Proceedings of SALT 31: 426–445, 2021

Might and might not: Children's conceptual development and the acquisition of modal verbs*

Brian Leahy
Harvard University

Eimantas Žalnieriūnas
University of Bath
Harvard University

Abstract When a child acquires her first epistemic modal vocabulary, is she learning how to map words in the language she is learning onto preexisting concepts of possibility, necessity, and impossibility? Or must she also construct the concepts themselves? If the concepts are constructed, does learning to talk about possibilities play a role in the construction process? Exploring this hypothesis space requires testing children's comprehension of modal vocabulary alongside nonverbal tests of their modal concepts. Here we report a study with 103 children from 4;0 through 7;11 and 24 adults. We find evidence of comprehension in only a few 4-year-olds. We argue that the data fit best with the hypothesis that acquisition of modal language and development of modal concepts proceed hand-in-hand. However, more work is needed to refine our methods and extend them to 3-year-olds.

Keywords: modal acquisition, conceptual development, development of modal concepts

1 Introduction

Modal vocabulary helps us communicate with one another about the modal status of states or events, to mark them as necessary, possible, or impossible. Since we can think about modal propositions without talking about them, our cognitive systems must also have tools that track the modal status of states and events, to mark them as necessary, possible, or impossible. We will call those tools *modal concepts*; the tools that cognition uses to mark states and events as merely possible are *possibility concepts*. This paper will explore the relationship between learning to talk about possibilities using possibility vocabulary and learning to think about possibilities using possibility concepts. We will argue that children use possibility language before they develop possibility concepts. But our claim should not be construed too broadly. We do not make claims, for example, about whether young children understand the root modals their parents use to give and withold permission. Our

* We thank Susan Carey, Scarlett Close, Ailís Cournane, Kathryn Davidson, Stephen Ferrigno, Michael Huemer, Sabine Iatridou, Sydney Levine, Jesse Snedeker, and Nancy Soja.

claim is that children use possibility language before they can differentiate states and events that obtain from states and events that merely might obtain, that is, from states and events that cannot yet be ruled in, but cannot be ruled out either.

When do children start thinking about possibilities? We know from looking time experiments that infants have expectations, and hence represent states that are merely possible (e.g., Baillargeon, Spelke & Wasserman (1985)). Yet behavioral measures have shown striking failures to reason with possibilities in preschoolers younger than 4 (Leahy & Carey 2020). If an experimenter drops a marble into a tube shaped like an upside-down Y, 2- and 3-year-olds who want to catch the marble often cover just one branch, even though it might go out the other way (Redshaw & Suddendorf 2016). In the 3-cups task (Mody & Carey 2016), participants see a stage with 3 cups, organized into a pair and a singleton. One of the sets–say, the singleton–is occluded, and a sticker is hidden there. Then the pair is occluded, and a sticker is hidden in one of those cups. The occluder is removed, and participants are given one chance to pick a cup and receive its contents. The smart choice is the singleton cup, since it certainly holds a sticker, while the other cups are 50/50 risks. Over several replications and manipulations, strong convergence has been found: 2.5-year-olds choose the singleton half of the time; 3-year-olds 60% of the time (Grigoroglou, Chan & Ganea 2019, Leahy Submitted). Children who have modal concepts should always pick the singleton cup. Why is performance so poor? Why do they fail so abjectly at these tasks, which seem to measure children's ability to represent alternative possibilities, if representing possibilities is an innate ability?

One explanation for children's poor performance is that the ability to represent possibilities is not innate; most children younger than 4 lack possibility concepts (Leahy & Carey 2020). They can only represent the world as it is, not as it merely *might* be. They can extrapolate, make predictions, simulate outcomes, and make guesses. We will say that they can *make assumptions*. Their assumptions yield expectations, and those expectations can be violated. But their beliefs are all about how the world *is*, never about how it merely *might* be.

Lacking possibility concepts, these children cannot recognize that a marble dropped into a Y-shaped tube *might* come out the left and *might* come out the right. So they make an assumption, and act accordingly. Thus they only cover one branch. Nor do they understand that a sticker hidden in the pair of cups might *and might not* be in the left member of the pair, and might *and might not* be in the right. So they make an assumption about which cup it went into, and treat that assumption as fact. Then the participant "knows" each sticker's location (the singleton, and their chosen member of the pair); they choose at random between the two cups that they "know" hold a sticker. Over many trials, they choose the singleton about 50% of the time. Three-year-olds' consistent 60% performance is expected if a small proportion have developed possibility concepts.

While their assumptions represent states that are merely possible, those representations are not marked as merely possible. Mere assumptions are not distinguished from representations derived from more trustworthy sources like perception. *Qua* representations of states that *are* merely possible but *are not marked* as merely possible, we call these 'Minimal representations of possibility'.

Comprehension of modal verbs A response to this proposal is that many children use epistemic possibility language by their second birthdays (Van Dooren, Dieuleveut, Cournane & Hacquard 2017, Dieuleveut, van Dooren, Cournane & Hacquard 2019, Cournane 2020, Van Dooren, Dieuleveut, Cournane & Hacquard Forthcoming). If children have mastered this vocabulary in an adult way, then they have possibility concepts, and the proposal is falsified.

However, it is not clear that these children have mastered possibility language in an adult way. Do 2- and 3-year-olds differentiate necessity modals from possibility modals? Does a 2-year-old who says, "Maybe we get ice cream" understand that maybe we won't get ice cream either? By analogy, consider that a 2-year-old, when asked to give all the crackers, typically hands them all over; does she understand 'all'? Perhaps not, since when asked to give some of the crackers, she also gives all of them (Barner, Chow & Yang 2009).

Assessing whether children have correctly analyzed their epistemic modal vocabulary requires comprehension studies. This section reviews existing comprehension studies. Signs of comprehension in children younger than 4 are virtually absent. Existing studies rarely report findings with 3-year-olds; when 3-year-olds have been tested, findings have been either a failure to differentiate modals or, in one case, weak differentiation that could be generated by the small proportion of 3-year-olds that we think have developed possibility concepts. This section, then, shows that existing literature is compatible with the proposal in (Leahy & Carey 2020).

