Three-year-old children's reasoning about possibilities

Stephanie Alderete & Fei Xu

University of California, Berkeley

Word count: 8,668 (abstract + text)

Address correspondence to: Stephanie Alderete, 2121 Berkeley West Way, Department of Psychology, UC Berkeley, Berkeley, CA 94720. E-mail: salderete@berkeley.edu.

Abstract

Recent studies in cognitive development suggest that preschoolers may not be able to represent multiple possibilities, therefore may lack modal concepts such as *possible, impossible,* and *necessary* (Leahy & Carey, 2020). We present two experiments adapted from previous probability studies but have a similar logical structure as those used in the previous modal reasoning tasks (Leahy, under review; Leahy et al., 2022; Mody & Carey, 2016). Three-year-old children have to choose between a gumball machine that *must* produce the desired gumball color and a gumball machine that merely *might* produce the desired gumball. Results provide preliminary evidence that three-year-old children can represent multiple incompatible possibilities, therefore they have modal concepts. Implications for the study of modal cognition, and how possibility and probability may be related are discussed.

Key Words: modal concepts, possibility, probability, cognitive development

1. Introduction

An important aspect of abstract human thought is our ability to represent possible alternatives (Shtulman, & Phillips, 2018; Vaidya, 2009). One-way humans can make decisions about the future and contemplate the past is by reasoning about possible alternatives. For example, imagine you are walking down the street when suddenly you come across a fork in the road. Now you have to consider two possibilities: Do you take the left path or the right path? The ability to represent alternative possibilities is contingent on the development of modal concepts, or conceptual representations of what is *possible*, *impossible,* and *necessary*. These conceptual distinctions are also reflected in our use of modal language (e.g., must, may, might) (Leahy & Zalnieriunas, 2021). Recently, cognitive scientists have debated whether young children and non-human primates have the capacity to think about possible alternatives and if modal concepts are a part of the cognitive repertoire that learners have early in ontogeny (Engelmann et al. 2021; Mody & Carey, 2016; Redshaw & Suddendorf, 2020; Shtulman & Phillips, 2018).

Recent research has suggested that children do not acquire modal concepts (i.e., reason about multiple possibilities) until at least four or five years of age (Leahy & Carey, 2020). Evidence for this claim comes from preschoolers' failure on a three-containers task (Mody & Carey, 2016). Children were presented with three opaque containers divided into a singleton side and a doubleton side. Each side was covered with an occluder. A reward (e.g., sticker) was placed into the singleton container behind the occluder, and a reward was placed into one of the two containers behind the other occluder (Figure 1). The occluders were lifted and children were asked to pick a container with the goal of finding a reward. Success only required that children recognize the singleton container *must* have a reward in it while each of the two containers on the doubleton side only *might* have a reward in it. Children of all age groups chose the singleton

container at above chance level (33%): two-and-a-half-year-olds: 47%, three-year-olds: 60%, four-year-olds: 71%, and five-year-olds: 72%, but even five-year-olds' performance was far from a robust success. This was surprising since surely adults would perform at ceiling, choosing the singleton container at close to 100% (barring some performance error). In another task testing children's understanding of possibility, Leahy and Zalnieriunas (2021) used two slides, one with a single exit and the second one with two possible exits (Figure 6), and children had to decide where to place a wagon in order to catch a marble. This task is conceptually similar to the three-containers task, and preschoolers performed at a level comparable to that of the three-containers task. It is not until about age 7 that children performed at ceiling on this slides task (Leahy & Zalnieriunas, 2021).

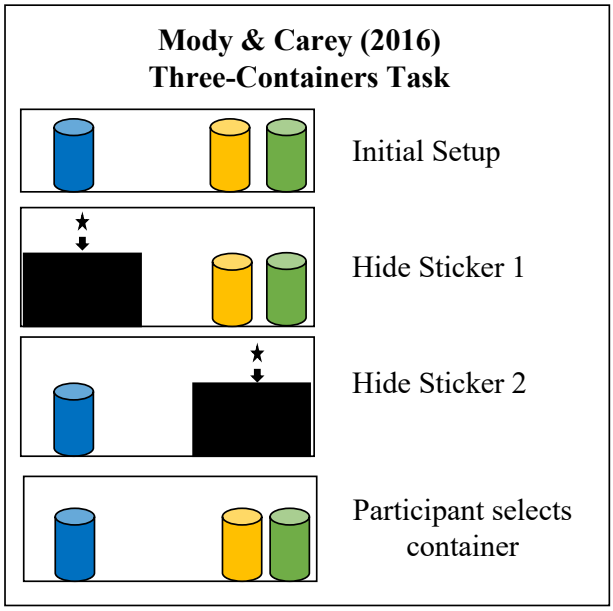

Figure 1. Mody & Carey (2016) three-containers task

Some have suggested that children's poor performance on the three-containers task was due to their lack of the modal concept of possibility (Leahy & Carey, 2020). A child who has not developed modal concepts would be unable to represent the alternative locations of the reward

on the doubleton side (e.g., the reward *might* be in the left container, or it *might* be in the right container). Without the modal concept of possibility, the child does not mark a representation of the reward on the doubleton side as merely possible (e.g., the reward *might* be in the left container). Since the representation is not marked as a possibility, the child conflates a mere possibility with reality (e.g., the reward *is* in the left container). When asked to pick a container, the child chooses at random between the singleton container and the container they believe the reward is in on the doubleton side (e.g., left container), resulting in the roughly 50% selection of the singleton. This behavior has been dubbed a "minimal representation of possibility" (Leahy & Carey, 2020) (Figure 2).

