an intuition about whether something is possible before they can think about how it could occur. This suggests that people's explanations for why they believe a phenomenon to be possible may sometimes be post hoc constructions.

Keywords: possibility, modal reasoning, intuition, reflection, development

This article was published Online First October 16, 2023.

Brandon W. Goulding https://orcid.org/0000-0002-4176-4993

We thank the parents and children for their participation and the amazing research assistants of the Childhood Learning and Development Lab at the University of Toronto Mississauga. This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to Samuel Ronfard (RGPIN-2020-04842) and a University of Toronto Mississauga Postdoctoral Fellowship to Brandon W. Goulding. We have no conflicts of interest to disclose. The supplemental materials for this article, as well as materials, data, and analysis scripts for all experiments, are available on the Open Science Framework (OSF) page at https://osf.io/5pdjf/.

Brandon W. Goulding and Farishteh Khan contributed equally to this work and shared first authorship. Brandon W. Goulding served as lead for formal analysis and writing–original draft. Farishteh Khan served as lead for project administration and served in a supporting role for formal analysis, investigation, methodology, and writing–review and editing. Keisuke Fukuda served as lead for drift-diffusion model analysis and write-up and served in a supporting role for writing–review and editing. Jonathan D. Lane served in a supporting role for funding acquisition, investigation, methodology, and writing–review and editing. Samuel Ronfard served as lead for funding acquisition, resources, and supervision and served in a supporting role for formal analysis, project administration, and writing–original draft. Brandon W. Goulding and Farishteh Khan contributed equally to data curation. Brandon W. Goulding and Samuel Ronfard contributed equally to methodology, writing–review and editing, and investigation. Jonathan D. Lane and Samuel Ronfard contributed equally to conceptualization.

Correspondence concerning this article should be addressed to Brandon W. Goulding, Department of Psychology, University of Winnipeg, 515 Portage Avenue, Winnipeg, MB R3B 2E9, Canada or Samuel Ronfard, Department of Psychology, University of Toronto, 3359 Mississauga Road, CCT Building, Room 4059, Mississauga, ON L5L 1C6, Canada. Email: b.goulding@uwinnipeg.ca or samuel.ronfard@utoronto.ca

Germs. Evolution. God. Terraforming. Time travel. Purple rabbits. Most people have never seen these phenomena and entities firsthand. Despite this fact, we can reason about the likelihood of their existence. By drawing upon our prior experiences and our intuitions about how the world works, we can determine whether these things are probable, plausible, or impossible. This kind of modal reasoning, where people judge whether actions and events are possible or permissible, is regularly deployed across domains. For example, people can make these distinctions when reasoning about other people, animals, physical phenomena, and supernatural entities. It is a core aspect of human cognition that impacts our decisions to trust what we are told, what we explore, and how we invest our time and efforts.

Even nonverbal infants have strong intuitions about what events are likely to occur, and they react with surprise when those intuitions are violated (Spelke & Kinzler, 2007). Three-year-olds can articulate their modal intuitions about the likelihood that events will occur—for example, they judge that making an object disappear is impossible but changing the color of an object is not (Johnson & Harris, 1994). Intriguingly, preschool-age children are especially conservative in their reasoning about what can exist or occur, often reporting that obscure (but possible) events cannot occur and that obscure (but real) entities cannot exist (Lane & Harris, 2014; Woolley & Ghossainy, 2013). For example, they report that it is impossible to find an alligator under a bed, make pickle-flavored ice cream, grow a beard down to one's toe, or walk on a telephone wire (Shtulman, 2009; Shtulman & Carey, 2007). Older children are more likely to report that these events can indeed occur, and adults are even more inclined to do so. This pattern holds both within and across the physical, psychological, and biological domains (Shtulman, 2009) and has been replicated using a diverse set of items by multiple labs (Cook & Sobel, 2011; Goulding & Friedman, 2020; Lane et al., 2016; Nancekivell & Friedman, 2017; Nolan-Reyes et al., 2016; Weisberg & Sobel, 2012).

We explore two accounts for these developmental differences. The first account posits that age-related differences in possibility judgments are driven by an emerging ability to reflect on modal intuitions before judging possibility (development = reflection, e.g., Shtulman, 2009). In contrast, the second (and novel) account posits that these age-related differences in possibility judgments are driven by changes in modal intuitions themselves, as a result of age-related changes in what people know and how individuals search their knowledge base to generate these intuitions (development = intuition). Below, we further describe each account and outline how we test for developmental patterns that are either consistent or inconsistent with each account.

Development = Reflection

Shtulman (2009) proposed a cognitive model for how children and adults reason about the possibility of events. This model includes two potential steps: (a) people form a modal intuition about whether an event is possible and (b) people reflect on this intuition before judging possibility. A version of this model is depicted in Figure 1. In this model, intuitions are assumed to be formed quickly and automatically. How these intuitions are generated is unclear yet important (we address this in the next section). Reflection, however, is framed as an active reasoning process where a person explicitly considers whether their automatic intuition is justified or correct. On Shtulman's (2009) account, the development of possibility judgments is driven by an increased tendency to reflect on, or reason about, modal intuitions. Importantly, this account proposes that children and adults usually form the same intuitions about events, but only adults take the time to reflect on whether their intuition is correct. This is especially impactful when reasoning about events that are strange or unfamiliar. For instance, both children and adults might share the intuition that a person having a pet lion is impossible, but only adults will take the time to reflect on this intuition and realize that it is incorrect, and thereby judge that a person really could obtain a pet lion if they wanted—despite this being an unorthodox and perhaps ill-advised decision. This tendency to reflect is hypothesized to increase with age and to explain developmental increases in judging improbable events as possible.

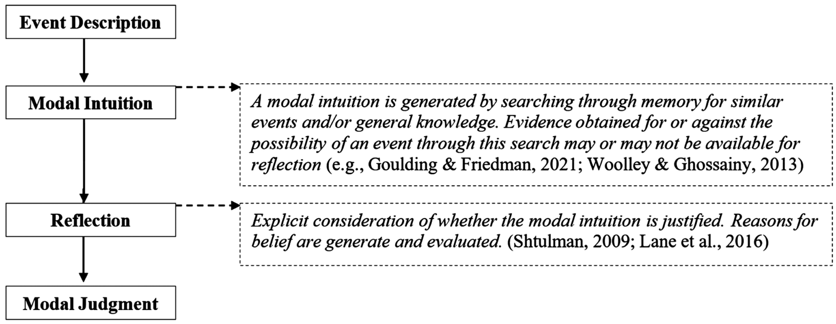
This development = reflection explanation is only one of several accounts positing that reasoning about possibility involves a two-stage process, whereas in young children it is initially restricted to one stage of processing (Harris, 2021; Lane et al., 2016; Shtulman & Carey, 2007; Woolley & Ghossainy, 2013). On this view, we should expect that reflection is most often engaged when reasoning about improbable events, because such events are unfamiliar but cannot be dismissed on principle. Older children and adults may reflect on the possibility of improbable events—and ultimately affirm them—by simulating how the events could come to be (Shtulman & Carey, 2007), or by deciding that the event does not violate a known causal constraint (e.g., solid objects cannot pass through each other; see Harris, 2021). If, according to this two-stage process, young children are not inclined to reflect on their intuitions, then young children's judgments of event possibility should require no additional time to reason about improbable events relative to ordinary or impossible events. Older children and adults—who are more inclined to engage in reflection—should reach different judgments of events' possibility depending upon the amount of time they are given to reflect on the events, and this should be especially true of improbable events.

Development = Intuition

According to our novel development = intuition account, what develops is not children's ability or tendency to reflect on, or reason about, their modal intuitions; it is the intuitions themselves which are changing. Children form incorrect intuitions about improbable events, and deny them; adults form correct intuitions, and affirm them. On this view, we do not require a second reflective step to explain the development of possibility judgments. Instead, this developmental shift may be a result of changes in the process through which such intuitions are generated, which we explore below. People may generate a modal intuition about whether an event could occur by sampling their knowledge (automatically, perhaps nonconsciously), and their eventual modal judgment may largely reflect the output of this sampling process rather than the output of a secondary deliberate reflective process. In other words, much of the developmental shift in modal judgments could be explained by changes in the first stage of generating a possibility judgment, rather than the second (see Figure 1). To clarify, the ability to reflect, the tendency to reflect, and the ways in which we reflect may still contribute to judgments of possibility (Shtulman et al., 2023). However, this account proposes that the well-documented developmental shift in people's modal judgments is not a function

Figure 1

How Modal Intuitions and Reflection Might Contribute to Judgments of Possibility



Note. This model is adapted from "The Development of Possibility Judgment Within and Across Domains," by A. Shtulman, 2009, *Cognitive Development, 24*(3), pp. 293–309 (https://doi.org/10.1016/j.cogdev.2008.12.006). Copyright 2009 by Elsevier. Adapted with permission.

of developmental shifts in reflection so much as it is a function of developmental shifts in modal intuitions.

When people are asked whether an event could happen, they may sample and retrieve relevant knowledge and memories to arrive at an intuition. This process should be quick for ordinary events, which are familiar and easily accessed, and for impossible events, which violate well-known causal laws and have no close counterpart to the events people actually experience. But it may proceed more slowly for improbable events, because it often takes more time to retrieve information about unfamiliar than familiar things (e.g., McCloskey, 1980). When generating a modal intuition for improbable events, people may sample their knowledge of how events could occur, or their memories of similar things that have happened (Goulding et al., 2022; Shtulman & Carey, 2007; Shtulman & Tong, 2013). For instance, when asked whether a person could walk on a telephone wire, a person could automatically retrieve semantic knowledge about tightrope walkers, or an episodic memory of a person performing at a circus, and then quickly form a modal intuition that walking on a telephone wire could occur. This sampling begins while forming the intuition itself, rather than during a second stage of processing where the intuition is reflected upon.

