Title
It doesn't matter how you ask: Many 4-year-olds don't quantify possibilities

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Leahy, Brian
Close, Scarlett A. D.
Carey, Susan

Publication Date
2026

It doesn't matter how you ask: Many 4-year-olds don't quantify possibilities

Brian Leahy,[†,1] **Scarlett Close,**[2] **Susan Carey,**[3]
[1]Massachusets Institute of Technology, [2]University of Bath, [3]Harvard University
[†] brianlea@mit.edu

Abstract

Many 4-year-olds give adultlike responses to questions about what "can" happen and the same responses to questions about what "has to" happen. Why? We argue that children repeatedly apply minimal representations of possibility to answer "can" and "have to" questions. This algorithm (a) only requires evaluating the single possibility each question asks about; (b) generates correct responses to "can" questions and (c) generates the same responses to "have to" questions. We show that half of 4-year-olds answer "have to" questions as adults do, correctly say what all the possibilities are and what the only possibilities are, and show sensitivity to multiple possibilities in nonverbal behavior. The other half make mistakes on "have to" questions, fail to quantify possibilities using "all" and "only", and deploy minimal representations of possibility on the nonverbal task. They make their decisions by evaluating a single possibility.

Keywords: Possibility; modal concepts; modal language

Introduction

When making decisions under uncertainty, it's often worth keeping a couple possibilities in mind: possibilities that we can't yet rule out, but that we can't yet rule in either. This helps us form contingency plans that will succeed no matter which possibility is actual. Our languages also provide tools for communicating about mere possibilities. Our *modal vocabulary* encodes the *mode* of some proposition's truth: as a possibility or a necessity. This vocabulary allows us to coherently describe states of affairs that are otherwise incoherent. It's impossible for Sydney to be in Somerville and in New York, and it's incoherent for me to say that she is. But there's nothing incoherent about saying that Sydney *might* be in Somerville and *might* be in New York. Words that perform this function may be universal among human languages. These words express concepts that play the same role in cognition, allowing us to keep incompatible possibilities in mind without making our beliefs incoherent.

The current paper describes variability in how 4-year-olds use these concepts. We find patterns in how they quantify incompatible possibilities in their nonverbal behavior, how they understand modal quantifiers, and how they use general-purpose quantifying vocabulary to quantify possibilities.

Preschoolers behave in nonadult ways in some tasks that check whether they can tell the difference between an established fact and a mere possibility. In the 3-cups task (Mody & Carey, 2016), participants see 3 cups organized into 2 sets: a pair and a singleton. One set (e.g., the singleton) is occluded, and a prize is hidden in it; then the other set is occluded, and a second prize is hidden there. The participant's information supports the inference that there is a prize in the singleton, and that there is a prize in the pair–but nothing more specific. They can't infer where the prize in the pair is; it might be the left and it might not–it might be the right. A child who reliably chooses the singleton shows that they can tell the difference between a cup that *holds* a prize and the two cups that merely *might* hold a prize.

Older 2-year-olds choose the singleton 50% of the time. This is surprising because random guessing should yield 33% choice of the singleton, as there are three cups. Differentiating the sure thing from the mere possibilities should yield closer to 100% choice of the singleton. Analysis of how errors distribute over participants suggests that *each individual*, on *every trial*, chooses the singleton with probability .5 (Leahy, Huemer, Steele, Alderete, & Carey, 2022, SI). In particular, when each individual's proportion of singleton choice is calculated, the resulting distribution is not significantly different from the expected binomial distribution if all individuals choose the singleton with probability .5 on every trial. It is significantly different from the expected distribution under other plausible hypotheses about how a mean of .5 could arise. As children age, a subgroup who reliably chooses the singleton emerges. About 20% of 3-year-olds and half of 4-year-olds seem to deploy modal concepts (Leahy et al., 2022; Leahy, 2024, 2025; Leahy et al., 2025). Children who do not deploy modal concepts choose the singleton with probability .5 on every trial.