In one paradigm, participants observe a situation and then answer a modal question or evaluate a modalized description of that situation. For example, in Öztürk & Papafragou (2015) Exp. 1, 4.5- to 6-year-olds saw a stage with a yellow box and a purple box. A toy entered the stage, the curtain fell, and the toy was hidden in one of the boxes. The curtain was raised and participants were asked a question. When asked "Agree or disagree: The toy may be in the yellow box", they agreed 86% of the time. When asked, "Agree or disagree: The toy has to be in the yellow box" they agreed only 47% of the time. Four and a half- to 6-year-olds differentiated these words. However, the 47% agreement rate with "The toy has to be in the yellow box" when it was possible that the toy was in the purple box suggests that many children have not differentiated the words. If not, it is not clear that they understand either verb. Other studies in this paradigm have found evidence that even at age 7 children did not always deny necessity claims when they were false (Noveck 2001).

A third study using this paradigm did not reveal comprehension of 'must' and 'may' in 4.5- to 5.5-year-olds, though an eye tracking measure revealed that at least some of these participants differentiated these modals (Moscati, Zhan & Zhou 2017).

The question of interest for this paper is whether more than a handful of 3-year-olds differentiate possibility modals from necessity modals and understand that 'Maybe p' is compatible with 'Maybe not-p'. The studies reviewed above do not provide clear evidence that more than a handful of 4-year-olds comprehend their modal vocabulary. Is performance so poor because only a few 4-year-olds comprehend modal vocabulary? Or are the measures insufficiently sensitive, allowing performance factors to mask children's competence? Addressing this question requires formulating hypotheses about what the problematic performance factors might be. One potential limiting factor is that children might not notice the necessary contrasts. For example, after seeing that a toy was hidden in either the yellow or the purple box, they might agree with "The toy has to be in the yellow box" because they don't notice the contrasting possibility that it's in the purple box. This issue is addressed by studies in a second paradigm, which scaffold participants' performance by providing explicit contrasts.

In an early study with this structure, Moore and colleagues (1990) showed participants a stage with a red and a blue box; one puppet said, "The candy might be in the blue box" while another said, "The candy must be in the red box". When asked to choose a box to get what's inside, 3-year-olds chose the red box exactly half of the time, but 4-year-olds chose it significantly more often. Identical results were found for the contrast between "It must be in the red box' and "It could be in the blue box". By age 5 children chose the red box when one puppet said, "It is probably in the red box" and the other said "It is maybe in the blue box", and by age 6 they picked the red box when one puppet said "It is probably in the red box" while the other said "It is possibly in the blue box". Again, at least some 4-year-olds differentiate between at least some possibility modals; more subtle differentiations were observed only at later ages. But there is no sign of differentiation among 3-year-olds.

Using a similar design, Öztürk and Papafragou (2015, Exp. 2), showed participants a stage with a yellow box and a pink box. A toy horse had been hidden in one of the boxes. The yellow box was opened and found empty. Two characters then described the scene: one said, "The horse may be in the pink box". The other said, "The horse has to be in the pink box". Participants were asked which character gave a better answer; 4.5- to 6-year-olds selected the character who said "The horse has to be in the pink box" 66% of the time, significantly more than expected by chance, though still far from ceiling. Providing participants with the contrast between 'may' and 'has to' may have helped them see what's good about 'has to' in this context. Another study in the same paper showed that some 4- to 5.5-year-olds differentiate what *might be* the case from what *is* the case. But 3-year-olds were not tested in these

studies, so we cannot draw conclusions about how many 3-year-olds comprehend modal vocabulary.

In (Shtulman & Carey 2007), 4-year-olds were asked whether impossible events (eating lightening for dinner), improbable events (eating pickle flavored ice cream) and familiar events (wearing a baseball hat) "could happen in real life" and "would take magic to happen". Performance was adultlike for impossible events and familiar events, but improbable events were treated like impossible events. Their understanding of what is possible is again not adultlike; it became increasingly adultlike with age, though even 8-year-olds showed some deficiencies. Scaffolding their performance by providing a forced choice between an image of an improbable event and an impossible event and asking which can happen in real life resulted in improved performance in 4-year-olds, but the effect was still not large, as would be expected if only some 4-year-olds properly evaluate what can happen in real life. Again, this study did not report results with 3-year-olds.

To our knowledge, (Armstrong 2020) reports the only study that finds any comprehension of modals in 3-year-olds. Participants were asked to decide which of two characters should be invited to a birthday party. When one character said they might give broccoli as a gift while the other said they will give broccoli as a gift, 3-year-olds chose to invite the person who merely might give broccoli. However, the effect was very small, and could be driven by a small proportion of 3-year-olds who have developed possibility concepts and learned the corresponding possibility vocabulary. Moreover, when one character said they will give a copy of the Frozen DVD and the other said they might give a copy of the Frozen DVD, 3-year-olds chose to invite both characters equally often. At age 4 there is a significant but still small effect for both postively and negatively valenced items. At age 5 the effect sizes become substantial, suggesting that a large proportion of 5-year-olds are differentiating the verbs.

We want to know when children differentiate possibilities from necessities, and when they understand that 'maybe p' is consistent with 'maybe not-p'. Studies with scaffolding like Armstrong's might be thought to address this issue since it is not clear why someone would prefer, for example, a person who might give you broccoli over a person who will give you broccoli unless you understand that the first person might not give you broccoli. However, a shortcoming of this design is that participants might also prefer a person who *dax* give you broccoli over a person who will give you broccoli, where 'dax' is a nonce word. The participant might simply understand that broccoli is forthcoming from one person, and choose against them. It is not clear, from the fact that participants prefer the person who *might* give broccoli, that they understand that that person *might not* give broccoli.

This review finds little evidence that many children have differentiated epistemic modal forces before age 4. There is one study where at most a small proportion of

3-year-olds differentiated 'will' and 'might', and more work is required to show that even those 3-year-olds understood that "I might give you broccoli" is consistent with "I might not give you broccoli".