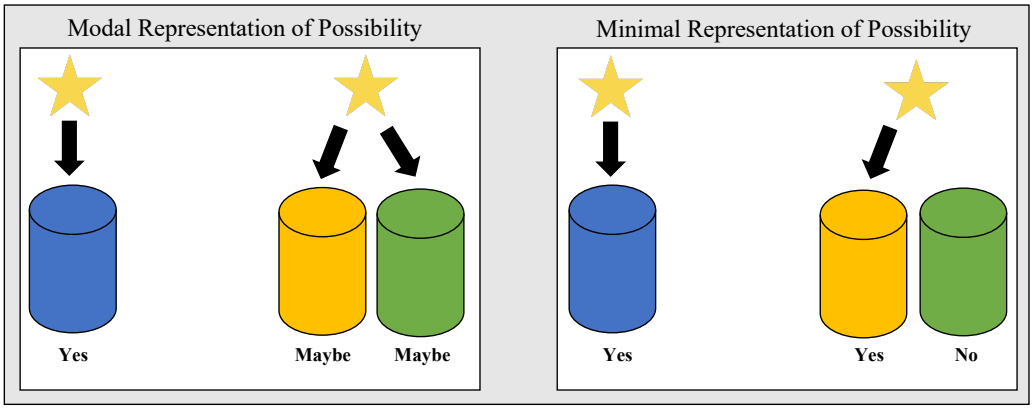

Figure 2. Modal representation of possibility versus Leahy & Carey (2020) Minimal representation of possibility in Mody & Carey (2016) three-containers task. Labels at the bottom indicate whether the child believes there is a reward in that container: yes, no, maybe

Several recent studies have provided converging evidence for the minimal representation hypothesis (Leahy, under review; Leahy et al., 2022). Leahy and colleagues, (2022) created an adapted version of the three-container task that presented children with the same setup as in the original task: three containers divided into a singleton and a doubleton side, each side was occluded, and a reward was hidden in one of the containers on each side. Rather than picking a container, three-year-old children were asked first to pick a container that they believe is empty

to throw away, and then pick one of the two remaining containers to look inside with the goal of getting a reward. Children chose to throw away a container from the doubleton side 81% of the time. When they were left with a choice between the singleton container that *must* contain the reward and the remaining container on the doubleton side that *might* contain a reward, children chose to look inside the singleton container 51%. These findings support the minimal representation hypothesis that children were making an assumption about where the reward was on the doubleton side (see also Grigoroglou et al., 2019; Leahy, under review).

These previous studies provide evidence that preschoolers are acting in accordance with the minimal representation hypothesis. They attend to only one of two possible outcomes and treat it as reality. Yet, the reason why children make this assumption is unclear. It may be that preschool-age children use a minimal representation of possibility because they have not developed modal concepts, or the capacity to represent two alternative events. However, there may be other reasons for children's failure on these tasks. One common critique of the three-container task is the presence of occlusion. Covering the containers with occluders before the experimenter places a reward in the container may add to a child's working memory load. However, Leahy (under review) modified the three-containers task into a slides task (Figure 6) that did not involve any occlusion and found results consistent with those of Mody and Carey (2016). But, in both tasks, children were given feedback that may be confusing over multiple trials (e.g., showing them what was inside the container they chose, which contained a reward half of the time). Thus, potential task demands may have prompted children to make an assumption about the location of the reward on the doubleton side.

The literature on early probabilistic reasoning, however, may provide suggestive evidence for the early emergence of modal concepts. Probability and possibility appear to be

closely related. What is probable ($p > 0$) must also be possible. Infants as young as six to 12 months can use probabilities to infer outcomes (Denison & Xu, 2010, 2014; Xu & Garcia, 2008; see Denison & Xu, 2019 for a review). Indeed, success on the three-containers task can be achieved by computing the relative probability of finding the reward on each side (i.e.., there is a 100% chance of finding the reward in the singleton container and a 50% chance of finding a reward in either of the two containers on the doubleton side). The close relationship between probability and possibility prompted us to see if three-year-old children, who perform poorly at the three-containers task, might show competence in modal reasoning in a task similar to those used in the probability studies.