Subsequent work showed that children who do not deploy modal concepts on the 3-cups task deploy an algorithm called *minimal representation of possibility*. They simulate one trajectory that the prize in the pair could have taken, and adopt the result of that simulation (e.g., a prize in the left member of the pair) as the fact of the matter. They do not deploy a modal concept to flag that state as merely possible; they do not differentiate their belief that the left member of the pair holds a prize from their belief that the singleton holds a prize. So they have beliefs about the locations of each prize; they choose at random between those two locations. This yields the observed mean (50%) and the observed binomial distribution centered on .5 (Leahy et al., 2022).

One candidate explanation for *why* children deploy minimal representations of possibility is that these children do not have modal concepts that allow them to distinguish what is from what merely might be (Leahy & Carey, 2020). The

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

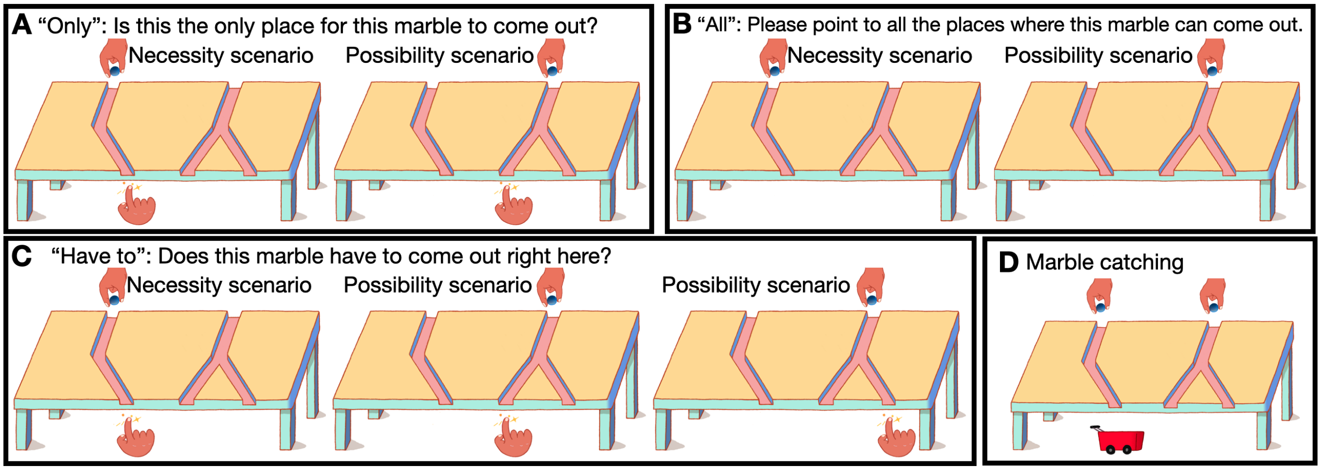

Figure 1: Components of the test phase. A. The 2 "only" questions. B. The 2 "all" questions. C. The 3 "have to" questions. D. A marble-catching trial with the wagon placed so as to guarantee a catch.

proportion of children who have developed modal concepts increases with age, accounting for age-related improvement on these tasks.

A challenge for this hypothesis is that children begin producing modal adverbs like "maybe" around their second birthdays, and modal auxiliaries like "can" a few months later. If children understand this modal vocabulary as adults do, they must have the concepts that those words express. Do children understand modal language?

Production does not guarantee comprehension, as shown by the case of number words. Many children produce number words for years before they draw appropriate distinctions between the meanings of those words (Wynn, 1992). Comprehension studies provide a clearer picture of children's conceptual resources because researchers can probe what children mean by modal words. Cournane and colleagues (2025) found indications that 3- to 5-year-olds understand what can and what cannot happen: when the puppet Logan said that something impossible can happen, participants said Logan was being silly. They did not say Logan was being silly when he said that merely possible events or necessary events can happen. This suggests that they understand "can". However, a separate group of children gave the *same pattern of responses* to questions about what "has to" happen. In particular, they said that Logan was *not* being silly when he claimed that a merely possible event had to happen. Similar results are reported by Leahy & Žalnieriūnas (2021): 4-year-olds answer questions about what "can" happen the same way adults do, and they give the same pattern of responses to questions about what "has to" happen.