Children and adults both necessarily sample their knowledge to infer possibility, but older children and adults should have a more robust knowledge base by virtue of having more semantic knowledge and episodic memories to draw upon when judging what is possible. Adults and older children also have better executive abilities than younger children, which should facilitate improved knowledge and memory retrieval while generating intuitions (Hjuler et al., 2023; Krøjgaard et al., 2022; Levy & Anderson, 2002). In sum, according to this account, changes in one's knowledge base and changes in retrieving that knowledge may largely account for developmental changes in children's modal judgments.

Testing Both Accounts Using Reaction Time (RT) Data

Here, we seek direct evidence for the development = reflection account and the development = intuition account by measuring how long children and adults take to affirm or deny events. Each account makes different predictions about how quickly children and adults should respond to questions about the possibility of ordinary events, impossible events, and improbable events, and how their response times for these events should vary with age. As a reminder: children usually reject improbable events, while adults usually affirm them (e.g., Shtulman & Carey, 2007). If reflection is the primary driver behind the developmental shift in judgments of improbable events, then we should find that adults take longer to affirm improbable events than ordinary and impossible events, and that children (who engage in little reflection) affirm all kinds of events at similar speeds. We expect this because reflective responding often takes longer than responding based on mere intuition (Shtulman, 2017). For instance, adults take longer to affirm well-known facts that compete with their intuitions, such as plants being alive (Goldberg & Thompson-Schill, 2009) or the Earth revolving around the Sun (Shtulman & Harrington, 2016). We may observe similar patterns for judgments of improbable events.

Conversely, if development is better explained by differences in how intuitions are generated, then we should not observe increased response times for improbable events with age. Indeed, we may even observe a decrease in response time, since adults may be better at retrieving relevant memories and knowledge (e.g., Levy & Anderson, 2002). We test these competing predictions in Experiment 1 by measuring the speed at which children and adults judge the possibility of ordinary, improbable, and impossible events.

Further, adults often reveal their intuitions when responding under time pressure, without the opportunity to reflect (Shtulman, 2017). For instance, placing people in speeded conditions has led to physicists more often endorsing teleological explanations (Kelemen et al., 2013) and lay adults more often denying that immoral actions are possible (Phillips & Cushman, 2017). We, therefore, anticipate that asking participants to make speeded judgments should further reveal whether they need to reflect on the possibility of improbable events to affirm them. We test this in Experiment 2 by forcing adults to make speeded judgments about the possibility of ordinary, improbable, and impossible events.

Experiment 1

Method

Participants

We tested 123 children: 39 5-year-olds ($M_{\text{age}} = 5;7$ [years; months], range = 5;0–5;11, 20 girls), 42 7-year-olds ($M_{\text{age}} = 7;5$, range =

7;0–7;11, 23 girls), 42 9-year-olds (M_{age} = 9;6, range = 9;0–9;11, 18 girls). These children were part of a shared participant database at the University of Toronto Mississauga and were recruited via emails to parents. We also tested 40 Canadian and American adults (M_{age} = 25 years, range = 18–30 years, 21 female, 17 male, two identified as other or preferred not to answer). An additional 32 children were tested but excluded due to parental interference, failing to complete the experiment, switching to a tablet or phone during the session, or taking too long to complete the study (>3 *SD* above the mean duration). We also tested two additional adults but excluded them from analysis for taking an especially long time to complete the study—more than 2 *SD* above the mean duration or beyond the 30-min time limit.

Our child sample was 49.60% White, 24.40% mixed race, 6.50% South Asian, 5.70% Chinese, 2.43% South-East Asian, 2.43% Black, and 1.62% other; 7.32% declined to respond. Children's household income was greater than $90,000 CAD for 65.04% of participants, and 83.60% of children had at least one parent who attended university. Our adult sample was 50% White, 20% Chinese, 7.5% mixed race, 7.5% South Asian, 5% Arab, 5% Filipino, 2.5% Latin American, and 2.5% Black.

Children were tested online via Zoom and were guided through the call by an experimenter; adults were tested online via Prolific and completed the experiment on their own. Adults were paid £2.91 GBP (∼5.00 CAD) for their participation.

Stimuli

We created 26 homologous triads of events for a total of 78 testing items. These items were either borrowed from or inspired by previous work on beliefs about possibility (Browne & Woolley, 2004; Cook & Sobel, 2011; Lane et al., 2016; Nolan-Reyes et al., 2016; Shtulman, 2009; Shtulman & Carey, 2007; Shtulman & Phillips, 2018; Weisberg & Sobel, 2012) and spanned multiple domains of reasoning (i.e., physical, biological, social, conventional, psychological). For our purposes, ordinary events were meant to be familiar to participants and not violate any apparent physical, biological, or psychological constraints; improbable events were designed to be unfamiliar, strange, or unconventional, but without any apparent physical, biological, or psychological constraints on their occurrence; and impossible events were potentially constrained by (known) physical, biological, or psychological limitations and perhaps because of this have never occurred (to our knowledge). Each triad consisted of an ordinary event, improbable event, and impossible event, where the root action was a uniform event (e.g., staying awake for *x* amount of time), but its probability of occurrence varied (*x* could be 5 hr, 5 days, or 50 days). Events within each triad were also matched for duration (mean duration across triads = 1,750 ms, mean *SD* within triads = 83 ms), and a univariate analysis of variance (ANOVA) confirmed that average duration did not differ across event types, $p = .903$.

We also created five nonhomologous training items (three ordinary and two impossible) for the training block administered prior to testing. Improbable items were not included in the training block to allow participants to calibrate their answers, and to ensure their ability to distinguish between the clearly possible and the clearly impossible.

Procedure

The experiment was conducted using the Gorilla platform (www.gorilla.sc). We selected the Gorilla platform because it provides near millisecond precision on several browser/operating system combinations (Bridges et al., 2020) and is thus well suited for studies measuring participants' RTs.

Prior to viewing the test trials, all participants underwent a brief training phase consisting of five training items. These items included common, possible events (e.g., "Could someone pet a cat?") and impossible events (e.g., "Could someone jump up and touch the moon?"). This ensured that participants understood what "possible" and "impossible" meant. It also served to familiarize them with the response buttons.

Participants then listened to a series of ordinary, improbable, and impossible events and were asked whether they thought each event was possible or impossible. The events were presented one-at-a-time. On each test trial, participants first saw a screen saying "Ready?" and were prompted to proceed by clicking a red "Next" button. They then saw a blank white screen and heard an audio recording of the static prompt ("Could someone…"), followed by an audio recording of one of the events (e.g., "…stay awake for 5 days?"). When the audio clip ended, a green "Yes" button and a red "No" button appeared in the center of the screen. Participants clicked "Yes" or "No" to indicate whether they thought the event could happen; we also recorded the time taken to click one of these buttons after they appeared. After responding, the "Ready?" screen with the "Next" button appeared again to ensure that participants placed their mouse at the bottom center of the screen at the start of each trial.

We randomly divided the 78 events into three blocks of 26 items each, such that each block had exactly one item from each homologous triad. Each block also had a balanced number of events of each type (ordinary, improbable, impossible). The presentation of the testing blocks was randomly generated to appear in one of three possible orders (ABC, BCA, and CAB). Both the order of the testing blocks and the order of the items within each testing block were randomized. The buttons and the static audio prompt ("Could someone…") were the same on each trial; only the audio differed. Trials were presented in a random order, and all participants completed all trials. There was no option to skip a trial to proceed. We also included a 2-min break at the end of each testing block.

For the adult participants, who completed the study without a researcher present, we also included implicit and explicit attention checks to ensure that participants were attending to the study and answering to the best of their ability. There were four implicit attention checks (i.e., checks that participants were not made aware of) which were formatted identically to testing items but were not part of a triad, and which always appeared at the end of each block. They were designed to be judged possible (e.g., "Could someone become sick?") or impossible (e.g., "Could someone become younger?") by all participants who were responding attentively; participants were to be excluded if they failed to respond correctly. There were also six explicit attention checks. These were made readily apparent to the participants, and participants were informed that they would be excluded if they failed these checks too often. On these trials, participants heard "Attention Check!," followed by either "Please click 'Yes'" or "Please click 'No.'" Two explicit checks were included in each test block. Participants who answered fewer than four of six of explicit attention checks correctly would have been excluded. No participant failed either the implicit or explicit checks, so none were excluded on these bases.

Transparency and Openness

The stimuli, data, and analysis scripts for our experiments are available on OSF at https://osf.io/5pdjf/. We have also included a document where we report the mean judgments and RTs for all events across age-groups in Experiment 1, which may be of interest to researchers who wish to use our materials. The version of Experiment 1 shown to children can be completed here: https://app.gorilla.sc/openmaterials/301970. All studies received approval from the Office of Research Ethics at the University of Toronto Mississauga (Protocol: 37628)

Results

All analyses were performed in R. We used linear mixed-effects regressions (LMERs) models to analyze continuous outcomes (i.e., RT) and generalized linear mixed-effects regressions (GLMERs) models to analyze binary outcomes (i.e., yes/no possibility judgments). These were run with lme4 (Bates et al., 2015) and passed through the joint_tests function from emmeans to produce a Type III omnibus test; any post hoc pairwise tests or trend analyses were also run with emmeans (Lenth et al., 2019). All models were fitted maximally with subject and event entered as random intercepts. The means and 95% confidence intervals (CIs) shown in each figure are outputted directly from these models.