The present study was inspired by the design and simplicity of the infant probability tasks. We adapt an infant probability task (Denison & Xu, 2010, 2014) so it has a similar logical structure to the three-containers task (Mody & Carey, 2016) but does not involve occlusion or feedback. In our task, three-year-old children are asked to help the experimenter choose a gumball machine that will give the experimenter the color gumball they desire. Children were asked to choose between a gumball machine that *must* produce the desired gumball and a gumball machine that *might possibly* produce the desired gumball. For example, the experimenter says that they want a purple gumball, and the child is asked to choose between a blue gumball machine containing a single purple gumball, and a green gumball machine containing a purple and a red gumball (Figure 4, 1 vs. 2).

The current study aims to test two contrasting hypotheses on early modal reasoning. If children have modal concepts, they should always choose the machine which contains only the desired gumball. A **modal representer** simulates the handle press for the one-gumball machine and understands that the desired gumball *must* come out. She also simulates the handle press for

the two-gumball machine and understands that the gumball machine that contains one desired gumball out of two could produce a desired gumball *or* the other gumball. The child understands that there are two possible outcomes in the machine with two different colored gumballs (Figure 3).

In contrast, if children lack modal concepts and are minimal representers, they will choose the machine which contains only the desired gumball 75% of the time. A **minimal representer** simulates the handle press for the one-gumball machine and knows that the desired gumball will come out. She also simulates the handle press for the two-gumball machine: half of the time she will get the desired gumball (Figure 3, A), therefore, choosing 50-50 between the two gumball machines, and half of the time she will get the other gumball, therefore, choosing the one-gumball machine (Figure 3, B).

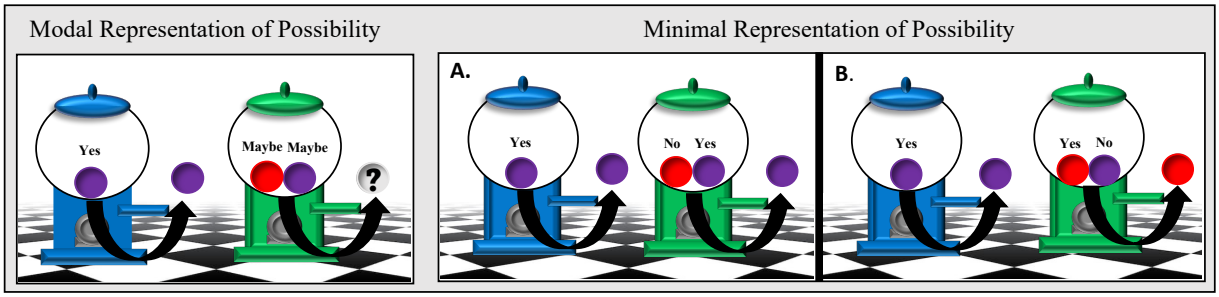

Figure 3. Modal versus Minimal Representation of possibility in the gumball study. Labels over gumballs indicate whether the child believes that the gumball will come out: Yes, no, or maybe. A modal representer always simulates that a purple gumball will come out of the blue machine when the handle is pressed, while a purple gumball or red gumball might come out of the green machine when the handle is pressed. A minimal representer always simulates that a desired purple gumball must come out of the blue machine when the handle is pressed. For the green machine, half the time the child believes the desired purple gumball will come out when the handle is pressed, the other half of the time they simulate that the undesired gumball will come out when the handle is pressed.

1. Experiment 1

The study was approved by the University of California, Berkeley institutional review board and follows all ethical & legal guidelines.

2.1 Participants

Sixteen three-year-old children (10 Female, mean age = 3.34, range = 3.05 - 3.92, sd = 0.25) were included in our final sample. An additional 4 children were excluded for failing the warm-up trials: 2 children were excluded because they failed the probability trials (see below), and 2 children did not complete the task due to inattention. The sample size was chosen before testing started, and we stopped collecting data once we reached this target sample size. Given the small sample size in Experiment 1, in Experiment 2, we replicate the results of Experiment 1 with a larger sample size of 36. We chose to test three-year-old children based on Leahy et al. (2022) and Leahy (under review), which tested three-year-olds on an adapted three-containers task and the slides task. All children were recruited from the Bay Area, California via an online laboratory database. Demographic information was not formally collected; however, most children were from Alameda County, which consists of 47.8 % White, 33.8% Asian, 22.4% Hispanic, and 10.7% Black (Alameda County Census Bureau). Additionally, the 2021 median household income of the county is $112,000 and 49.6% of the population holds a bachelor's degree or higher. Due to the Covid pandemic, all participants were tested online via zoom. Families consented before the study. Children were given a certificate for their participation.

2.2 *Materials*

The stimuli were animated images of gumball machines of different colors, including blue, green, and red, and the machines contained different numbers of gumballs. All animations were created with Microsoft PowerPoint.

2.3 Design

Study 1 had a repeated measure design with one main dependent variable: which gumball machine handle (blue machine or green machine) the participant chose to press. The side the target gumball machine was on in the first trial (left or right) was counterbalanced across participants. Within participants, the side the target gumball machine was on, and what color gumballs were in the machines varied randomly across trials.

2.4 Procedure

Participants were asked to sit in front of a computer screen at a comfortable distance next to their parents or on their lap. The experimenter told the participant that they were going to play a fun game. The experimenter also asked the parent not to talk or point during the study.