The absence of studies showing comprehension in 3-year-olds is striking. However, it is unclear whether 3-year-olds are rarely studied or if they have been studied and nonsignificant results have gone unreported. It is also possible that existing methods are not appropriate for such young children, who might need even more scaffolding and support to demonstrate their competence.

So it remains an open question which children differentiate necessity modals from possibility modals, and which children understand that 'p *can* happen' is consistent with 'p *can fail* to happen'. The experiment described below addresses this issue for children aged 4 through 7, scaffolding performance by repeating questions about contrasting *situations*, rather than contrasting modal verbs about the same situation. We add additional scaffolding by giving children a fun, interactive experience with an apparatus that should make the possibilities, impossibilities, and necessities apparent to anyone with the tools to notice them. This also tests the same children with a nonverbal measure of their modal concepts. Future work will extend this project to 3-year-olds.

Experimental prospectus We developed an apparatus that tests children's comprehension of 'can' and 'have to' while also testing their modal concepts nonverbally.[1] Scaffolding was provided in two ways. First, we used a simple and intuitive apparatus. Children had extensive opportunity to interact with the apparatus in a fun game before they were asked any modal questions; we gave them every opportunity to notice the possibilities that the apparatus affords. Second, for each modal verb we tested, we asked multiple questions about a single event. Our goal was to draw children's attention to multiple possibilities by asking multiple questions about the same event. The questions should draw participants' attention to each relevant possibility. This should help them to answer the questions correctly, and to help them to solve the nonverbal problem (if they haven't already).

The resulting experiment compares two hypotheses about the relationship between the acquisition of modal language and the development of modal concepts.
Hypothesis 1. Learning modals is a matter of mapping words onto concepts that are either innate or early-developing, constructed prior to production of modal language. Children will solve the nonverbal problem regardless of their comprehension of modal verbs.

1 Almost all existing experimental studies for children's acquisition of modal force have held modal flavor fixed, and have typically used epistemic modal flavor. The study we report here is not clearly either epistemic or root flavor; both epistemic and root interpretations seem to be available.

Hypothesis 2. Learning modals and the development of modal concepts goes hand-in-hand. Children who do not comprehend modals will not perform well on nonverbal measures of modal concepts. Only children who comprehend modals will solve the nonverbal problem.

2 Methods

Participants Participants were recruited through our lab database (n = 94), through childrenhelpingscience.com (n = 13), and through the Harvard Psychology Study Pool. They were 26 4-year-olds (12 f, mean 4.51, sd .31), 27 5-year-olds (16 f, mean 5.27, sd .19), 24 6-year-olds (9 f, mean 6.36, sd .25), 26 7-year-olds (16 f, mean 7.47, sd .3), and 24 adults (13 f, mean = 20.8, sd = 6.47). One 4-year-old, 2 5-year-olds, and 1 6-year-old were excluded for parental interference. Two 6-year-olds and one 7-year-old withdrew consent. Participants received a $5 gift card.

Procedure We collected data via Zoom using the apparatus in figure 1. In the training phase children learned how to catch marbles in a wagon, that marbles dropped into branching slides can go out left or right, and that when they catch a marble in the wagon, a cute monkey emerges to fetch it. This was intended as a reward for catching the marble, and children clearly enjoyed seeing the monkey.

After training, participants saw the testing apparatus (Figure 1). The test phase had a catch marbles-answer questions-catch more marbles structure, to check if the questions helped children solve the problem. Participants saw 16 catching trials and answered 18 modal questions, 6 each for 'can', 'have to', and 'will'. The 'will' results failed a manipulation check–adults did not answer as we expected–and so we will not report that data here.

Before the first catching trial, participants were told that two marbles would be dropped at the same time. The experimenter said, "This marble [indicating marble over the nonbranching slide] will come out here [indicating the exit of the nonbranching slide]. This marble [indicating the marble over the branching slide] can come out here or here [sequentially indicating the exits for the branching slide]". Participants were prompted to help the monkey get a marble, and were reminded to pick the best slide to put the wagon under to be sure that the monkey would get a marble. Then they were asked, "Can you pick a slide?". The experimenter placed the wagon in the indicated location and dropped both marbles into the apparatus.

The apparatus was flipped over from trial to trial, as illustrated in figure 1. This allowed us to distinguish side biases from correct responses (side biases were not observed). The marble's path through the branching slide was pseudorandomized.

After 6 trials of marble catching, the modal questions phase began. The experimenter held a marble above one of the entrances, pointed to one of the exits,

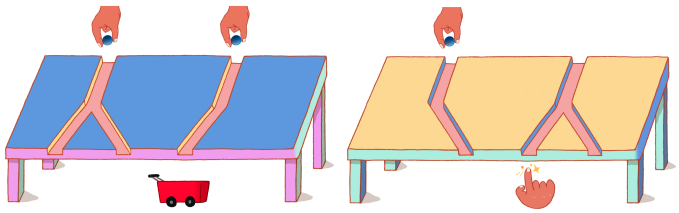

Figure 1: The main testing apparatus in its two configurations. (Left) A catching trial; the wagon is under the target slide, where it is sure to catch a marble, but the marbles have not yet been dropped. (Right) The apparatus has been flipped. A modal question trial is illustrated. The experimenter holds a marble above an entrance, points to an exit, and asks, "If I drop a marble in here, [verb] it come out here?"

and asked, "If I drop a marble in here, [verb] it come out here?" The question was repeated for all three exits, with the entrance held constant. Then the entrance was changed, and all three questions were repeated. After asking all 6 questions (2 entrances x 3 exits), children saw two trials of catching marbles. Then all 6 questions were repeated with a new verb. This was followed by 2 more trials of catching marbles, and then 6 more modal questions with the final verb. The final round of modal questions was followed by 6 more trials of catching marbles. The apparatus was always flipped between catching trials.