The literature provides two hypotheses for why children seem to understand "can" and provide the same pattern of responses when probed with "have to". On the *force error* hypothesis children understand that modal auxiliaries quantify possibilities. They have correctly mapped "can" to a possibility modal. However, they incorrectly treat "have to" as a possibility modal as well, due to the *subset problem* (Cournane et al., 2025): Whenever *every* possibility has some property, then *some* possibility has that property. Children who map "have to" to an existential quantifier over possibilities never get counterevidence that tells them they got the mapping wrong from unembedded modal claims. While counterevidence could be found when modal claims are embedded under negation, modals often have queer scoping behavior that may make it difficult to learn the meanings of modals from their behavior when negated.

On the *minimal representation of possibility* hypothesis children answer modal questions by evaluating the single possibility the question asks about. They check whether they can simulate the possibility described by the question. If they can, they answer affirmatively; if not, they answer negatively. Repeatedly applying this algorithm in response to a sequence of questions about incompatible possibilities will result in the child evaluating incompatible possibilities in sequence without ever holding multiple incompatible possibilities in mind. This would generate adultlike responses to "can"-questions and the *same* pattern of responses to "have to"-questions. When children are asked about possible or necessary events, the simulation goes through, and they answer affirmatively, no matter whether the verb used is "can" or "have to". When they are asked about an impossible event, the simulation does not go through, and they answer negatively, no matter whether the verb used is "can" or "have to". The algorithm does not require evaluating more than one possibility at a time, so it can be implemented without deploying modal concepts.

The current paper teases these two hypotheses apart by giving the same children (1) tests of the ability to quantify possibilities using "all" and "only" and "have to" and (2) a variant of the 3-cups task known as the 3-slides task (Figure 1D). Under the force error hypothesis, children should be able to quantify possibilities using "all" and "only", since we already know that by age 4 children can use those words to quantify concrete objects (Barner, Chow, & Yang, 2009; Barner, Brooks, & Bale, 2011). And since these children all have the modal concepts needed to contrast sure things with mere possibilities, no relationship is predicted between the comprehension measures and the 3-slides task. But under the minimal representation of possibility hypothesis, children have a hard time quantifying possibilities, no matter the vocabulary used. Children will either show comprehension on all three vocabulary items or on none of the three. And children who make errors on the vocabulary questions will deploy minimal representations of possibility on the 3-slides task: they will place the wagon under the nonbranching slide 50% of the time, with individual proportions binomially distributed around .5. Children who answer questions correctly will place the wagon under the nonbranching slide more than 50% of the time, significantly more than children who make mistakes on the vocabulary batteries, and individual proportions will not be binomially distributed around .5.

Methods

Participants were 57 4-year-olds, recruited through our lab database and through childrenhelpingscience.com. Four children were excluded for parental interference, 4 for refusing to complete the study, and 1 for experimenter error. The final sample was 48 children (22 f; mean age 4.48 years, sd = .24 years, range = 4.03-4.92 years). Participants received a $5 gift card. Children were tested by a female experimenter in online zoom sessions.

Stimuli and Procedure Participants played a game with an experimenter, modified from Leahy (2023) and extending Leahy & Žalnieriūnas (2021). The experimenter used an animated hand to drop marbles onto tables, and participants tried to catch them. She also controlled virtual pointer fingers that could be used to draw attention to any exit from the table she chose. Each table had channels carved into its surface that would guide the marble's path; participants placed a small wagon under the table to try to catch a marble when it rolled off the table (Figure 1D). Some channels were nonbranching, so that it was possible to predict where the marble would come out. Other channels branched, so that it was not possible to predict where the marble would come out.

Each table that had a nonbranching slide was constructed in two versions. Each was a mirror image of the other, with a change in color. The experimenter would switch between these tables from trial to trial, so that the nonbranching channel alternated sides from trial to trial. This allows us to distinguish adultlike responding from side biases that are accidentally correct. The experimenter could use a keystroke to control the path of any marble dropped into a branching slide. The path of each marble was pseudorandomized in advance. When a marble landed in the wagon, an animated monkey came to retrieve it, giving the participants a wink and a wave. This was intended as a reward for catching the marble, and children clearly enjoyed it.