Processing of RT Data

Our primary aim was to explore how RT related to judgments of possibility by comparing participants' RT across age-groups (5-year-olds, 7-year-olds, 9-year-olds, and adults) and event types (ordinary, improbable, impossible). However, there were large differences in overall RT across age-groups, with older participants responding faster on average than younger participants; likely due to age-related differences in overall processing speed (Kail, 1991) and familiarity with computers or tablets. However, we wanted to look for meaningful differences in RT across event types within each age-group, as this would tell us something about the specific relation between reflection and possibility judgments. We also wanted to ensure that average RTs were not influenced by extreme outliers, while retaining a maximal number of responses that could potentially reflect genuine reflection. But, we could not use overall RT scores to determine outliers due to the variance in our sample. This approach would lead to eliminating some responses (e.g., 5-year-olds judging improbable events) much more often than others (e.g., adults judging ordinary events).

Thus, we opted for the following procedure. We defined outliers as any RT greater than two standard deviations above the mean RT within each age-group and event type. This approach removes only 3.04% of the responses. We also standardized RT within each age-group to account for more general differences in RT that emerge with age. Our main analyses use raw RTs with outliers removed, though we also report analyses with standardized RTs where relevant.

Overview of Analyses

We report four sets of analyses testing the development = reflection account's hypothesis that older children and adults explicitly reflect on their intuitions more than younger children, and as a result should take longer to affirm improbable events than deny them. First, we examine possibility judgments across age-groups and event types to ensure that our triadic item-set captures typical developmental trends in possibility judgments (e.g., Shtulman, 2009; Shtulman & Carey, 2007). Second, we explore RT differences across age-groups and event types to determine whether older children and adults were more likely to take time to reflect than younger children, and most likely to take time to reflect before judging whether improbable events could happen. Third, we examine differences in RT when participants affirm versus deny events to identify whether greater reflection predicts judging improbable events as possible rather than impossible. Finally, we present an analysis that examines how raw RT predicts possibility judgments for each age-group.

Possibility Judgments

We first explored how participants judged the possibility of ordinary, improbable, and impossible events, and whether judgments varied with age. Our findings replicate those of Shtulman and Carey (2007). All main effects and interactions were significant; see Figure 2 for a summary of the results. Judgments varied across event types, $F(2) = 196.81$, $p < .001$: participants more often affirmed ordinary events ($M = 0.98$, 95% CI [0.97, 0.99]) than improbable ($M = 0.52$, [0.40, 0.64]) or impossible events ($M = 0.04$, [0.02, 0.06]); and affirmed improbable events more than impossible ones, $ps < .001$. Judgments also differed across age-groups, $F(3) = 31.30$, $p < .001$: while 5-year-olds ($M = 0.34$, 95% CI [0.26, 0.44]) and 7-year-olds ($M = 0.40$, [0.31, 0.50]) affirmed events at similar rates, $p = .697$, 9-year-olds ($M = 0.63$, [0.53, 0.72]) affirmed more events than 5- and 7-year-old children, $ps < .001$; adults ($M = 0.85$, [0.78, 0.91]) affirmed more events than children in all age-groups, $ps < .001$.

The interaction between event type and age-group emerged because judgments for some event types were more stable across age-groups than others, $F(6) = 61.80$, $p < .001$. This is evident from Figure 1. Older participants were more likely than younger children to affirm improbable events, $ps \leq .020$, and ordinary events, $ps \leq .001$. For impossible events, 5-year-olds were more likely than 7- and 9-year-olds to affirm them, $ps \leq .007$, but judgments for these events were otherwise similar across ages, $ps \geq .106$.

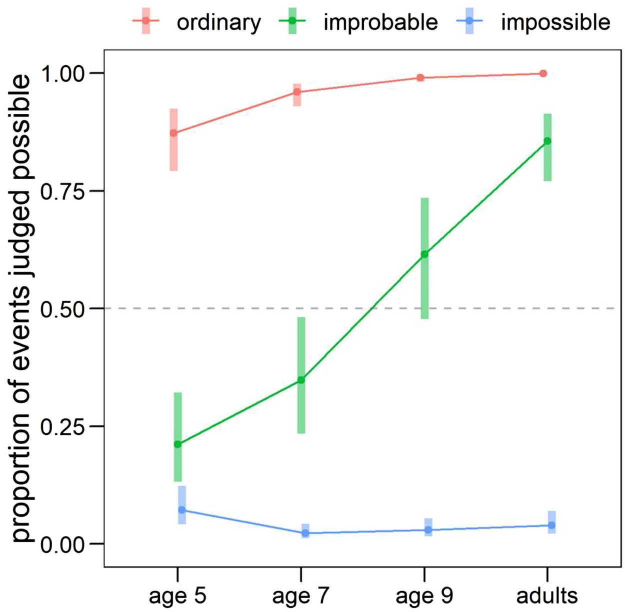
RT

The development = reflection account of modal reasoning predicts an interaction between age and event type. Young children should respond at similar speeds to all event types, whereas older children and adults should respond more slowly to improbable events than to ordinary and impossible events. To test these predictions, we explored whether participants took more time to judge the possibility of some event types than others, and whether this varied with age; see Figure 3. Our results do not support this prediction.

We first looked at differences in raw RTs (see Figure 3, left panel). Here, all effects were significant. RT differed across event types, $F(2) = 41.13$, $p < .001$: participants responded slowest to improbable events ($M = 2{,}950$, 95% CI [2,768, 3,133]), then impossible events ($M = 2{,}266$, [2,084, 2,448]), and were fastest for ordinary events, ($M = 2{,}130$, [1,948, 2,312]), $ps \leq .001$. Participants also generally responded more quickly with age, $F(3) = 70.88$, $p < .001$: 5-year-olds ($M = 3{,}734$, 95% CI [3,466, 4,001]) were the slowest,

Figure 2

Proportion of Improbable, Impossible, and Impossible Events Affirmed by Each Age-Group



Note. Dots show means, whiskers show 95% CIs. CIs = confidence intervals. See the online article for the color version of this figure.

ps < .001; 7-year-olds (*M* = 2,621, [2,362, 2,879]) and 9-year-olds (*M* = 2,383, [2,125, 2,642]) responded at similar rates, *p* = .548, while adults (*M* = 1,057, [793, 1,322]) responded faster than children in all age-groups, *ps* < .001.

The interaction between event type and age-group emerged because RT did not decrease evenly with age across all event types, $F(6) = 7.40$, $p < .001$. While 7- and 9-year-olds responded at similar rates for improbable events, $p = .998$ and impossible events, $p = .592$, judgments for these events otherwise sped up between Ages 5 and 7, *ps* ≤ .001, and between Age 9 and adulthood, *ps* < .001. In contrast, judgments for ordinary events sped up at every age, *ps* < .034.

We then reran these analyses on our standardized RT scores (see Figure 3, right panel). Here, we found that response time varied across event types, $F(2) = 50.03$, $p < .001$: improbable events were judged slowest and ordinary events were judged fastest, with impossible events falling in between. Predictably, there was no effect across age-groups, $F(3) = 0.01$, $p = .999$. But a significant interaction effect again emerged, $F(6) = 21.41$, $p < .001$. As shown in Figure 2 (right panel), relative RTs for improbable and impossible events did not significantly differ as a function of age, *ps* ≥ .063. For ordinary events, adults responded more quickly than 5-year-olds, $p = .006$; standardized RTs for ordinary events were otherwise comparable across age-groups, *ps* ≥ .092.

RT Differences When Affirming or Denying Events

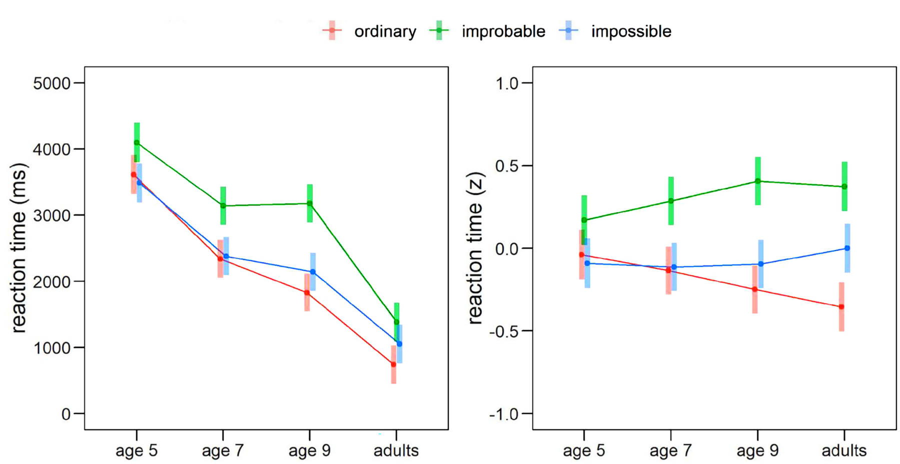
The development = reflection account of modal reasoning predicts that participants who reflect on their intuitions about improbable events should more often affirm them as possible. We explored this by assessing whether participants took longer to affirm improbable events than to deny them. This hypothesis was not supported by our data.