2.4.1 Training trials

Gumball Machine Demonstration

Each session began with two demonstrations to familiarize participants with the game. The first trial showed participants how the gumball machine worked. Participants were shown a single red gumball machine filled with eight pink gumballs (Figure 4). Then the experimenter pressed the handle on the gumball machine, triggering a pink gumball to exit the machine. For the second trial, participants were shown a red gumball machine filled with eight blue gumballs. This time participants were asked what color gumball will come out of the machine. After the participant answered, the experimenter pressed the handle, triggering a blue gumball to exit the machine.

Probability Trials

After the familiarization trials, the participants did two *probability trials*. The purpose of the probability trials was to see if participants could correctly infer which gumball was more likely to come out of the machine. These trials also primed participants to think about different possible outcomes. In the first probability trial, participants were shown a red gumball machine filled with six black gumballs and two pink gumballs (Figure 4). Participants were asked, "What color gumball do you think will come out?". After they answered, the experimenter pressed the handle, triggering the more probable gumball (black) to exit the machine. The second trial was the same except the colors were reversed. For each probability trial, if the participant did not answer correctly, they were given a second chance with a new gumball machine that had orange and blue gumballs in it. Participants were excluded if they failed both trials. Importantly, with the demonstration trials and the probability trials, participants also learned that each time the handle is pressed, *only one gumball* will come out of the machine.

Introduction to two machines

After the probability trials, participants were introduced to two gumball machines: a blue one with two pink gumballs inside, and a green one with two black gumballs inside (Figure 4). The purpose of these trials was to train participants to choose a gumball machine that will give them the desired gumball color. The experimenter told each participant, "I really want a pink gumball. Which handle should I press to get a pink gumball, the blue handle or the green handle?". The participant gave a verbal answer, and then the experimenter asked the participant to point to the machine. Parents were asked to confirm that the color the participant stated corresponded to the machine they pointed to.

2.4.2 Test Trials

Participants completed a total of six test trials. The trials were divided into two trial types. The first four trials were 1 vs. 2 gumball trials (Figure 4). Participants were presented with two gumball machines: a blue gumball machine and a green gumball machine. For each trial, one of the machines contained a single gumball with the desired gumball color (e.g., purple), and the other machine contained two gumballs: a gumball of the desired color (e.g., purple) and a gumball of a different color (e.g., red). In each trial, the experimenter told each participant, for example, "I really want a *purple* gumball. Which handle should I press to get a *purple* gumball, the blue handle or the green handle?". Participants responded verbally and then were asked to point to the machine they chose. The parent was asked to confirm the participant's response. Participants did not see a gumball exiting the gumball machine, and there was no feedback during test trials. After the participant chose a machine for each test trial, the experimenter said, "Thank you! Let's do another one."

The last two trials were 2 vs 2 gumball trials (Figure 4). These trials were the same as the 1 vs. 2 gumball trials, except each gumball machine had two gumballs. Participants were shown, for example, a blue gumball machine with an orange gumball and a black gumball, and a green gumball machine with two black gumballs. The purpose of these trials was to control for the possibility that participants, for some reason, may have simply preferred a gumball machine with a single gumball.

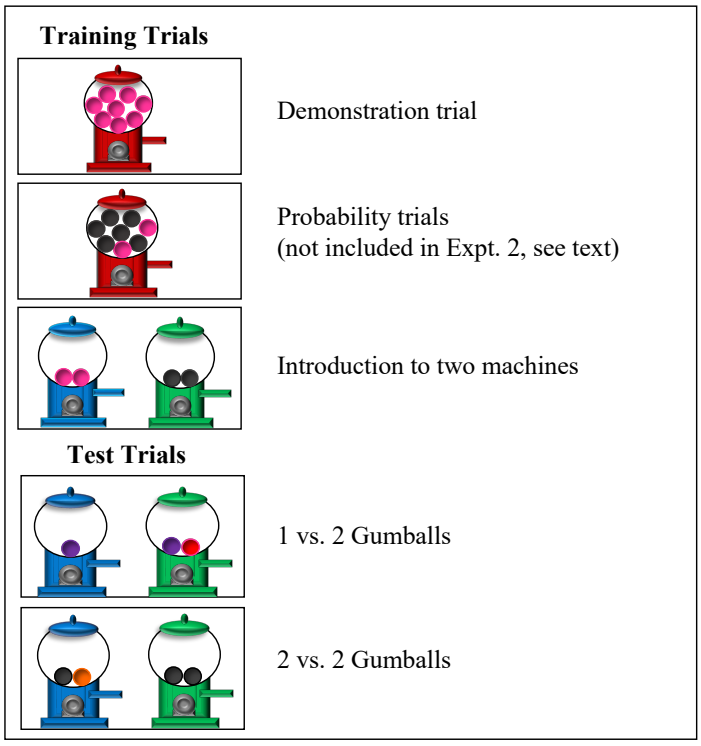

Figure 4. Gumball task methods. Note: The color of the gumballs inside each gumball machine varied in each of the test trials.