We grouped the 6 modal questions into 4 categories (figure 2B): is the marble held above a branching or a nonbranching entrance, and are we indicating a branching or a nonbranching exit? Three of these question-types do not distinguish among verbs. When a marble is held above the nonbranching entrance, it can and has to come out the nonbranching exit; it cannot and doesn't have to come out either branching exit. When a marble is held above the branching entrance, it cannot and doesn't have to come out from the nonbranching exit. These questions allow us to identify children who are biased to respond 'yes' or 'no'. (No child displayed any such bias.) The fourth question-type distinguishes between the verbs. A marble dropped into a branching slide can but does not have to emerge from either branching exit. We call questions of this fourth type "discriminating questions".

This test for the comprehension of modals scaffolds children's performance by asking multiple questions about the same situation (one marble drop into a single entrance), and checking multiple contrasting outcomes, some necessary, some possible, some impossible. It tests whether children differentiate possibility modals from necessity modals, since we ask the same children both what *can* happen and what *has to* happen in the same range of situations. And finally, it asks them to identify multiple possible outcomes of a single event, to evaluate whether children

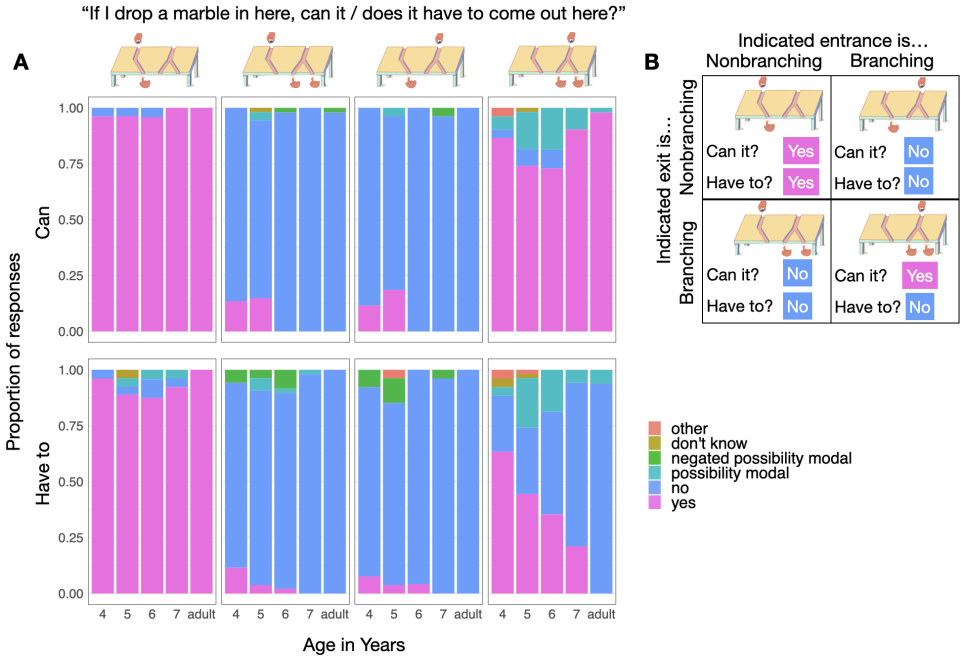

Figure 2: **(A)** Proportion of each response type, broken down by modal verb, question type, and age group. **(B)** We group the 6 modal questions into 4 categories. The figure color-codes the expected response for each question-type. Three of the question-types do not discriminate between the modal verbs, but still require participants to switch between "yes" and "no" responses; we call questions of these types "nondiscriminating questions". The fourth question-type ("discriminating questions") discriminates between the verbs.

understand that what *might* happen *might not* happen either. This is the strongest test currently available for children's comprehension of modal vocabulary.

3 Results

Analysis script and raw data are at Harvard Dataverse (Leahy & Žalnieriūnas 2021).

3.1 Analysis of modal comprehension

Descriptive analysis of modal auxiliary comprehension We coded children's responses to modal questions into 6 categories, developed post hoc: yes, no, possibility modal ("maybe", "it might", etc), negated possibility modals ("it can't"), I don't know, and other. Necessity modals and negated necessity modals were rare (5

of each). We coded these as "yes" and "possibility modal", respectively.

Figure 2 displays the descriptive proportion of each response type for each question type, verb, and age group.

Nondiscriminating questions provide a sanity check, showing us that children switch flexibly between 'yes' and 'no' when answering nondiscriminating questions. From age 4, performance was near ceiling on nondiscriminating questions.

On discriminating questions, for the verb 'can', both "yes" and "possibility modal" are pragmatically appropriate responses (cf."If I drop a marble in here, can it come out here?" "It *can*."). These response types account for more than 90% of responses to discriminating "can"-questions in all age groups. By age 4, children seem to affirm that there are two possible versions of a single event: one marble, dropped into a branching slide, might come out left and might not–it might come out right. However, before we draw this conclusion, we must check children's responses to discriminating "have to"-questions.

On discriminating questions for the verb 'have to', both "no" and "possibility modal" are pragmatically appropriate responses (cf."If I drop a marble in here, does it have to come out here?" "It *can*."). Pragmatically appropriate responses were rare among 4-year-olds (29%), though they increased with age (5-year-olds: 52%; 6-year-olds: 65%; 7-year-olds: 79%; adults: 100%). The most common error in all age groups was "yes" (4-year-olds 63% yes; 5-year-olds 44% yes, 6-year-olds: 35% yes, 7-year-olds: 21% yes, adults: no errors). The fact that for 'have to' (1) pragmatically appropriate answers are rare among 4-year-olds and (2) their errors were overwhelmingly the same responses they provided to can-questions suggests that many 4-year-olds have not differentiated 'have to' from 'can'. Further, this raises doubts about whether they really understand 'can', and hence whether they really understand that there is more than one possible version of a single event. Perhaps children have found a strategy for answering modal questions that happens to yield correct responses to 'can' questions in the absence of possibility concepts. For example, perhaps they merely check whether there is a path from the indicated entrance to the indicated exit; if so, they answer affirmatively, if not, they answer negatively. But they needn't understand that the presence of multiple paths means that there are multiple open possibilities. They can repeatedly apply the same procedure when each question is asked.