We recoded our data such that responses were labeled "correct" or "incorrect"; for ordinary and improbable events this meant affirming them, and for impossible events this meant denying them. These labels were derived from how adults typically respond to these kinds of events, both in the present study and in past work (e.g., Shtulman, 2009; Shtulman & Carey, 2007). We then ran a model to evaluate whether standardized RT was predicted by response type (correct, incorrect), age-group, and event type.

Participants were overall slower to respond when giving incorrect ($M_Z = 0.38$, 95% CI [0.27, 0.49]) than correct ($M_Z = 0.02$, [−0.05, 0.09]) answers, $F(1) = 65.03$, $p < .001$. Their response time also differed across event type as expected, $F(2) = 12.41$, $p < .001$, and did not differ with age, $F(3) = 1.09$, $p = .353$. However, there

Figure 3

RT Across Event Types and Age-Groups



Note. Left panel: raw RT; right panel: standardized RT. Dots show means, whiskers show 95% CIs. RT = reaction time; CIs = confidence intervals. See the online article for the color version of this figure.

were significant interactions between response type and age-group, $F(3) = 10.87$, $p < .001$, response type and event type, $F(2) = 56.26$, $p < .001$, and a three-way interaction between response type, event type, and age-group, $F(6) = 2.15$, $p = .045$. The interaction between age-group and event type was not significant, $p = .241$.

We first probed the three-way interaction between response type, event type, and age-group interaction by exploring how relative RTs for each response type varied with age within each event type; the general trends are depicted in Figure 4. For correct responses, RTs remained stable with age across all event types, $ps \geq .193$, with one exception: adults were quicker than 5-year-olds when responding correctly to ordinary events, $p = .014$. For incorrect responses, RTs for ordinary events remained stable across age, $ps \geq .270$, and adults were slower than 5-year-olds when responding incorrectly to impossible events, $p = .018$; RTs were otherwise stable for these two event types, $ps \geq .327$. For incorrect responding to improbable events, RTs were stable between 5- and 7-year-olds, $p = .608$, between 7- and 9-year-olds, $p = .054$, and between 9-year-olds and adults, $p = .140$. Both 9-year-olds and adults were slower than 5-year-olds to respond incorrectly to improbable events, $ps \leq .001$, and adults were slower than 7-year-olds, $p = .001$.

We then looked at whether RTs differed between correct and incorrect responses for each event type at each age. For ordinary events, 5-year-olds' rate of responding did not differ based on accuracy, $p = .055$, but older participants were always faster to give correct than incorrect responses, $ps \leq .039$. For impossible events, correct responses were provided faster at all ages than incorrect responses, $ps < .001$. For improbable events, the pattern in participants' judgments reversed with age: 5- to 7-year-olds were faster to give incorrect than correct responses, $ps < .001$, 9-year-olds' RTs did not differ based on accuracy, $p = .383$, and adults were faster to give correct than incorrect responses, $p = .001$.

Do Possibility Judgments Increase With Greater Reflection Time?

Finally, we explored how RT predicted overall responding. We examined this with a model in which event type (ordinary, improbable, impossible), age-group (5-year-olds, 7-year-olds, 9-year-olds, and adults), and standardized RT predicted affirmations of possibility. According to the development = reflection account of modal reasoning, taking more time to respond should predict greater affirmations of possibility for improbable events. This hypothesis was not supported by our data.

There were significant effects of event type, $F(2) = 201.55$, $p < .001$, and age-group, $F(3) = 29.21$, $p = .001$, but no effect of RT, $F(1) = 0.70$, $p = .404$. However, all interactions were significant, including the three-way interaction between event type, age-group, and RT, $F(6) = 2.232$, $p = .037$.

We unpacked this by separately analyzing each slope shown in Figure 5. In short: responses for each event type usually moved toward chance (0.5), representing a reversal from whatever pattern emerged at the lowest RTs for each age-group. For ordinary events, 5- and 7-year-olds' responses were unrelated to their RT, $ps \geq .081$, whereas 9-year-olds and adults were more likely to deny ordinary events as their RTs increased, $ps \leq .032$. For impossible events, all age-groups were more likely to affirm events as RTs increased, $ps < .001$. For improbable events, the nature of the shift depended on whether each age-group was more likely to affirm or deny the event at lower RTs: 5- and 7-year-olds became more likely to affirm these events, $ps \leq .006$. Adults became more likely to deny them, $p = .006$, and 9-year-olds' judgments did not appreciably shift over time, $p = .508$.

Discussion

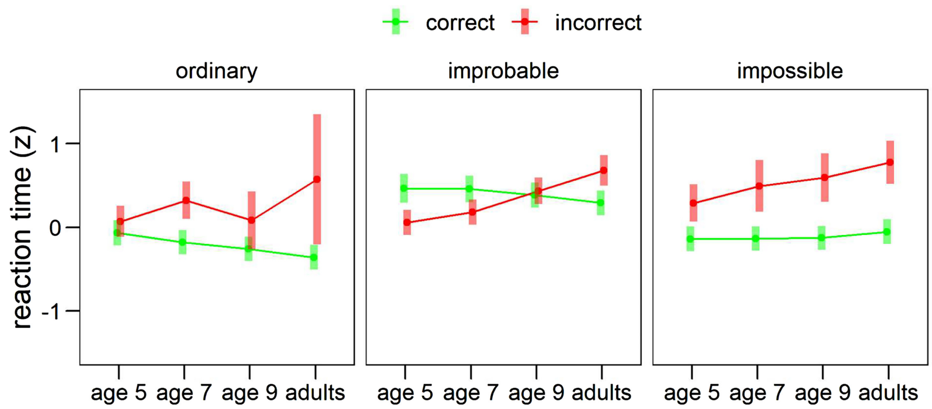
In Experiment 1, we replicated the now-classic finding that, with increasing age, children are more likely to affirm the possibility of improbable events (Shtulman & Carey, 2007). We achieved this with a novel item-set composed of 26 homologous triads of ordinary, improbable, and impossible events; this item-set is larger and broader than any previously used to investigate children's beliefs about possibility. Crucially, this item-set allowed us to carefully explore how long it took participants to decide whether these events could happen. We therefore explored whether we would capture a developmental increase in reflection—as indexed by RT—when judging improbable events relative to ordinary or impossible events. If adults spent more time before judging improbable events and children did not, then we would have evidence that adults (but not children) engage in reflective reasoning about their modal intuitions before affirming these events.

Our findings so far do not suggest that adults are more likely to reflect than children when judging possibility. First, all participants took longer to judge the possibility of improbable events than ordinary or improbable ones—even the youngest children, who usually respond as if improbable events are impossible. Second, after accounting for age differences in general processing speed, participants at all ages did not differ in the time taken to affirm improbable events. We also found that younger children were quicker to deny rather than affirm improbable events, whereas adults were quicker to affirm rather than deny such events. Crucially, this trend was driven by an age-related increase in denial RTs, rather than affirmation RTs. Thus, this pattern of result is inconsistent with the development = reflection account which predicts an age-related increase in affirmation RTs due to an age-related increase in reflection when judging the possibility of improbable events. Third, the main developmental shift in RTs was observed for ordinary events, as adults judged these events faster than 5- and 7-year-olds. Finally, participants provided the modal responses for their age-group at the lowest RTs; increases in RT predicted responses moving toward chance responding or ambivalence for all event types.

Together, these results suggest that affirming improbable events is largely not a reflective process for adults, and that the development of modal judgments cannot be attributed to the development of reflective processes. Reflection may allow people to provide responses that go against their intuitions, as we found that responses at higher RTs were likely to be the opposite of whatever the modal response was for any given age-group and event type. But our results suggest that changes in how people's initial intuitions are generated may better explain the development of the patterns in modal judgment usually observed when children and adults judge the possibility of strange and unfamiliar events. Experiment 2 tests this further by asking adults to make possibility judgments under time pressure. If adults must reflect on their intuitions to affirm improbable events, then they should more often deny improbable events when they are deprived of the opportunity to reflect. But if adults' beliefs are instead driven by their intuitions, we should find that limiting their opportunity to reflect has little impact on their possibility judgments.

Figure 4

RT Differences Across Correct and Incorrect Responses



Note. Relative RTs for correct and incorrect responding are shown within each event type. Dots show means, whiskers show 95% CIs. RTs = reaction times; CIs = confidence intervals. See the online article for the color version of this figure.

Experiment 2

Method

Participants

We tested 52 Canadian and American adults (M_{age} = 24 years, range = 18–32 years, 24 female, 27 male, one identified as other or preferred not to answer) through Prolific. We tested an additional three participants but excluded them for failing an attention check (two) or exceeding the time limit (one). Our sample was 42.32% White, 15.38% South Asian, 7.69% Chinese, 7.69% mixed race, 5.77% Black, 5.77% Korean, 3.85% Arab, 3.85% Filipino, 3.85% Latin American, 1.92% South-East Asian, and 1.92% Central Asian.