2.5 Results

Our analysis had two parts. First, we conducted a preliminary analysis that tested for gender, order, and trial effects. Second, we compared the average participant's performance in the test trials to 50% (chance performance) and to 75% (the performance predicted under the minimal representation of the possibility hypothesis). In both Experiment 1 and Experiment 2, the preliminary and main analyses were conducted using generalized linear mixed effect models (GLMM) from the lme4 package in R (Douglas et al., 2015). Datasets and R codes for both Experiments can be found on the Open Science Framework (OSF): (redacted).

2.5.1 Preliminary Analysis

A GLMM was fit predicting the probability of choosing the target gumball machine from age (younger 3.0-3.50 vs. older 3.51-3.99), gender (male vs. female), order (which side the target gumball was on in the first trial (left or right)), and trial type (1 vs. 2, vs., 2 vs. 2) with a random intercept for participant id. Our model had dummy-coded fixed effects for age, gender, order, and trial type. There was no effect of gender, ($\hat{\beta}$ = -1.28, *SE* = 1.38, *z* = -0.93, *p* = 0.35), order ($\hat{\beta}$ = -.83, *SE* = 1.41, z = -0.59, *p* = 0.56), or age ($\hat{\beta}$ = -.08, *SE* = 1.18, *z* = 0.00, *p* = 1.00). There was also no significant effect of trial type on performance ($\hat{\beta}$ = -.23, *SE* = .84, *z* = -0.27, *p* = 0.79). The mean performance for the 1 vs. 2 trials and 2 vs. 2 trials were 92.18% and 90.6%, respectively. Thus, having a single desired gumball in a machine versus two desired gumballs in a machine did not affect performance.

2.5.2 Test Trials

Across all six trials, participants on average chose the gumball machine which contained only the desired gumball 91.66% of the time (Figure 5). A GLMM was fit predicting the probability of choosing the correct gumball machine from the intercept. We then used the emmean package (Lenth, 2022) in R to get the estimated marginal mean performance on the gumball task from the regression model. We compared that mean to chance performance (50%) and the performance predicted under the minimal representation hypothesis (75%). The average participant's performance (Probability correct = 96%, 95% CI = [.80, .99]) was significantly above 50% (z = 3.53, *p* < .001), and significantly above 75% (z = 2.30, *p* = .021). We also provide converging evidence from a one-sample Wilcox test which compared participants'

proportion correct on all six test trials to 50% ($t(15) = 120$, $p < .001$, $d = .91$) and, in a separate test, to 75% ($t(15) = 123$, $p = .003$, $d = .66$)[1].

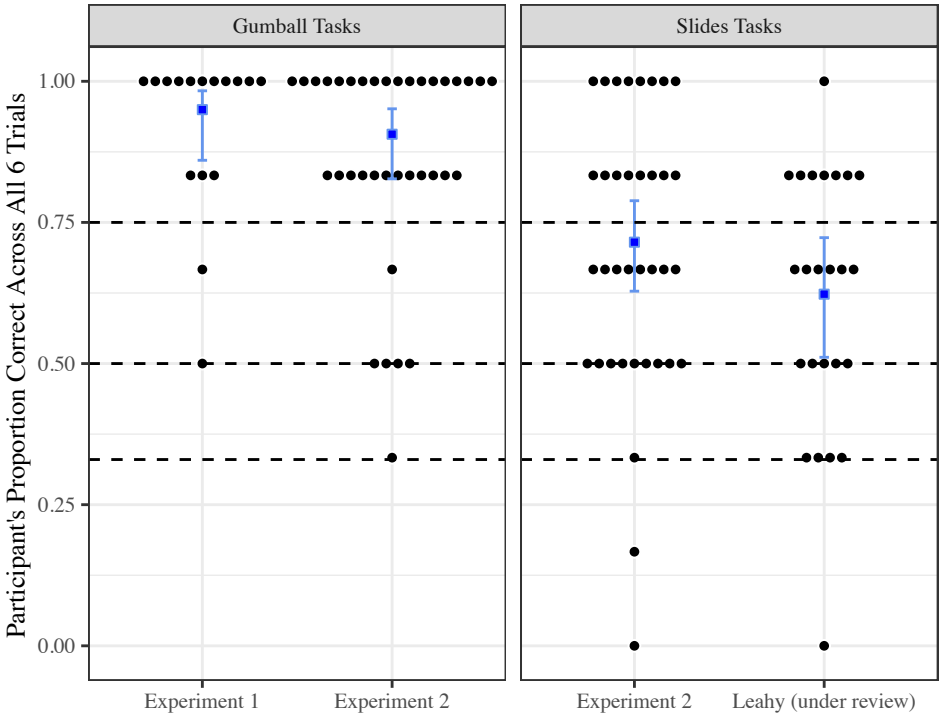

Figure 5. Left side depicts results from the gumball task in Experiments 1 and 2. Right side depicts results from the slides task in Experiment 2 and Leahy (under review) The squares depict the mean performance with 95% confidence intervals. Each dot represents a participant's performance across all six test trials. The Y-axis represents the proportion of correct response on the six test trials. The dash line at 75% represents the performance predicted by the minimal representation of possibility hypothesis in the gumball tasks. The dash line at 50% represents chance performance for the gumball task. The dash line at 33% represents chance performance for the slides task.