Alternatively, perhaps performance factors are masking children's competence.

The grouped data in figure 2 holds at the individual level: participants almost always gave the same response to both discriminating questions with the same verb. We discuss individual level performance in the error analysis, appendix 1.

In summary, even 4-year-olds provide pragmatically appropriate responses to 'can' questions, seeming to affirm that there are multiple places where the marble, on a single drop, can come out and other places where it cannot come out. However,

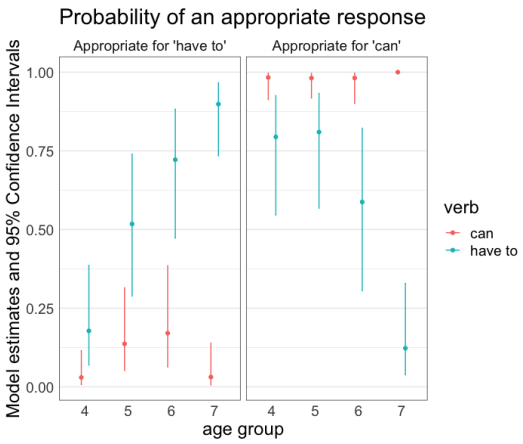

Figure 3: Probability of a have to-appropriate (*no*, *possibility modal*) and can-appropriate (*yes*, *possibility modal*) response to discriminating questions. Error bars represent 95% confidence intervals.

most 4-year-olds and many older children provide the same pattern of responses to 'have to' questions. This error decreases with age, to the point that 7-year-olds are nearly adultlike. We see two possible explanations: either most 4-year-olds and many older children do not understand 'have to' (and perhaps not 'can'), or else our task is too difficult, and fails to reveal competences that are in fact present.

Inferential analysis of modal comprehension The second analysis provides inferential statistics that help us gauge the evidence that children have differentiated 'can' and 'have to' in each age group. We tested whether children's responses to discriminating questions were more often appropriate for 'have to' questions when the verb was 'have to' than when the verb was 'can'. Similarly, we tested whether their responses were more often appropriate for 'can' when the verb was 'can' than when it was 'have to'. Results are plotted in figure 3. Since both 'no' and 'possibility modal' are pragmatically appropriate responses to the discriminating 'have to'-questions, we coded these responses as "have to-appropriate" and coded all other responses as "have to-inappropriate". We fit a random-intercept binary logistic mixed effects model to evaluate the probability of a have to-appropriate response to discriminating questions from the verb used in the question (can, have to), age group (categorical), and their interaction. The grouping variable was participant id.

Results (figure 3, left) show that 4-year-olds gave have to-appropriate responses significantly more often to discriminating have to-questions than to discriminating can-questions (odds ratio 1 / 7.14, p = .006). Some 4-year-olds appear to have differentiated 'have to' from 'can'. However, appropriate responses were rare

(estimated probability .18). The differentiation between ‘can’ and ‘have to’ became more pronounced with age, and the probability of a have to-appropriate response to discriminating questions with ‘have to’ as verb increased. These results show that while some 4-year-olds have differentiated these verbs, many have not. The proportion of children who differentiate these verbs increases with age.

‘Yes’ and ‘possibility modal’ are pragmatically appropriate responses to the discriminating ‘can’ questions. We coded these responses as “can-appropriate” and coded all other responses as “can-inappropriate”. We fit a random-intercept binary logistic mixed-effects model to evaluate the probability of a can-appropriate response from the verb used in the question (can, have to), age group (categorical), and their interaction; the grouping variable was participant id (figure 3, right). Again, even 4-year-olds gave can-appropriate responses significantly more often to discriminating questions when the verb was ‘can’ than when it was ‘have to’ (odds ratio: 15, $p = .001$). Again, this supports the view that some 4-year-olds have differentiated ‘have to’ and ‘can’. Notably, though, the probability of a can-appropriate response was quite high among 4-year-olds when the verb was ‘have to’ (estimated probability .79). Again, while some 4-year-olds showed signs of differentiating these verbs, many did not. These effects increased with age: the differentiation between ‘can’ and ‘have to’ became more pronounced,and the probability of a can-appropriate response when the verb was ‘have to’ fell.

These inferential statistics reveal evidence that at least a small subgroup of 4-year-olds have differentiated ‘can’ from ‘have to’. Most give the same pattern of responses to both ‘can’ and ‘have to’; for the majority of our participants it is not clear whether they understand either of these verbs. However, there are many features of this task that may account for 4-year-olds’ errors as performance errors rather than competence errors. We mention some of these in the discussion.

We turn to the results of our nonverbal measure of modal concepts.

3.2 Nonverbal measure of modal concepts

Group-level analysis Children who can see that there are two possibilities for the marble in the branching slide but only one possibility for the marble in the nonbranching slide should place the wagon under the nonbranching slide. If they reliably place the cup under the nonbranching slide, we infer that they see that there are two possibilities on the branching side, and they wish to avoid the attendant risks.

While there are three places where a child can place a wagon (and hence the probability of choosing the target at random is .33), we will test whether performance differs from .5. This is because we wish to compare the hypothesis that participants have modal concepts to the hypothesis that they are using minimal representations of possibility. Children using minimal representations of possibility make an as-

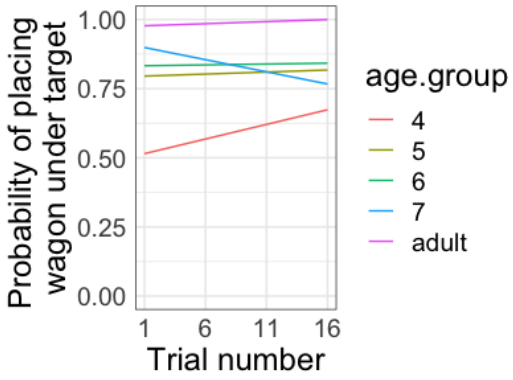

Figure 4: The probability of choosing the target across trials.

sumption about which path the marble in the branching slide will take. They do not differentiate that assumption from other beliefs, such as their belief about where the marble in the nonbranching slide will emerge. Consequently, their beliefs specify two locations where marbles will emerge; they choose at random between those locations. Over many trials, participants choose the target 50% of the time and each of the other locations 25% of the time.