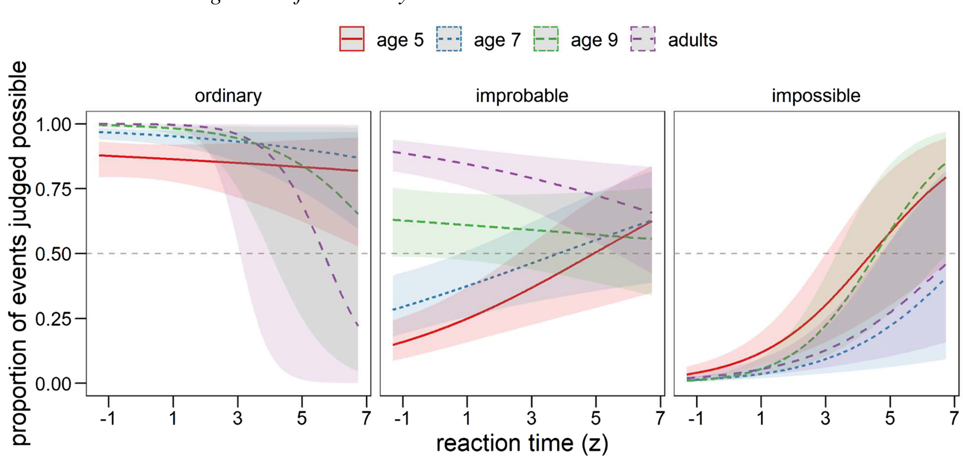
We compared the responses from this speeded group to the adults from Experiment 1, who completed the same task without time limits.

Procedure

The experiment was identical to the version of Experiment 1 prepared for adults, with one crucial change: participants were given only 1,000 ms to respond on all trials. In Experiment 1, adults' average response time was 736 ms for ordinary events and 1,374 ms for improbable events. We therefore set a 1,000 ms cutoff with the aim of giving adults enough time to respond while diminishing their opportunity to reflect on their modal intuitions. The training trials allowed participants to become accustomed to the time demands

Figure 5

RT in Relation to Judgments of Possibility



Note. Lines show predicted values that extend beyond the maximum RTs for each age-group and event type. They were generated by passing an LMER through ggeffects (Lüdecke, 2018). Ribbons show 95% CIs. RT = reaction time; LMER = linear mixed-effects regression; CIs = confidence intervals. See the online article for the color version of this figure.

of the task. If participants failed to respond within 1,000 ms on a test trial, their response was logged as a time-out; these trials are therefore absent from our main analyses. Only 8.14% of responses were time-outs, and we discuss the characteristics of these time-outs below.

Results

For our analyses, we included all adult responses from Experiment 1 that were within two standard deviations of the mean RT for each event type as an unspeeded comparison group. The speeded data includes all data points except one: a single glitched trial that allowed a participant to respond after 1,000 ms had elapsed.

We first checked whether adults' judgments for ordinary, improbable, and impossible events differed under time pressure, compared to when they were given unlimited time to decide. But they did not: overall, adults' judgments under time pressure did not differ from those made with no time constraints, $F(1) = 0.13$, $p = .715$. As before, adults' judgments differed across event types, $F(2) = 177.27$, $p < .001$. But there was also an interaction between time pressure and event type, $F(2) = 4.96$, $p = .007$; see Figure 6. This emerged because adults under time pressure were more likely to affirm impossible events ($M = .06$, 95% CI [0.03, 0.10]) than adults who had unlimited time to respond ($M = .03$, [0.02, 0.06]), $p = .041$. Time pressure had no effect on judgments for ordinary or improbable events, $p\text{s} \geq .100$.

We next checked for variations in RT across event types. Given that RTs will inevitably differ across the unspeeded and speeded versions of the experiment, we report the findings for the standardized RTs; both raw and standardized data are shown in Figure 6. We find no effect of condition, $F(1) = 0.19$, $p = .666$, and a main effect of event type, $F(2) = 23.80$, $p < .001$. But we also find an interaction between the two, $F(2) = 24.14$, $p < .001$.

As reported for Experiment 1, relative RTs for all event types pulled apart under unspeeded conditions: participants were fastest for ordinary and slowest for improbable, with impossible events falling in between, $p\text{s} < .001$. But when participants were under time pressure, relative RTs for improbable and impossible events did not differ, $p = .662$, though participants were still significantly faster when judging ordinary events, $p\text{s} < .001$.

Participants in the speeded condition occasionally timed-out and were unable to provide a response. Participants timed-out on 12.50% of improbable trials, 8.65% of impossible trials, and 3.25% of ordinary trials. This pattern may explain why RTs for improbable and impossible events did not pull apart in the speeded condition, as the events that took the longest to reason about were ultimately timed-out and omitted from our analyses. However, we can speculate from the findings in Figure 5 that these timed-out responses would more likely be incorrect than correct. This is opposite the pattern expected by a reflection account, which anticipates that more reflection should lead to more correct responding.

Discussion

In Experiment 2, we tested whether adults under time pressure would be more likely to deny the possibility of improbable events, as they would have little time to reflect on their intuitions before responding. We found that these time constraints had no impact: adults who responded within one second provided judgments that fully aligned with judgments made by adults with unlimited time to reflect before responding. This suggests that adults do not typically reflect on their modal intuitions before judging possibility.

Our findings so far suggest that the development of possibility judgments is not driven by increased reflection, and that these judgments may largely reflect quickly-generated intuitions. Experiment 1 found that children and adults deliberate more before judging improbable events than other events, but this pattern exists at all ages—even 5-year-olds take longer to reason about these events, despite usually (incorrectly) denying them. Further, Experiment 2 shows that adults' possibility judgments are not impacted by time pressure, suggesting that their judgments are driven by quick intuitions rather than prolonged reflection. Together, this suggests that the development of possibility judgments is driven by changes in modal intuitions themselves rather than an increased tendency to reflect on those intuitions.

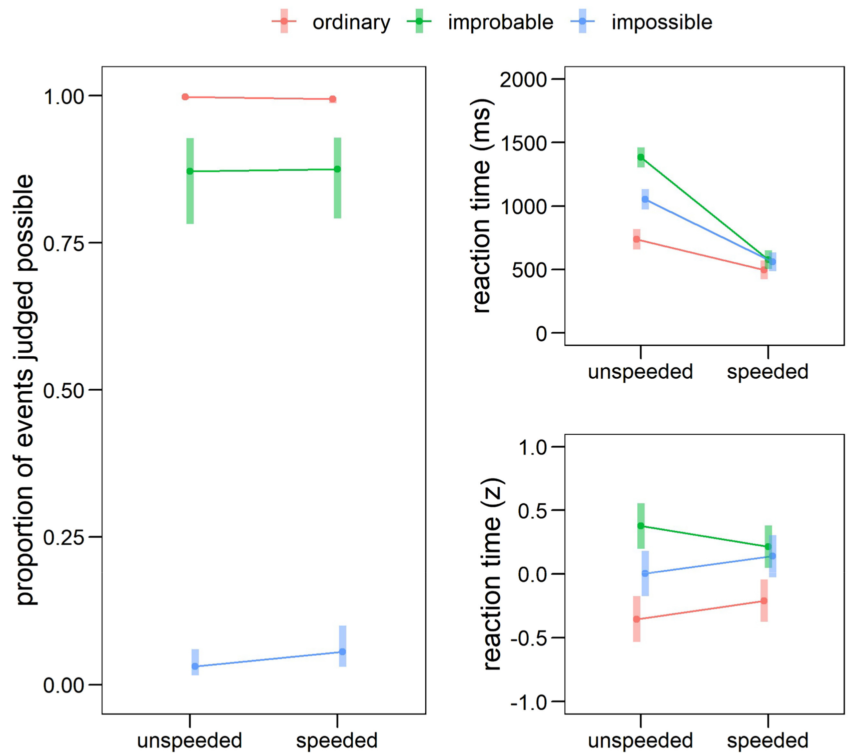
After discovering evidence in favor of a developmental shift in how intuitions are generated, we sought to gain some insight into how this process changes with age. In a final set of exploratory analyses, we explored the kind of "evidence" that children and adults consider when forming modal intuitions by fitting a drift-diffusion model (DDM) to the findings from Experiment 1.

Modal Intuitions, Sequential Sampling, and DDMs

As outlined in the Introduction, we propose that modal intuitions could emerge from the sampling of relevant knowledge and memories, and that age-related differences in possibility judgments might be explained by changes in how children and adults access their memories. Specifically, we suggest that modal intuitions might arise from a sequential sampling process that seeks to find affirming or disconfirming evidence of an event's possibility. In sequential sampling models of decision making, evidence is accumulated until a sufficient threshold of evidence is reached and a decision is made (Forstmann et al., 2016). These decisions are not based on absolute certainty, but instead employ a more frugal "take-the-best" approach to the problem that favors efficiency over accuracy (Gigerenzer & Goldstein, 1996; Lee & Cummins, 2004). Sequential sampling accounts have primarily been used to explain how people make decisions based on perceptual evidence, but recent work has argued that the same process may apply to obtaining evidence from memories (see Shadlen & Shohamy, 2016), as it would in the case of modal judgment.