The study had a training phase and a test phase. Tables used in the training phase are not shown in Figure 1. The training phase was identical to that in Leahy and Žalnieriūnas (2021): Children learned how to put out a wagon to catch marbles in a simple, deterministic situation; they were familiarized with branching slides and were shown that a marble can come out of either branch; finally they were introduced to the test apparatus and told that one marble follows a deterministic path ("This marble (holding a marble above the nonbranching entrance) will come out here (pointing to the exit from the nonbranching slide)!") while the other follows an indeterministic path ("This marble (holding a marble above the branching entrance) can come out here or here (pointing to each exit from the branching slide)!").

The test phase had 3 subparts. First children saw 6 marble-catching trials (Figure 1D). Second was the question phase; children were asked 2 "only" questions and 2 "all" questions (4 counterbalanced orders). For the "only-necessity scenario" question (Figure 1A), the experimenter held a marble above the nonbranching slide, indicated its exit, and said, "I'm going to drop this marble in here right now. Is this (pointing to the exit from the nonbranching slide) the only place for this marble to come out?" For the "only-possibility scenario" question, the experimenter held a marble above the branching slide and said, "I'm going to drop this marble in here right now. Is this (pointing to the exit from the branching slide at the center of the table) the only place for this marble to come out?". For the "all-possibility scenario" and "all-necessity scenario" questions (Figure 1B) the experimenter held a marble above a slide and said, "I'm going to drop this marble in here right now. Please point to all the places where this marble can come out."

After the "all" and "only" questions, children answered 3 "have to" questions (1C). For the "have to-necessity scenario" question (Figure 1C) the experimenter held a marble above the nonbranching slide and said, "I'm going to drop this marble in here right now. Does this marble have to come out right here (indicating the exit from the nonbranching slide)?" For the two "have to-possibility scenario questions, she held a marble above the branching slide and asked the same question, indicating one of the exits from the branching slide. This was repeated for the other possible exit.

We did not include "can"-questions in the current study or "have to" questions in impossibility scenarios because 4-year-olds are at ceiling on "have to" questions in impossibility scenarios and on "can"-questions in all scenarios (Leahy & Žalnieriūnas, 2021). After answering all of these questions, children saw 6 more marble-catching trials (Figure 1D).

Results

First we assess comprehension of "have to". The question was, "I'm going to drop this marble right here. Does this marble have to come out right here?". Most responses were either "yes" or "no" (85%). These coded as "correct" in necessity and possibility scenarios, respectively; otherwise they were

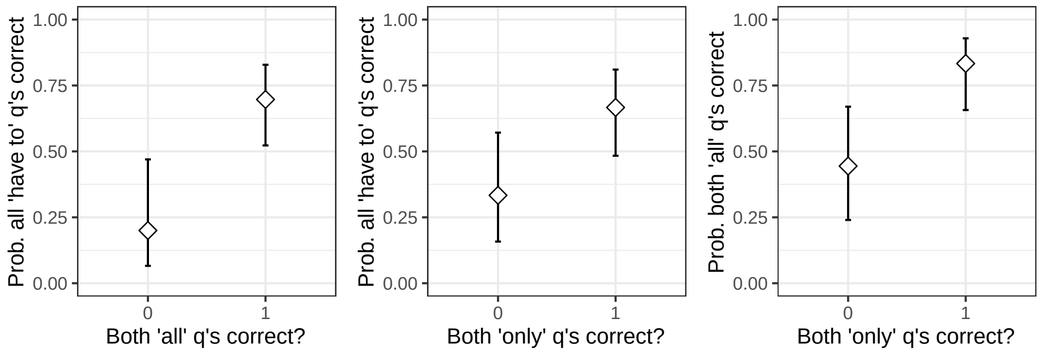

Figure 2: Pairwise relationships between batteries of questions. Left: performance on the "all" battery significantly predicts performance on the "have to" battery. Center: performance on the "only" battery significantly predicts performance on the "have to" battery. Right: performance on the "only" battery significantly predicts performance on the "all" battery.