Results are plotted in figure 4, broken down by age group. The probability of choosing the target was modeled with a random-intercept binary logistic mixed-effects model. Fixed effects were trial number (continuous) and age group (categorical). The grouping variable was participant id.

For 4-year-olds the estimated probability of a correct response on the first trial was .51, as predicted if all 4-year-olds use minimal representations of possibility. Their rate of increase was small and is not unlikely to have arisen through random sampling variance (odds ratio = 1.04, 95% CI [0.997, 1.10], p = .07). This analysis yields no evidence that any 4-year-olds have modal concepts. Five- and 6-year-olds chose the target more often from the first trial (5-year-olds estimated probability correct .80; 6-year-olds estimated probability correct .83); their performance did not improve across trials (5-year-olds: odds ratio = 1.01, 95% CI [0.96, 1.07], p = 0.73; 6-year-olds: odds ratio = 1.00, 95% CI [0.95, 1.06], p = 0.88). Five-year-olds were more likely that 4-year-olds to choose the target on the first trial (odds ratio = 3.66, 95% CI [1.21, 11.28], p = .02), as were older participants; however, the older groups of children did not differ significantly from one another. At least some 5- and 6-year-olds seem to have modal concepts; not all are using minimal representations of possibility. Seven-year-olds were very likely to choose the target on the first trial (estimated probability correct .90). Performance declined significantly across trials (odds ratio = 0.94, 95% CI [0.89, 0.99], p = 0.02), possibly due to boredom. It is

clear that many 7-year-olds are not using minimal representations of possibility.

If most participants have modal concepts before they start using modal language, these data are surprising. We expect that 4-year-olds, who use modal language, will either solve the problem from the outset, or else learn how to solve the problem. But there are no clear signs of learning in these data. The data are consistent with the hypothesis that most 4-year-olds use minimal representations of possibility.

Individual-level analysis The previous analysis found no evidence of modal concepts among 4-year-olds; there was evidence of modal concepts in older children. To find additional evidence of modal concepts, we looked at individual performance. Since each participant saw 16 trials, we can check how many participants, in each age group, chose the target significantly more than half the time (13 or more target choices of 16 trials: binomial test n = 16, k = 13, H_0 probability = .5, p = .02). Then we can check whether the number of children who satisfy this criterion was greater than expected by chance (second-order binomial test). Five of 26 4-year-olds chose the target branch 13 or more times; this is more than expected by chance (n = 26, k = 5, H_0 probability = .02, $p < .001$), suggesting that at least some 4-year-olds have modal concepts. Among 5-year-olds 14 of 27 were significant at an individual level, which is more than expected by chance (n = 27, k = 14, H_0 probability = .02, $p < .001$); among 6-year-olds 12 of 24 were significant at an individual level, more than expected by chance (n = 24, k = 12, H0 probability = .02, $p < .001$). Finally, 13 of 26 7-year-olds were significant at an individual level, which is more than expected by chance (n = 26, k = 13, H0 probability = .02, $p < .001$). This analysis sets a lower bound on the proportion of children in each age group who have modal concepts: at least 11% of 4-year-olds and at least 41% of the older children.

These data are not expected if all children have modal concepts. That hypothesis does not explain 4-year-olds' poor performance. Of course, it is always possible that performance factors are masking children's competence. We need to formulate and test hypotheses about what those performance factors might be.

4 Discussion

We analyzed children's comprehension of modal verbs and the nonverbal signs of children's modal concepts. From age 4, children's responses to 'can'-questions were adultlike. However, many 4-year-olds gave the same pattern of responses to 'have to' questions. This undermines our confidence that most children understand 'can', just as learning that most 2-year-olds give all the crackers in response to a request for some of the crackers undermines our confidence in their comprehension of 'all', even though they give all the crackers in response to a request for all the crackers (Barner et al. 2009). However, a small proportion of 4-year-olds differentiate 'have to' from

'can'. That proportion grows with age. This extends existing research, which has shown that some 4-year-olds distinguish what *must be* from what *might be* and from what *could be* (Moore et al. 1990, Moscati et al. 2017). Some 4.5- to 6-year-olds distinguish what has to be the case from what may be the case (Öztürk & Papafragou (2015), experiments 1 and 2), and some and 4- to 5.5-year-olds distinguish what *is* the case from what *may be* the case (Öztürk & Papafragou (2015), experiment 3). A small proportion of 3-year-olds might distinguish what *will* happen from what *might* happen (Armstrong 2020). We have shown that some 4-year-olds distinguish what *can* happen from what *has to* happen.

Many 4-year-olds, but fewer older children, performed poorly on the nonverbal measure of modal concepts; they either lack modal concepts or fail to deploy them.

These data address our hypotheses. Modal concepts either (1) are innate or constructed prior to the onset of modal language production, or (2) are constructed after children start using modal language. Our data provide qualified support for hypothesis 2. Children produce modal language by age 3 (Cournane 2021, Van Dooren et al. 2017). On the first hypothesis 3-year-olds should have modal concepts; hence they should solve our nonverbal task. This was not observed. While the difficulties in figuring out which words go with which concepts can explain their struggles with production and comprehension tasks, they cannot explain their struggles with a task that does not require speech. However, to fully address this question, we need individual-level analyses that check whether the children who comprehend modal language perform better on the nonverbal task. Too few 4-year-olds in the current dataset differentiate 'can' from 'have to' for an informative individual-level analysis.

If many 4-year-olds do not understand 'can', how do they generate the correct response pattern to all 6 'can'-questions? One hypothesis is that when asked whether something can happen, they look for something that prevents that event from happening, and respond accordingly. Asked whether a marble dropped into a branching slide can come out a branched exit, they check whether there is a path from entrance to exit, and if so, they affirm. If not, they deny. More generally, asked whether p is possible, they check whether not-p, and respond accordingly. Being able to check whether not-p does not require modal concepts; it does not require understanding that 'maybe p' is compatible with 'maybe not-p'.