2.6 Discussion

These data provide preliminary evidence against the hypothesis that children cannot represent alternative possibilities or have a minimal representation of possibility. In the gumball

[1] In an exploratory analysis, we wanted to examine whether the probability training trials affected performance on the test trials. However, we discovered there were only two participants that were excluded due to failing the probability trials. Thus, we included the 2 participants who were excluded because they failed the probability trials. With this new sample of 18, participants chose the target gumball machine 89%. A one-sample Wilcox test was conducted to compare participant's performance across all trials to 50% (chance) and 75% (the performance predicted by the minimal representation of the possibility hypothesis). Results were significantly above 50% ($t(17) = 136$, $p < .001$, $d = .89$) and significantly above 75%, ($t(17) = 146$, $p = .006$, $d = .79$)

task, participants overwhelmingly chose the target gumball machine above chance (50%) and above 75% (the performance predicted under the minimal representation of possibility hypothesis). Thus, children are not deploying a minimal representation of possibility on this task. Rather, three-year-old children may be succeeding by using modal concepts to distinguish between the gumball machine that *must* produce the desired gumball and the gumball machine that *might possibly* produce the desired gumball. The results support the hypothesis that three-year-olds are capable of representing two alternative possibilities.

The results from Experiment 1 contradict the results of the previous modal reasoning tasks (Leahy, under review; Leahy et al., 2022; Mody & Carey, 2016). What are the differences between the tasks that could explain the performance differences? One possibility is that it may be easier to quantify over objects (as in the gumball task) rather than over locations or trajectories (as in the three-container tasks and the slides task). In the gumball task, the gumballs are laid out in front of the children and are visible throughout the trials. The child may be able to succeed in the task by quantifying over the gumballs in the gumball machines (i.e., one possible outcome vs. two possible outcomes). In the three-containers tasks or the slides task, success requires that the child tracks the reward as it goes behind the occluder or the marble as it goes down the slides. Thus, perhaps it is easier for young children to quantify over objects like gumballs compared with locations or trajectories.

This hypothesis motivated the design of Experiment 2. Children were asked to complete the gumball task first followed by the slides task. Experiment 2 had two aims: first, to replicate the main results of Experiment 1 with a larger sample size, and second, to test whether playing the gumball task prior to the slides task may improve three-year-old children's performance on the slides task. We chose the slides task because it did not have any of the working memory

demands that were present in the three-containers task (i.e., hiding and occlusion), but yielded the same roughly 60% performance as in the original three-containers task (Leahy, under review).

3. Experiment 2

In Experiment 2 three-year-old children were tested in the gumball task first, then the slides task. We hypothesized that the gumball task may prime children to think about multiple possibilities, which may improve their performance on the slides task. However, if quantifying over objects (as in the gumball task) is easier than quantifying over locations (as in the slides task), we may not see any improvement on the slides task after completing the gumball task.

Experiment 2 was preregistered on aspredicted.org (redacted), was approved by UC Berkeley institutional review board, and follows all ethical & legal guidelines.

3.1 Participants

Participants were 36 three-year-old children (23 female, mean age = 3.56, range = 3.04-3.99, sd = .32). An additional five participants were excluded due to failing the training trials, and one child was excluded due to fussiness. The sample size was chosen before testing started, and pre-registered. The effect sizes in Experiment 1 ranged from medium to large. However, the sample size (N = 16) was fairly small. One of the goals of Experiment 2 is to see if success on the gumball task scaffolds performance on the slides task. If children improved on the slides task, we would like to be able to compare their performance to the original slides task (Leahy, under review). We conducted a power analysis with this goal in mind, assuming a medium effect size and a power of .8. By testing 36 children in Experiment 2 we are also able to provide a replication of Experiment 1 by doubling the sample size (a rule of thumb often used in

replications). All participants were recruited from the Berkeley area via the same online laboratory database used in Experiment 1, with similar demographics. All participants were tested online via zoom after a parent gave informed consent. Children were given a certificate for their participation.

3.2 Materials

The stimuli for the gumball task were the same as in Experiment 1. The slides task stimuli were taken from Leahy (under review), which used an animated computer game with slides, marbles, wagons, and a monkey.[2]

3.3 Design

The design of the gumball task was the same as in Experiment 1. The slides task had a repeated measure design with one main dependent variable: which slide exit the child chose to place the wagon under. The side the single slide was on in the first trial (left or right) and which exit on the Y-shape slide the marble came out of were counterbalanced across participants. Within participants, the side the single slide was on, and which exit on the Y-shape slide the marble came out of varied on the test trials. Each participant was randomly assigned one of two orders for which exit the marble came out of on the Y-shape slide in each trial: order one (left side, left side, right side, left side, right side, right side), or order two (right side, right side, left side, right side, left side, left side).