coded as incorrect. On 12 possibility scenario questions, the participant used a possibility modal or other expression that expressed awareness that there were two possibilities (e.g., "Or right here!"). These were coded as "correct". One child said, "Yes, maybe" in a necessity scenario. We coded this as correct: "yes" is correct, and while "maybe" is not maximally informative, it is not false. Three "I don't know" responses were given in necessity scenarios; these were coded as incorrect. Three times in possibility scenarios children offered predictions about where the marble would come out (e.g., "It's gonna come out this way"). These were coded as incorrect. Three nonresponses (all in necessity scenarios) were coded as incorrect (e.g., "Let's see"). Responses in the necessity scenario were 90% correct. Responses in the possibility scenarios were 69% correct. Note that if all participants have a belief about where the marble will come out in the possibility scenario, we would expect 50% correct responses in the possibility scenarios, for they would incorrectly say the marble *has to* come out from their chosen exit but correctly *deny* that it has to come out from the other possible exit. So it is important to check how these 69% correct responses break down over participants.

The 27 children (56%) who were coded as correct on all 3 questions have denied that both possible events have to happen (i.e., they know that neither has to happen) while affirming that a necessary event has to happen. This is the strongest evidence of comprehension of "have to" in children this young currently available. The most common error (9 children, 19%) was responding affirmatively to all three questions. Note that this is unlikely a simple yes-bias: Leahy & Žalnieriūnas (2021) also asked "have to" questions in impossibility scenarios, where 4-year-olds almost always said "no". Rather, as in (Leahy & Žalnieriūnas, 2021) and (Cournane et al., 2025), it seems that most 4-year-olds who don't understand "have to" use the same algorithm to answer both "can" and "have to" questions. Children's responses to "all" and "only" questions will shed light on what that algorithm is.

For completeness: 3 of the remaining children responded as though they knew where the marble would come out on the branching side, 3 responded "no" to all 3 questions, and 7 gave response patterns that were hard to interpret (e.g., denying that a marble dropped into the nonbranching slide has to come out from the nonbranching slide).

Next we assess the ability to use "all" to quantify possibilities. Thirty-three children (69%) answered both "all" questions correctly: When asked to indicate all the places where a marble dropped into the nonbranching slide could come out, they indicated the exit from the nonbranching slide. When asked to indicate all the places where a marble dropped into the branching slide could come out, they indicated both exits from the branching slide. The most common error (6 children, 13%) was indicating exactly one exit from the branching slide when asked about the branching slide, as though there were only one place for that marble to come out (while indicating the exit for the nonbranching slide when asked about the nonbranching slide). This error is suggestive of minimal representation of possibility. Also common (4 children, 9%) was indicating all 3 exits for both questions. This might arise, for example, if children just think their job is to indicate all the places for marbles to come out, without restricting attention to where the marble is being dropped in.

Next we assess the ability to use "only" to quantify possibilities. Thirty children (62%) answered both "only" questions correctly: they affirmed that the bottom of the nonbranching slide was the only place for a marble dropped into the nonbranching slide to come out and denied that the center exit was the only place for a marble dropped into the branching slide to come out. The most common error (10 children, 21%) was to answer both questions affirmatively; this pattern is expected of children who deploy minimal representations of possibility. Four children (8%) answered both questions negatively; 3 children responded "I don't know" to one or both questions, and one participant indicated a second "possible" exit for the marble dropped into the nonbranching slide.

The most common error in all three cases suggests that children are deploying minimal representations of possibility.

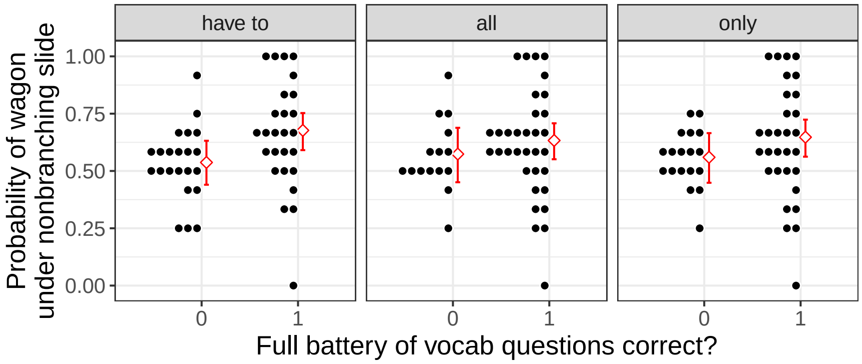

Figure 3: Relationships between answering question batteries correctly and catching marbles. Each facet shows one battery. On the x-axis is whether all questions in that battery were answered correctly. On the y-axis is the probability of placing the wagon under the nonbranching slide. Error bars are 95% CIs. Small black dots are individual proportions.