We attempted to capture this process with a DDM. This kind of model can infer whether people retrieve evidence in favor of or against an event's possibility, and can approximate the duration of this retrieval process separate from overall processing speed. In other studies, this evidence is often visual and obtained from the stimuli itself. Here, we only presented participants with an audio prompt, and began measuring their response time after the prompt ended. So, any "evidence" our participants sought must have been obtained by retrieving their own relevant knowledge and memories. This model can also approximate the relative amount of evidence that must be accumulated before an event is affirmed or denied (e.g., Ratcliff, 1978; Ratcliff et al., 2016). We propose that when people have accumulated enough evidence from memory to cross a decision threshold about an event's possibility, this is experienced

Figure 6
Findings of Experiment 2



Note. Left panel: possibility judgments among adult participants for the speeded (Study 2) and unspeeded (Study 1) versions of the task. Top-right: RT data across event types; bottom-right: standardized RTs. Dots show means, whiskers show 95% CIs. RT = reaction time; CIs = confidence intervals. See the online article for the color version of this figure.

as a modal intuition and—in most cases—expressed as a modal judgment. Here, we explore whether there is a developmental shift in (a) the kind of evidence that is accessed (confirming or disconfirming), (b) the duration of search, and (c) the amount of evidence needed to cross the "affirm" or "deny" thresholds.

Drift-Diffusion Analyses of Experiment 1

Here, we report key details of our model and the main findings; for our full method and analyses including additional figures and statistical tables, see the supplementary materials available on the OSF page (https://osf.io/5pdjf/).

First, we pooled the single-trial RT data across participants separately for each age-group. We eliminated both fast and slow outliers by trimming the fastest and slowest 2.5% RTs for each age-specific pooled data, and we adopted bootstrapping techniques to estimate drift-diffusion parameters for each age-group. We then fit a DDM by allowing starting point (*zr*), drift rate (*v*), threshold separation (*a*), and nondecision time (*t0*) to vary. Importantly, we allowed *v* to vary across three conditions (ordinary, impossible, improbable), and *a* to vary between certain questions (i.e., ordinary and impossible questions) and uncertain questions (i.e., improbable questions). The results are summarized in Figure 6.

We found a monotonic decrease in nondecision time (*t0*) with age, showing that the age-related decrease in RT was partly due to speed of response executions, that is, moving the cursor and clicking on the response options. The 95% CI for the starting point (*zr*) included 0.5 for all ages, indicating that no age-group exhibited a statistically significant bias to affirm or deny events before accumulating evidence (95% CI: [0.44, 0.53], [0.47, 0.53], [0.46, 0.53]. [0.47, 0.54] for 5-, 7-, and 9-year-olds and adults, see the online supplementary materials); the starting point also did not differ across age-groups. Threshold separations (*a*), on the other hand, were relatively stable across development. If anything, the decision thresholds decreased as a function of age; in other words, adults accumulated less evidence than 5-, 7- and 9-year-olds before making their judgments (certain decisions [95% CI]: adults [1.98, 2.48] < 5-year-olds [3.11, 4.06], 7-year-olds [2.85, 3.76], and 9-year-olds [2.77, 3.62]; uncertain decisions [95% CI]: adults [1.88, 2.46] < 5-year-olds [3.11, 4.15], 7-year-olds [2.73, 3.51], and 9-year-olds [2.64, 3.55]). However, this age-related reduction should be interpreted with caution, as a different model fitting procedure showed statistically-equivalent threshold separations across age-groups (see the online supplementary materials).

The absolute value of drift rate (*v*) for possible and impossible conditions increased with age, indicating that older participants were faster at accumulating evidence than younger ones toward correct boundaries (possible [95% CI]: 5-year-olds [0.28, 0.58], 7-year-olds [0.61, 0.97], 9-year-olds [0.86, 1.35], and adults [1.61, 2.24]; impossible [95% CI]: 5-year-olds [−0.76, −0.29], 7-year-olds [−1.06, −0.72], 9-year-olds [−1.21, −0.64], and adults [−1.53, −0.91]). However, for improbable items, we found that 5-year-olds (95% CI: [−0.42, −0.09]) accumulated evidence toward denying the events, whereas adults accumulated evidence toward affirming the events ([0.37, 0.87]), with 7-year-olds ([−0.34, 0.00]),) and 9-year-olds ([−0.02, 0.31]) falling in between.

Together, the results of this model suggest that adults and children may differ in their ability to access relevant knowledge, and may differ in the kinds of knowledge they sample when forming intuitions. Depending on how the model is run, we find that adults either sample less knowledge than children before judging possibility, or sample comparable amounts of knowledge. Both results suggest that adults are better able to access relevant knowledge and memories than children when forming their intuitions, as they ultimately (correctly) affirm more improbable events as possible. The model also suggests that adults and children sample different kinds of knowledge when forming their intuitions: for improbable events, adults retrieve evidence of possibility, while children retrieve evidence of impossibility. These findings align with the development = intuition explanation to account for age-related increases in the affirmation of improbable events.

General Discussion

Across two studies, we tested two accounts for the development of modal judgments by measuring how long children and adults took to judge the possibility of ordinary, impossible, and improbable events. The development = reflection account proposes that people first generate a modal intuition about whether events could happen; since improbable events are unfamiliar, this intuition suggests that the events are impossible. However, proceeding to actively reflect on these inaccurate intuitions leads to affirmations of possibility (Shtulman, 2009). The model explains the development of possibility judgments by proposing that older children and adults engage in this reflection while young children do not. In contrast, the development = intuition account posits that these age-related differences are driven by changes in modal intuitions themselves, as a result of age-related changes in what people know and in how knowledge is retrieved when generating these intuitions.

The development = reflection account predicts that adults should take longer to affirm improbable events than ordinary or impossible ones, whereas young children (who do not reflect) should take similar amounts of time to affirm or deny all kinds of events. It also predicts that placing adults under time pressure should lead them to more often deny improbable events, since this should deprive them of the opportunity to reflect. Conversely, the development = intuition account predicts that the time taken to affirm improbable events should either stay the same or decrease with age—all participants should use knowledge retrieval processes to generate modal intuitions, and these processes may speed up with age. Further, a development = intuition account predicts that speeding-up modal judgments (thereby limiting reflection time) should have little influence on participants' affirmation of improbable events.

We tested these predictions by measuring how quickly children and adults made judgments about 26 triads of homologous ordinary, improbable, and impossible events. Crucially, we anticipated that comparing RT across ordinary, improbable, and impossible events would index whether additional processing was performed when judging improbable events possible or impossible (Goldberg & Thompson-Schill, 2009; Shtulman & Harrington, 2016). Our within-subject design, large number of items, and the fact that we matched the length, complexity, and content of each of our 26 items across event type allowed for a strong test of these competing predictions.

Our findings from Experiment 1 suggest that increased reflection cannot explain the development of possibility judgments. We found that all age-groups took the longest to respond to improbable events, and the speed at which they did so remained constant with age after controlling for general processing speed. Further, after accounting for age-related differences in general reaction speed, we did not observe a reliable developmental increase in RTs when affirming improbable events. Together, these findings suggest that either participants of all ages reflected before judging improbable events, or that participants' judgments were driven by non-reflective processes at all ages.

Further, we found no evidence that taking longer to respond (a proxy for reflection) is overall associated with event affirmation over denial. On average, longer RTs were associated with incorrect responding; this was true at all ages for impossible events, and this pattern emerged with age for improbable events. We also found that the typical developmental pattern for possibility judgments (i.e., children deny improbable events while adults affirm them) emerged at the lowest RTs. These findings suggest that participants were most likely to reflect when they were responding counter to their intuitions; for instance, when taking time to decide whether it really is impossible for a person to eat a large rock or count all the grains of sand in the world. Or, participants may have simply taken a long time to respond when they were confused or especially uncertain. In any case, our evidence suggests that children's denials and adults' affirmations toward improbable events are often made quickly, potentially without requiring any active reflection.

Experiment 2 provides strong evidence for this proposal. One method for revealing whether adults hold intuitions that differ from their more reflective (and typical) judgments is to force them to make judgments under time pressure (Kelemen et al., 2013; Phillips & Cushman, 2017; Shtulman, 2017). But here, adults' judgments about the possibility of improbable events did not differ when they were under time pressure and deprived of the opportunity to reflect. This suggests that adults' possibility judgments were often driven by their initial intuitions even when they could reflect, and that reflection is not essential for adults to affirm the possibility of improbable events. Indeed, Experiment 2 revealed that in adulthood possibility judgments can be made in approximately half a second without any loss in accuracy—on average, adults made speeded judgments about ordinary events in 493 ms and judgments about improbable and impossible events in 560 and 549 ms, respectively. Although we cannot entirely rule out the possibility of reflection within these short timeframes, we argue it is more plausible that adults' speeded judgments were not the output of a reflective two-stage process. Instead, they may have quickly retrieved relevant knowledge to form an intuition, and this intuition was directly outputted as a possibility judgment.

We see further evidence for the developing role of knowledge retrieval in our RT data across ages. For instance, we found that judgments for ordinary events sped up with age, even after accounting for age-related increases in processing and movement speed. Children and adults likely have comparable knowledge about whether a person could have a pet cat or eat a sandwich, and they almost certainly do not have to reflect on these events to acknowledge their truth. So, adults may have judged these events faster simply because they were able to more quickly and efficiently retrieve this knowledge and form an intuition (Hjuler et al., 2023; Krøjgaard et al., 2022; Levy & Anderson, 2002).

In sum, our results are most consistent with the proposal that modal judgments often directly reflect modal intuitions, and that age-related differences in possibility judgments reflect developmental changes in how memories and knowledge are accessed and retrieved. Adults likely have more knowledge to draw upon when inferring possibility than children; for instance, they are likely to have seen more people riding strange animals or eating strange foods. But they may also be more efficient than children at retrieving relevant memories, and may even retrieve different kinds of evidence when forming their intuitions. To begin to understand what might be changing in the process of knowledge and memory retrieval we fit a DDM to our RT data from Experiment 1. This exploratory analysis allowed us to begin to understand what may and may not be changing with age.