3.4 Procedure

All participants completed the gumball task first, then the slides task.

[2] We would like to thank Brian Leahy for allowing us to use his stimuli for the slides task.

3.4.1 Gumball Task

The methods for the gumball task were the same as in Experiment 1, except that we removed the probabilities trials. Our exploratory analysis (see footnote 1) did not find an effect of the probability trials, and by removing these, we can also rule out the alternative interpretation that these trials may have primed children to use probability in the gumball task.

3.4.2 Slides Task.

After participants completed the gumball task, they did the slides task. The slides task procedure was the same as in Leahy (under review). To ensure we performed the slides task exactly as it was performed in the original study, we worked closely with Dr. Brian Leahy who provided us with the study script, coding sheet template, and computer-animated slides task that he used in his original study.

In the slides task, participants were asked to sit next to a parent, or on their lap. Participants were asked to give their responses by pointing to a slide exit. The parent or guardian told the experimenter where the participant had pointed.

3.4.3 Training Trials

Single Slide

The single-slide demonstration was used to familiarize participants with how to catch marbles. Participants were shown a single slide with one branching exit, with a marble held at the top of the entrance (Figure 6). The demonstration had three parts. In the first part, the experimenter told the participant to watch the marble roll down the slide: “Watch where the marble comes out” (*marble rolls down the slide*). “It came out here” (*animated finger points to*

single slide exit). The second part taught participants how to catch a marble. The experimenter placed an animated red wagon under the slide exit, the marble was dropped and landed in the wagon. The third part taught participants to point to the slide exit, where they want to place the wagon in order to catch the marble: "Okay, so in this game, you have to point to where the marble's going to come out. Then I'll put this wagon (*red wagon appears under the slide*) where you pointed, and we'll see if it goes in the wagon! Wanna have a go?". Participants pointed to the slide exit, and a parent or guardian told the experimenter where the participant had pointed.

In these demonstrations, participants were also taught that when the red wagon catches the marble, a monkey comes out and grabs the marble from the wagon.

Y-shape Slide

The purpose of the Y-shape slide training was to familiarize participants with how the Y-shape slide worked. Participants were shown a Y-shape slide containing a single entrance connected to two branching exits (Figure 6). The experimenter asked the participant to watch the marble roll down the slide and then point to the slide exit where the marble came out. Before the marble was dropped, the experimenter told the participant the marble could come out of the left side exit or the right-side exit: "After I roll it, can you point to the side it comes out from? Here (*animated finger appears to point to left slide exit*) or here (*animated finger appears to point to left slide exit*)?" Then participants were shown two events: One where the marble exited the left side, and another where it exited the right side. No wagons were used in these trials, participants merely watched the marble exit the slide and pointed to where it came out.

3.4.4 Test Trials

Participants completed 6 test trials. In each trial, children were presented with two slides, a single slide with one exit, and a Y-shaped slide that had a single entrance point and two branching exits. A hand holding a marble was placed above each slide entrance point (Figure 6). Participants had to place a red wagon under one of the slide exits with the goal of catching the marble: “Let’s help the monkey get a marble! Remember, we have one wagon. So pick the best slide to put the wagon under so that we’re sure the monkey will get a marble!”

Participants indicated which exit they wanted to place the wagon under by pointing, and the parent told the experimenter which slide exit the participant selected. The experimenter then placed a red wagon under the exit the participant pointed to. After the wagon was placed, both marbles were dropped. If the wagon caught a marble, the monkey came out and grabbed the marble from the wagon. Otherwise, the marble simply dropped off the screen. This procedure was repeated for all 6 trials. On average, the wagon caught the marble 50% of the time if it was placed under the Y-shaped slide.