On the force error hypothesis, children should make mistakes on "have to" questions, not "all" or "only" questions. On the minimal representation of possibility hypothesis, the challenge is in quantifying possibilities, no matter which vocabulary item is used; performance on all 3 vocabulary items should pattern together. To test this three GLMs were fit, testing all 3 pairwise relationships. As Figure 2 shows, performance on the "all" battery predicts performance on the "have to" battery (logOR 2.22, 95%CI[0.85, 3.86], p = .003); performance on the "only" battery predicts performance on the "have to" battery (logOR 1.39, 95%CI[0.18, 2.68], p = .03); and performance on the "only" battery predicts performance on the "all" battery (logOR 1.83, 95%CI[0.54, 3.25], p = .007). Indeed, the two most common patterns were correct on all 3 batteries (19 children, 40%) and mistakes on all 3 batteries (8 children, 17%). The three verbal measures of the ability to quantify incompatible possibilities pattern strongly together. Some children fail to quantify possibilities, no matter which vocabulary item we use to formulate the question.

Next we evaluate the relationship between our language comprehension measures and the 3-slides task. First we analyze how success on the marble-catching task distributes over participants. Black dots in figure 3 visualize these distributions. In all 3 panels (have to, all, and only), among the children who make a mistake on the vocabulary battery, the distribution is not significantly different from the expected distribution if all of those participants choose the nonbranching slide with probability .5 on every trial (Multinomial tests: have to: p = .10; all: p = .11; only: p = .51). This is consistent with deployment of minimal representation of possibility. Among children who answer the complete battery correctly, the distributions are significantly different from the expected distribution if all of those participants choose the nonbranching slide with probability .5 on every trial (multinomial tests: have to: $p < .001$; all: $p < .001$; only: $p < .001$). In all three cases, children who fail to quantify possibilities verbally seem to deploy minimal representations of possibility when catching marbles. Some of the children who answer the vocabulary questions correctly are not deploying minimal representations of possibility.

Note that the lack of a significant difference between the observed and expected distributions among children who make mistakes on the vocabulary batteries is not due merely to the small sample size. The observed distributions *do* all differ, for example, from the distribution expected under the hypothesis that 50% performance arises because 75% of children choose randomly and 25% of children always choose the singleton (multinomial tests: have to: p = .009; only: p = .04; all: p = .04), which is another way one might expect to generate a mean of .5. (The proportions in this mixture are chosen because (.75 * .33) + (.25 * 1) = .5, where .33 is the rate of choosing the nonbranching slide among the 75% of children who choose randomly and 1 is the rate of choosing the nonbranching slide among the remaining 25% of children.)

Next we analyze the means (Fig. 3, red points). We fit 3 GLMMs, one for each vocabulary battery, predicting the probability of placing the wagon under the nonbranching slide from whether that vocabulary battery was answered correctly or not with a random intercept for participant. Among children who make mistakes on the vocabulary items, none of the estimates are significantly different from .5 (have to: logOR 0.15, p = .45; all: logOR 0.30, p = .24; only: logOR 0.24, p = .29). Among children who answer a battery correctly, all of the estimates are significantly greater than .5 (have to: logOR 0.74, $p < .001$; all: logOR 0.55, p = .002; only: logOR 0.61, $p < .001$). And children who answer "have to" questions correctly are significantly more likely to place the wagon under the nonbranching slide than children who make a mistake on the "have to" battery (logOR 0.59, 95% CI [0.05, 1.13], p = .03). The predictions of the minimal representation of possibility hypothesis are borne out in fine detail.