Perhaps most 4-year-olds used the same strategy to answer 'have to' questions, yielding the observed pattern of errors.

There are alternative interpretations for our finding that most 4-year-olds treat 'have to' like 'can'. For example, young children may not have attended to all the relevant facts when generating their answers to modal questions. Our first follow-up experiment will prompt children to attend to all relevant features of the apparatus before asking 'have to' questions. We will ask, "I'm going to drop a marble in here right now. Please show me all the places for it to come out" and "Is this the only place

for it to come out?". We will also adjust the phrasing of the "have to" question to rule out some alternative interpretations. We will now ask, "I'm going to drop this marble right here right now. Does it have to come out right here?" Initial results suggest that 4-year-olds will show much stronger signs of understanding 'have to'. If this result holds up, we will work downward to 3-year-olds to test their comprehension, and upward to 5-year-olds to see if we find near-ceiling performance.

When we are more confident that we are correctly measuring children's comprehension of modal verbs, we will be in a better position to test the relationship between answering modal questions correctly and nonverbal signs of modal concepts. Here we make an intriguing prediction: many 3-year-olds may answer 'can' questions appropriately. But they will not differentiate 'can' from 'have to', and answering 'can' questions correctly will be unrelated to performance on the nonverbal measure of modal concepts. This is because we suspect that 3-year-olds do not have modal concepts, and their success at answering 'can' questions is spurious, generated by repeatedly checking whether not-p and responding accordingly. But children of any age who answer both 'can' and 'have to' questions correctly may have developed modal concepts, and so they might succeed on the nonverbal task.

5 Conclusion

We tested children's comprehension of 'can' and 'have to', and applied a nonverbal measure of modal concepts. At age 4, where differentiation of 'can' and 'have to' was weakest, performance on the nonverbal measure of modal concepts was as predicted if most children use minimal representations of possibility. Differentiation increases slowly with age, as does performance on the nonverbal measure. This converges with many other findings (for a review see (Leahy & Carey 2020)).

However, it is not unlikely that task demands are obscuring children's performance. Next steps will attempt to reduce these task demands to clarify the picture. Once we are satisfied that we are accurately measuring children's comprehension of modal vocabulary and their modal concepts, we will be in a better position to evaluate which comes first.

The most important future step will be to implement a version of this task with 3-year-olds, children who produce a wide variety of modal language but, across many studies, have shown few signs of modal concepts.

Appendix 1: Error analysis

Only 2 response patterns to our 12 modal questions appeared with any frequency: the *Both correct* combination (appropriate responses to all questions), and the *Both as can* combination (responses treat all questions like like 'can').

Figure 5 shows the count of each response pattern in each age group. For each verb we asked 6 questions (top row, "Questions"). We associate 'can' with yellow and 'have to' with blue. We asked about 3 situation types (second row, "Situations"): necessity situations, where the event described by the question is the only possibility; impossibility situations, where the event described by the question is not among the possibilities, and possibility situations, where the event described by the question is one of multiple possibilities. The third row, "Pragmatically appropriate responses", lists the responses that we took to be appropriate for each combination of question and situation. Note that only possibility situations discriminate between the two verbs, as indicated by the red rectangle. In the bottom panel, yellow codes the number of children in each age group who gave "Can-appropriate" responses to all 12 questions. Green codes the number of children who gave "Can-appropriate" responses to all 6 'can'-questions (yellow) and "Have to-appropriate" answers to all 6 'have to'-questions (blue). Note that green = blue + yellow.[2]

If there was a way to interpret a response as appropriate, we coded it as appropriate. So, for example, if a child said "yes" in necessity situations and "no" in impossibility situations, and used possibility modals for all questions about possibility situations, they would be coded as "both verbs interpretable as correct". However, it would not be clear whether this child has differentiated the verbs. This strategy risks showing more differentiation than is actually present. We chose it because because differentiation was rare even with this potential inflation. The point here is that differentiation is rare.

Figure 5 shows us, first, that more than half of participants used one of these two strategies across all 12 questions. This is especially surprising for 4-year-olds, where even a brief lapse of attention over 12 questions could yield an uninterpretable error.

Second, many 4- to 6-year-olds treat "have to" the same way as they treat "can". Given our coding scheme, this means that 10 4-year-olds, 9 5-year-olds, 9 6-year-olds, and even 5 7-year-olds said 'yes' to both discriminating have to-questions. That is, when asked, "If I drop a marble in here [branching entrance], does it have to come out here [one of the branched exits]?", these children said 'yes' both times, while answering all 10 other questions appropriately. This is a striking pattern that demands explanation. One tantalizing possiblility is that many children even at age 6 have not yet mapped their modal vocabulary onto the appropriate modal concepts. Some of these children may have not yet developed modal concepts.

Third, some children give appropriate responses to all 12 questions. So some 4-year-olds seem to have differentiated 'can' from 'have to'. An important question for future research is whether children who have differentiated these words perform better on our nonverbal measure of modal concepts. But before we can ask that