A priori, we might expect that the development of possibility judgments can be explained by an age-related change in the evidence thresholds required to judge an event possible or impossible. Decision thresholds appear to shift with development in other domains of reasoning (Ratcliff et al., 2012), so we might also expect development in the thresholds for generating modal intuitions. Specifically, young children may have a much lower impossible threshold than adults, requiring far less evidence against an event's possibility before generating a negative intuition and dismissing the event as impossible. This would explain why children deny the possibility of most things that are unfamiliar, whereas adults only deny events that violate a causal law (Shtulman, 2009; Shtulman & Carey, 2007). But the results of our exploratory DDM appear to speak against this explanation. We found no bias in the starting point of evidence accumulation for determining the possibility of an event: zr included 0.5 for all ages (Figure 7), suggesting that no age-group exhibited a significant bias to affirm or deny events before accumulating evidence.

However, adults and children may be biased toward sampling different knowledge while attempting to generate modal intuitions. The absolute value of the drift rate (v) in our DDM describes the average speed of evidence accumulation, and the sign of the drift rate reflects the average direction of the judgment. If a participant was more likely to accumulate evidence in support of a "possible" judgment, then the sign of their drift rate was estimated positive; if they were more likely to accumulate evidence in support of an "impossible" judgment, then the sign of their drift rate was negative. The results of our model suggest that there is a shift in the kind of evidence children and adults search for when reasoning about improbable events: on average, the target events triggered a search for disconfirming evidence among 5-year-olds, and a search for confirming evidence among 9-year-olds and adults, with 7-year-olds' search strategy falling somewhere in between. For now, we can only speculate about how these search processes might differ with age. For example, if a scan for similar events is part of the input for feelings of possibility, then children and adults may retrieve different analogs despite possessing the same knowledge (Goulding et al., 2022). We might expect this since younger children focus on surface-level features whereas adults focus on deeper similarities when reasoning about analogs (see Chen, 1996; Chen & Klahr, 2008; Gentner & Hoyos, 2017), and this might result in children and adults searching their memories for different things when judging possibility. Thus, age-related increases in judgments about the possibility of improbable events may be in part due to differences in how young children and adults search through and sample their prior knowledge and experiences. Critically, this would also explain why adults sometimes judge improbable events to be impossible—like children, they may sometimes fail to retrieve relevant information even if they possess that information.

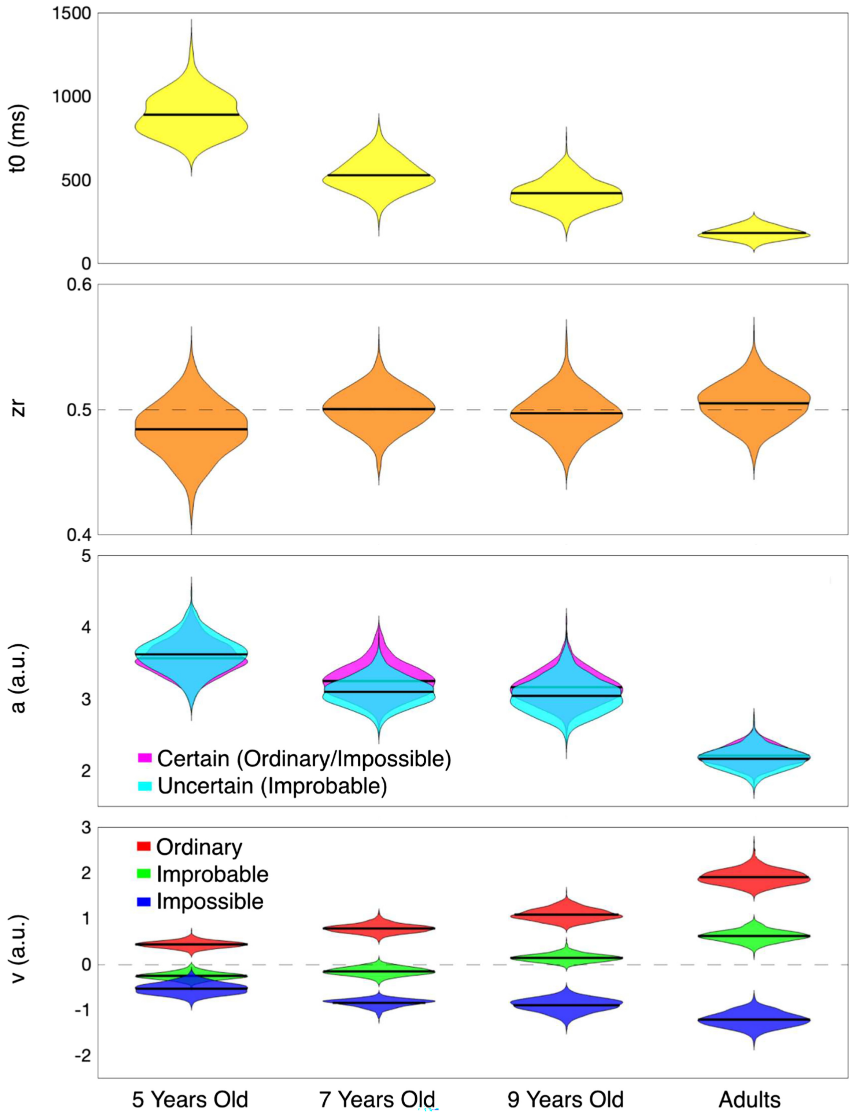
Another explanation is that while children and adults both sequentially sample their knowledge to infer possibility, how they search for information may be different. Adults have access to more knowledge than children and have better executive abilities, which should facilitate improved memory search (Hjuler et al., 2023; Krøjgaard et al., 2022; Levy & Anderson, 2002). Indeed, the results of our DDM (Figure 7) revealed that drift rate (v) increased with age, meaning older participants were faster at accumulating evidence than younger ones. However, we can only speculate as to what kinds of evidence might have contributed to participants' decisions, since participants themselves might have been unaware of how they sampled such data, especially given how quickly they formed modal intuitions. The model may also measure yet-unconsidered aspects of "evidence" collection besides causal knowledge or memory, or other cognitive processes entirely. Future work should explore whether children and adults search for different kinds of evidence when forming modal intuitions, and whether a developmental shift in memory sampling strategy or ability can help explain the development of possibility judgment.

Future studies could also be designed to explain why children and adults provide different answers when asked to explain why they believe an event is impossible. With increasing age, people justify their judgments of impossibility by appealing to causal constraints (e.g., "walls are solid") rather than by providing hypothetical explanations (e.g., "you could get a new apple") or redundant explanations (e.g., "that's not possible") (e.g., Shtulman & Carey, 2007). Given our data, such explanations appear to reflect post hoc reasoning about modal judgments and are unlikely to explain how many such judgments are made in the first place. Even so, why do post hoc justifications change with age and increasingly reference causal constraints? One explanation is that with increasing age people are more likely to think that causal constraint explanations are the best kind of explanations for modal judgments, or the kinds of explanations that people expect when asking for a justification. Indeed, children and adults usually find mechanistic explanations particularly satisfying, despite rarely possessing accurate mechanistic knowledge themselves (see Keil & Lockhart, 2021).

Alternatively, people of all ages may sample causal information to arrive at modal intuitions, but young children may be less aware of the content of this sampled knowledge than older children and adults, and therefore less able to accurately report the reasons behind their judgments. It may be that the causal information often present in people's explanations is identified through reflection on the intuition itself. People might feel like they have access to this relevant information when forming their intuitions, but may only fully access and reflect on this information when prompted to explain themselves, or when there is pressure to be certain about their judgments.

Figure 7

Drift-Diffusion Parameters Across Age-Groups



Note. Violin plots show the distribution of bootstrapped estimates of *t0* (yellow), *zr* (orange), *a* (cyan and magenta for the certain and uncertain decisions), and *v* (red, green, and blue for possible, improbable, and impossible conditions). The black line in each violin represents the mean estimate for the corresponding parameter. See the online article for the color version of this figure.

In this way, modal intuitions might operate like feeling-of-knowing judgments derived from apparent knowledge of relevant events or causal principles, which can be retrieved and accessed in full during later reflective processing (Cultice et al., 1983; Koriat, 1993; Wellman, 1977). Development in children's ability to access the data underlying their feeling of possibility could account for age-related improvements in the ability to justify possibility judgments.

Our data suggest that reflection does not explain the development of possibility judgments, and that these judgments may largely reflect quickly-generated intuitions. However, our data does not suggest that people are incapable of reflection, that reflection does not change with age, or that reflection does not matter for judgments of possibility. Indeed, our findings suggest that even young children might engage in reflection when judging possibility when they are confused or uncertain about their intuitions. Our findings also suggest that reflection may enable both children and adults to override their intuitions when judging possibility, as children became more likely to correctly affirm improbable events at higher RTs, whereas adults became more likely to incorrectly deny them. This aligns with recent work showing that children who are more likely to engage in cognitive reflection also show more mature modal reasoning, suggesting that reflection is one pathway to affirming unlikely possibilities (Shtulman et al., 2023). Thus, reflection is present at every age and matters for judgments of possibility. Two important steps for future research are to investigate the connection between the generation of intuitions and reflective processing and to investigate why reflecting on improbable events leads to different judgments for young children and adults—increased endorsement for the former and decreased endorsement for the latter.