Discussion

Many 4-year-olds make errors in quantifying possibilities, no matter which vocabulary item is used. But 4-year-olds already understand "all" and "only" and use them correctly to quantify concrete individuals. Their problem is in quantifying possibilities. Furthermore, the most common errors on the "all" and "only" batteries suggest that they deploy minimal representations of possibility. Finally, for all three batteries of questions, children who make a mistake on the battery seem to deploy minimal representations of possibility on the 3-slides task. Children who answer the batteries correctly do not. All of these findings speak for the minimal representations of possibility hypothesis over the force error hypothesis.

Half of the 4-year-olds in the current sample showed comprehension of "have to", converging with estimates of the proportion of 4-year-olds who bring modal concepts to bear on nonverbal tasks (Leahy, 2024, 2025; Leahy et al., 2025). These children are significantly more likely to put the wagon under the nonbranching slide in the marble-catching task than children who make mistakes on "have to" questions. As a group they show no signs of minimal representations of possibility. These children seem to bring modal concepts to bear in both language and behavior.

This addresses an open question in the literature, as previous studies have found much lower rates of comprehension of "have to": Only 15% of the 4-year-olds reported by (Leahy & Žalnieriūnas, 2021), and 85% incorrect responses in possibility scenarios in Cournane et al. (2025). Two procedural differences may account for these differences. First, we used "all" and "only" questions to scaffold performance on "have to", using vocabulary that we know these children understand to draw their attention to the facts they need to think about. Second, the questions in the current study were posed in a slightly different way: In the study reported by Leahy & Žalnieriūnas (2021), the experimenter held a marble above an opening and asked, "Does this marble have to come out here [indicating an exit]?" Children's responses raised concerns that they interpreted the question as "Does it *ever* have to come out here, over repeated drops?". For that reason we altered the question to be clear that we were asking about this *specific* drop: "I'm going to drop this marble in here right now. Does this marble have to come out right here?"

A limitation of the current study is that while children who answered "all" and "only" questions correctly placed the wagon under the nonbranching slide significantly more that 50% of the time, they did not place the wagon under the nonbranching slide significantly more often than children who made mistakes on "all" and "only" questions, respectively. That relationship was observed in the case of "have to". Furthermore, numerically more children were correct on the "all" battery (69% correct) and on the "only" battery (62% correct) than on the "have to" battery (56% correct). Notably, the "all" and "only" questions specifically ask about *places*. These data might be explained if some children who cannot quantify possibilities succeed on the "all" and "only" batteries because they quantify places, i.e., parts of the actual world. Alternatively, some children may figure out what "all" and "only" mean before they solve the mapping problem for "have to". One way to address this issue will be to follow up with younger children, to see if quantifying possibilities is easier with "all" and "only" than with "have to", and whether comprehension of any of these predicts performance on the marble catching task.

The current study does not address the growing body of research showing that infants and preschoolers make quite sophisticated inferences (recent examples include Alderete & Xu, 2023; Andreuccioli et al., 2024; Gautam, Suddendorf, & Redshaw, 2021; Cesana-Arlotti, Jáuregui, Mazalik, Nichols, & Halberda, 2025; Cesana-Arlotti & Halberda, 2024; Bleijlevens & Behne, 2024; Turan-Küçük & Kibbe, 2024, 2025a, 2025b, and many more). Children may draw on modal concepts in making these inferences. On the other hand, it is impossible to say how a participant formulates their premises when the only data are the conclusions they draw, since any conclusion can be drawn from many different sets of premises. Children might make these inferences without deploying modal concepts. Space precludes proper discussion of those ideas. A limitation of the current work is the lack of a close control showing that children succeed in a paradigm that is closely matched for task demands, and that they fail exactly when they need to represent two versions of one state of affairs. We are currently developing such a paradigm.