2 This figure was modeled after one designed by Ailís Cournane, who independently developed a similar experiment.

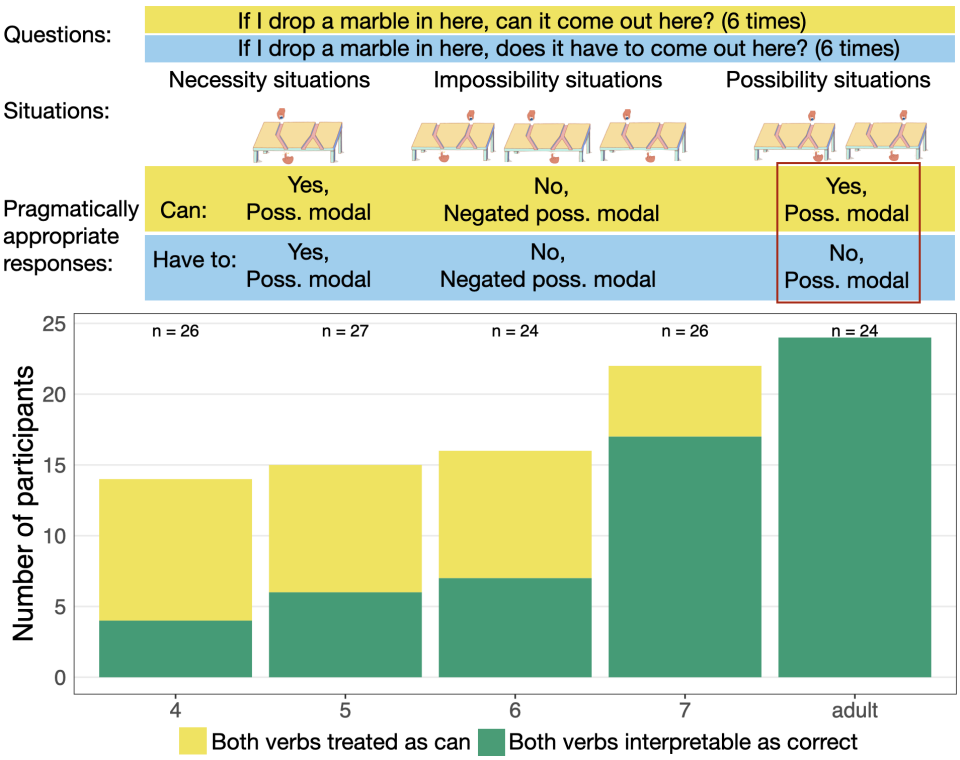

Figure 5: Top row: two question types. Second row: three kinds of situation. Third row: Responses that were counted as pragmatically appropriate for each combination of question and situation. Bottom: count of each response pattern in each age group.

question, we must be confident that performance factors are not limiting children's performance on "have to" questions.

With those qualifications in place, the observed conflation of these verbs raises two hypotheses to salience. On the first, many young children think 'can' and 'have to' both mean 'can'. On the second hypothesis they use a simple strategy for evaluating both "possibly p" and "necessarily p": check whether not-p; if not-p, then deny; if not-not-p, then affirm.

References

Armstrong, Meghan. 2020. Children's epistemic inferences through modal verbs and prosody. *Journal of Child Language* 47(6). 1132–1169. http://dx.doi.org/10.1017/S0305000919000916.

Baillargeon, Renee, Elizabeth S Spelke & Stanley Wasserman. 1985. Object

permanence in five-month-old infants. *Cognition* 20(3). 191–208. http://dx.doi.org/10.1016/0010-0277(85)90008-3.

Barner, David, Katherine Chow & Shu-Ju Yang. 2009. Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology* 58(2). 195–219. http://dx.doi.org/10.1016/j.cogpsych.2008.07.001.

Cournane, Ailís. 2020. Learning modals: A grammatical perspective. *Language and Linguistics Compass* 14(10). 1–22. http://dx.doi.org/10.1111/lnc3.12399.

Cournane, Ailís. 2021. Revisiting the epistemic gap: It's not the thought that counts. *Language Acquisition* 1–26. http://dx.doi.org/10.1080/10489223.2020.1860054.

Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane & Valentine Hacquard. 2019. Acquiring the force of modals: Sig you guess what sig means?

Grigoroglou, Myrto, Sharon Chan & Patricia A Ganea. 2019. Toddlers' understanding and use of verbal negation in inferential reasoning search tasks. *Journal of Experimental Child Psychology* 183. 222–241. http://dx.doi.org/10.1016/j.jecp.2019.02.004.

Leahy, Brian. Submitted. Don't you see the possibilities?

Leahy, Brian & Susan Carey. 2020. The acquisition of modal concepts. *Trends in Cognitive Sciences* 24(1). 65–78. http://dx.doi.org/10.1016/j.tics.2019.11.004.

Leahy, Brian & Eimantas Žalnieriūnas. 2021. Data and analysis script for Might and might not: Children's conceptual development and the acquisition of modal verbs. https://doi.org/10.7910/DVN/GABLSZ.

Mody, Shilpa & Susan Carey. 2016. The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition* 154. 40–48. http://dx.doi.org/10.1016/j.cognition.2016.05.012.

Moore, Chris, Kiran Pure & David Furrow. 1990. Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development* 61(3). 722–730. http://dx.doi.org/10.2307/1130957.

Moscati, Vincenzo, Likan Zhan & Peng Zhou. 2017. Children's on-line processing of epistemic modals. *Journal of Child Language* 44(5). 1025–1040. http://dx.doi.org/10.1017/S0305000916000313.

Noveck, Ira A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2). 165–188. http://dx.doi.org/10.1016/S0010-0277(00)00114-1.

Öztürk, Ozge & Anna Papafragou. 2015. The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development* 11(3). 191–214. http://dx.doi.org/10.1080/15475441.2014.905169.

Redshaw, Jonathan & Thomas Suddendorf. 2016. Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology* 26(13).

1758–1762. http://dx.doi.org/10.1016/j.cub.2016.04.062.
Shtulman, Andrew & Susan Carey. 2007. Improbable or impossible? how children reason about the possibility of extraordinary events. *Child Development* 78(3). 1015–1032. http://dx.doi.org/10.1111/j.1467-8624.2007.01047.x.
Van Dooren, Annemarie, Anouk Dieuleveut, Ailís Cournane & Valentine Hacquard. 2017. Learning what must and can must and can mean. In *21st Amsterdam Colloquium*, 225–234. Institute for Logic, Language and Computation.
Van Dooren, Annemarie, Anouk Dieuleveut, Ailís Cournane & Valentine Hacquard. Forthcoming. Figuring out root and epistemic uses for modals: The role of the input.

Brian Leahy
Department of Psychology
Harvard University
33 Kirkland Street Cambridge MA
...
brianleahy@g.harvard.edu

Eimantas Žalnieriūnas
Department of Psychology
University of Bath
10 West, Claverton Down, Bath
and
Department of Psychology
Harvard University
33 Kirkland Street Cambridge MA
...
ezalnieriunas@gmail.com