It is important to note that only a small selection of our improbable events were clearly immoral (e.g., cutting the line to go into a store). Morality and convention constrain children's judgments of possibility (Browne & Woolley, 2004; Chernyak et al., 2019; Kalish, 1998; Komatsu & Galotti, 1986; Miller et al., 2000; Shtulman & Phillips, 2018), thus young children judge actions that violate norms to be impossible, and even adults more often judge immoral events as impossible under time constraints (Phillips & Cushman, 2017). Importantly, and consistent with our results, prior work has also found that adults do not change their judgment of possibility about statistically unlikely items (Phillips & Cushman, 2017). Thus, reflection might play a greater role for thinking about the possibility of moral and normative actions, as modal intuitions for these events may be especially strong and may align less often with a more reflective response.

Our studies also employed fully binary judgments of possibility (yes/no), which allowed us to directly connect our findings with the large body of work showing that children mostly deny events while adults mostly affirm them (e.g., Goulding & Friedman, 2020, 2021; Shtulman, 2009; Shtulman & Carey, 2007; Shtulman & Tong, 2013). We argue that longer response times potentially indexed either uncertainty or a reversal in judgment (Figure 5). However, we did not ask about uncertainty, so we cannot know for sure when participants were genuinely uncertain, or when they were actually overriding their intuitions via reflection. Future work measuring RT could offer participants the opportunity to indicate uncertainty rather than making a judgment. This design could capture whether higher delays in responding are associated with uncertainty (i.e., clicking "I don't know"), an intuition override (i.e., judging an impossible event as possible), or both.

Constraints on Generality

Our studies looked at the possibility judgments of Canadian 5- to 9-year-olds from a large urban center, and at the judgments of Canadian and American adults. We cannot be certain that children and adults with different backgrounds would provide similar judgments, but our development = intuition account would largely account for any discrepancies, since intuitions should be driven by accessible knowledge. For instance, a child who grew up around farm animals might be more likely to agree that a person could ride a cow than a person who grew up in a city, and a child from a region with thatched-roofed houses, which are uncommon in Canada, might be more likely to agree that a person could make a house out of grass. Indeed, children with religious backgrounds often say that miracles can happen in real life (Davoodi et al., 2023; Payir et al., 2021), but these judgments—like the judgments of the children in our own study—appear to stem from differences in the knowledge they receive from testimony rather than core differences in reasoning compared to nonreligious children.

One important individual difference that might matter, though, is a person's tendency to engage in cognitive reflection, which might facilitate counterintuitive responding among both children and adults. Children who engage in more cognitive reflection demonstrate more sophisticated modal reasoning (Shtulman et al., 2023), and cognitive reflection predicts rational thinking in both Chinese and American children (Gong et al., 2021; Shtulman & Young, 2023). But some cultures might promote the importance of using intuition over reflection, or vice versa (Buchtel & Norenzayan, 2008; Norenzayan et al., 2002). Future work should explore whether the RT trajectory observed here replicates in populations that are either more or less likely to value intuition and reflection as pathways to sound judgment.

Conclusion

In sum, we find no evidence that the development of possibility judgments is driven primarily by an increasing tendency to reflect on modal intuitions. Children and adults seem to use the same cognitive processes to make decisions about the possibility of ordinary, improbable, and impossible events. These judgments are made relatively quickly within each age-group, and seem to preclude any time for reflection on those intuitions. Thus, our results support the hypothesis that a developmental shift in what children know and how knowledge is retrieved leads to a shift in modal intuitions with age, and that these intuitions are what drive possibility judgments.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, Article e9414. https://doi.org/10.7717/peerj.9414

Browne, C. A., & Woolley, J. D. (2004). Preschoolers' magical explanations for violations of physical, social, and mental laws. *Journal of Cognition and Development*, *5*(2), 239–260. https://doi.org/10.1207/s15327647jcd0502_4

Buchtel, E. E., & Norenzayan, A. (2008). Which should you use, intuition or logic? Cultural differences in injunctive norms about reasoning. *Asian Journal of Social Psychology*, *11*(4), 264–273. https://doi.org/10.1111/j.1467-839X.2008.00266.x

Chen, Z. (1996). Children's analogical problem solving: The effects of superficial, structural, and procedural similarity. *Journal of Experimental Child Psychology*, *62*(3), 410–431. https://doi.org/10.1006/jecp.1996.0037

Chen, Z., & Klahr, D. (2008). Remote transfer of scientific reasoning and problem-solving strategies in children. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 36, pp. 419–470). Elsevier.

Chernyak, N., Kang, C., & Kushnir, T. (2019). The cultural roots of free will beliefs: How Singaporean and U.S. children judge and explain possibilities for action in interpersonal contexts. *Developmental Psychology*, *55*(4), 866–876. https://doi.org/10.1037/dev0000670

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

Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling-of-knowing judgments. *Child Development*, *54*(6), 1480–1486. https://doi.org/10.2307/1129810

Davoodi, T., Jamshidi-Sianaki, M., Payir, A., Cui, Y. K., Clegg, J., McLoughlin, N., Harris, P. L., & Corriveau, K. H. (2023). Miraculous, magical, or mundane? The development of beliefs about stories with divine, magical, or realistic causation. *Memory & Cognition*, *51*(3), 695–707. https://doi.org/10.3758/s13421-021-01270-2

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, *9*(3), 672–693. https://doi.org/10.1111/tops.12278

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science*, *20*(4), 480–487. https://doi.org/10.1111/j.1467-9280.2009.02320.x

Gong, T., Young, A. G., & Shtulman, A. (2021). The development of cognitive reflection in China. *Cognitive Science*, *45*(4), Article e12966. https://doi.org/10.1111/cogs.12966

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., 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

Harris, P. L. (2021). Early constraints on the imagination: The realism of young children. *Child Development*, *92*(2), 466–483. https://doi.org/10.1111/cdev.13487

Hjuler, T. F., Sonne, T., Kingo, O. S., Berntsen, D., & Krøjgaard, P. (2023). "I can't remember!" Three-year-olds struggle to strategically access encoded and consolidated memories. *Cognitive Development*, *65*, Article 101292. https://doi.org/10.1016/j.cogdev.2022.101292

Johnson, C. N., & Harris, P. L. (1994). Magic: Special but not excluded. *British Journal of Developmental Psychology*, *12*(1), 35–51. https://doi.org/10.1111/j.2044-835X.1994.tb00617.x

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, *109*(3), 490–501. https://doi.org/10.1037/0033-2909.109.3.490

Kalish, C. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development*, *69*(3), 706–720. https://doi.org/10.2307/1132199

Keil, F. C., & Lockhart, K. L. (2021). Beyond cause: The development of clockwork cognition. *Current Directions in Psychological Science*, *30*(2), 167–173. https://doi.org/10.1177/0963721421992341

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*(4), 1074–1083. https://doi.org/10.1037/a0030399

Komatsu, L. K., & Galotti, K. M. (1986). Children's reasoning about social, physical, and logical regularities: A look at two worlds. *Child Development*, *57*(2), 413–420. https://doi.org/10.2307/1130597

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Krøjgaard, P., Sonne, T., Kingo, O. S., & Berntsen, D. (2022). Spontaneous verbal recall: A new look at the mechanisms involved in episodic memory retrieval in young children. *Developmental Review*, *66*, Article 101050. https://doi.org/10.1016/j.dr.2022.101050

Lane, J. D., & Harris, P. L. (2014). Confronting, representing, and believing counterintuitive concepts: Navigating the natural and the supernatural. *Perspectives on Psychological Science*, *9*(2), 144–160. https://doi.org/10.1177/1745691613518078

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

Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, *11*(2), 343–352. https://doi.org/10.3758/BF03196581

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.4.5). R Package. https://cran.r-project.org/web/packages/emmeans/index.html

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, *6*(7), 299–305. https://doi.org/10.1016/S1364-6613(02)01923-X

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), Article 772. https://doi.org/10.21105/joss.00772

McCloskey, M. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 485–502. https://doi.org/10.1016/S0022-5371(80)90330-8

Miller, S. A., Custer, W. L., & Nassau, G. (2000). Children's understanding of the necessity of logically necessary truths. *Cognitive Development*, *15*(3), 383–403. https://doi.org/10.1016/S0885-2014(00)00034-4

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

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

Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, *26*(5), 653–684. https://doi.org/10.1207/s15516709cog2605_4

Payir, A., McLoughlin, N., Cui, Y. K., Davoodi, T., Clegg, J. M., Harris, P. L., & Corriveau, K. H. (2021). Children's ideas about what can really happen: The impact of age and religious background. *Cognitive Science*, *45*(10), Article e13054. https://doi.org/10.1111/cogs.13054

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649–4654. https://doi.org/10.1073/pnas.1619717114

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. E. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, *83*(1), 367–381. https://doi.org/10.1111/j.1467-8624.2011.01683.x

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, *90*(5), 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

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. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Basic Books.

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., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*(1), 118–137. https://doi.org/10.1111/tops.12174

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

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

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

Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child Development*, *48*(1), 13–21. https://doi.org/10.2307/1128875

Woolley, J. D., & Ghossainy, M. E. (2013). Revisiting the fantasy–reality distinction: Children as naïve skeptics. *Child Development*, *84*(5), 1496–1510. https://doi.org/10.1111/cdev.12081

Received October 26, 2022
Revision received August 20, 2023
Accepted August 23, 2023 ■