Conclusion

Recent studies have found that comprehension of "can" is near-universal among 4-year-olds (Leahy & Žalnieriūnas, 2021; Cournane et al., 2025), and in one study 40% of 36-month-olds showed comprehension on a battery of "can" questions (Leahy, 2025). "Have to" seems to be much harder for children to understand, with only about half of 4-year-olds showing comprehension. Previous data were consistent with the hypothesis that children were capable of quantifying possibilities, but erroneously believed that "have to" expressed an existential quantifier. The current data speaks clearly against that interpretation: the 4-year-olds who cannot quantify possibilities using "have to" also struggle to quantify possibilities using "all" and "only", words that we know they understand by age 4. That is, half of 4-year-olds have a hard time quantifying possibilities, no matter which language we prompt them with. And those same children seem to deploy minimal representations of possibility on the 3-slides task, failing to differentiate the sure thing from the mere possibilities. These children have given no sign that they see multiple possibilities for a marble dropped into a branching slide.

If these children really do not see that there are multiple possibilities, if they lack modal concepts that are necessary for keeping track of incompatible possibilities, then children produce both "can" and "have to" before they really know what those words mean. They may have mapped both vocabulary items to minimal representations of possibility, a precursor to modal concepts. The next question becomes, What is the process by which they differentiate the meanings of those words, and how is that process related to the ontogenesis of modal concepts?

References

Alderete, S., & Xu, F. (2023). Three-year-old children's reasoning about possibilities. *Cognition*, *237*, 105472.

Andreuccioli, L., Mazor, S., Begus, K., Bonawitz, E., Denison, S., & Walker, C. M. (2024). Young children adapt their search behavior for necessary versus merely possible outcomes. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Barner, D., Chow, K., & Yang, S.-J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive psychology*, *58*(2), 195–219.

Bleijlevens, N., & Behne, T. (2024). Young children and adults use reasoning by exclusion rather than attraction to novelty to disambiguate novel word meanings. *Developmental Psychology*, *60*(7), 1313.

Cesana-Arlotti, N., & Halberda, J. (2024). A continuity in logical development: domain-general disjunctive inference by toddlers. *Open Mind*, *8*, 809–825.

Cesana-Arlotti, N., Jáuregui, S., Mazalik, P., Nichols, S., & Halberda, J. (2025). Logical concepts of (im) possibility guide young children's decision-making. *Developmental Science*, *28*(5), e70044.

Cournane, A., Dieuleveut, A., Repetti-Ludlow, C., & Hacquard, V. (2025). Word learning challenges explain nonadult possibility language comprehension in preschoolers. *Developmental Psychology*.

Gautam, S., Suddendorf, T., & Redshaw, J. (2021). When can young children reason about an exclusive disjunction? *Cognition*, *207*, 104507.

Leahy, B. (2023). Don't you see the possibilities? *Developmental Science*, *26*(6), e13400.

Leahy, B. (2024). Many preschoolers do not distinguish the possible from the impossible in a marble-catching task. *Joournal of Experimental Child Psychology*, *238*, 105794.

Leahy, B. (2025). Comprehension of modal vocabulary predicts performance on a nonverbal measure of explicit possibility concepts at 48 but not 36 months. *Under review*.

Leahy, B., & Carey, S. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78.

Leahy, B., Huemer, M., Steele, M., Alderete, S., & Carey, S. (2022). Minimal representations of possibility at age 3. *Proceedings of the National Academy of Sciences*, *119*(52), e2207499119.

Leahy, B., Vivanco, V., Cheyette, S. J., Smith, K. A., White, L. J., Feiman, R., ... Tenenbaum, J. B. (2025). Calculating probabilities from imagined possibilities: Limitations in 4-year-olds. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 47).

Leahy, B., & Žalnieriūnas, E. (2021). Might & might not. In *Semantics and linguistic theory* (Vol. 31, pp. 426–445).

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48.

Turan-Küçük, E. N., & Kibbe, M. M. (2024). Three-year-olds' ability to plan for mutually exclusive future possibilities is limited primarily by their representations of possible plans, not possible events. *Cognition*, *244*, 105712.

Turan-Küçük, E. N., & Kibbe, M. M. (2025a). Children's reasoning about possible outcomes of events in the present and the future. *Developmental Psychology*.

Turan-Küçük, E. N., & Kibbe, M. M. (2025b). Three-and four-year-old children represent mutually exclusive possible identities. *Journal of Experimental Child Psychology*, *249*, 106078.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive psychology*, *24*(2), 220–251.