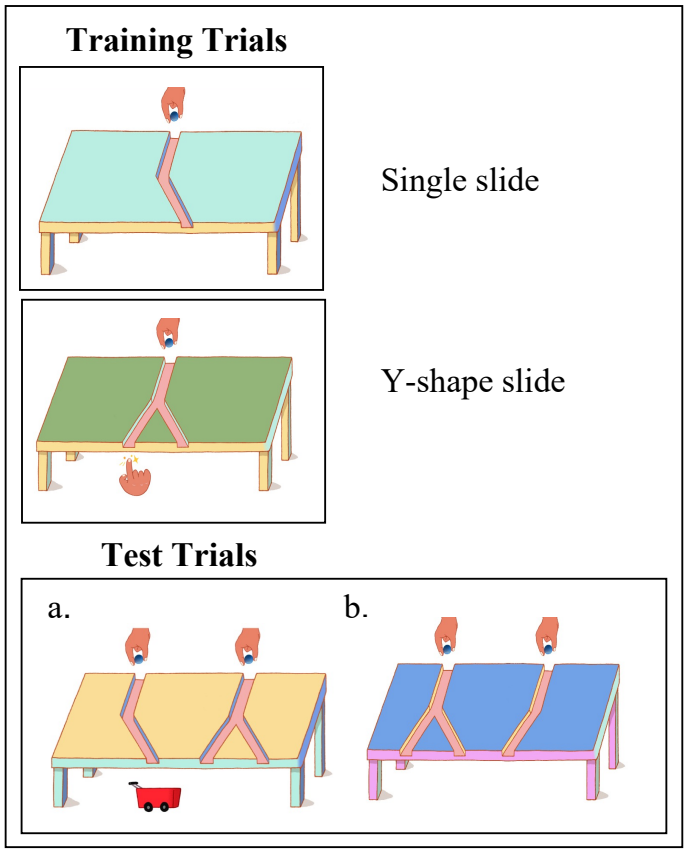

Figure 6. Experiment 2 and Leahy (under review) slides task. Test trials display counterbalanced orders: the single slide on left (Figure 6a), and the single slide on the right (Figure 6b).

3.5 Results

Our analysis has four parts. First, we conducted a preliminary analysis that tested for gender, order, age, and learning effects in the gumball task and, in a separate analysis, in the slides task. Second, we compared performance on the new gumball task (Expt. 2), to performance on the original gumball task (Expt. 1) to see if the study replicated. We then analyzed which gumball machine participants selected in the test trials of the new gumball task and compared it to chance (50%) and the performance predicted by the minimal representation of possibility hypothesis (75%). Third, we compared participant's performance on the Experiment 2 slides task to participant's performance in the original slides task data from Leahy (under

review)[3]. Finally, we conducted a secondary analysis to check whether performance on the gumball task is correlated with performance on the slides task[4].

3.5.1 Preliminary Analysis

Separate preliminary analyses were conducted on the gumball task and the slides task. In each analysis, we tested for gender, age, order, trial, and order effects in the data. Both analyses had dummy-coded fixed effects.

3.5.2 Preliminary Analysis: Gumball task

A GLMM was fit predicting the probability of choosing the target gumball machine from the fixed effects of gender (male vs. female), age (age 3.0-3.50 vs. 3.5-3.99), order (i.e., which side the target gumball was on in the first trial: left or right), and trial type (1 vs. 2, vs. 2 v. 2) with a random intercept for participant id. There was no effect of gender, ($\hat{\beta}$ = -1.28, *SE* = 1.38, *z* = -0.93, *p* = 0.35), order ($\hat{\beta}$ = -*.83, SE = 1.41, z* = -0.59, *p* = 0.56), or trial type ($\hat{\beta}$ = 0.86, SE = 0.52, z = 1.64, p = 0.10). There was an effect of age ($\hat{\beta}$ = 1.29*, SE* = 0.58, z = 2.22, *p* = 0.02), older three- year old's (range 3.5-3.99) selected the target gumball machine 89% while younger three-year old's (range = 3.0-3.5) selected it 84% of the time.

3.5.3 Preliminary Analysis: Slides task

---

[3] Thanks to Dr. Brian Leahy who provided us with the original slides data for our analysis.

[4] We originally pre-registered a Bayesian analysis to compare the gumball task performances from Experiments 1 and 2, and to compare the slides task performance from Experiment 2 and Leahy (under review). Given the amount of data we have, we decided to use frequentist testing and used a GLMM. We also conducted a Bayesian analysis, and it provided converging evidence with the frequentist analysis. See both analyses in our code script on OSF. Additionally, we pre-registered that we would conduct a secondary analysis to check whether the number of children who chose the target gumball machine on the first trial differs from the number of children who placed the wagon under the nonbranching slide on the first slides trial. We did not proceed with the analysis as the average first trial performance on the gumball task, and the slides task in Experiment 2 was the same, 75%.

A GLMM was fit predicting the probability of placing the wagon under the single slide from the fixed effects of gender (male vs. female), age (age 3.0-3.50 vs. 3.5-3.99), and order (i.e., which side the single slide was on, which side of the y-shape tube the marble exited on the first trial on participant's responses in the slides task) with a random intercept for participant id. There was no effect of gender, ($\hat{\beta}$ = 0.35, *SE* = 0.42, *z* = 0.85, *p* = 0.39), or order ($\hat{\beta}$ = 0.07, *SE* = 0.40, *z* = 0.18, *p* = 0.86). There was an effect of age ($\hat{\beta}$ = 0.98, *SE* = 0.41, *z* = 2.38, *p* = 0.02), older three-year old's (range 3.5-3.99) selected the single slide 71% while younger three-year old's (range = 3.0-3.5) selected it 68% of the time.

3.5.4 Experiment 2 Analysis: Gumball Task

The mean proportion of correct responses across participants in the Experiment 2 gumball task was 86%. A GLMM was fit predicting the probability of choosing the target gumball machine from study type (Experiment 1 vs. Experiment 2). There was no effect of study type ($\hat{\beta}$ = 0.63, *SE* = 0.62, *z* =1.02, *p* = 0.31). Thus, we replicated the results from experiment 1. We then used the emmean package (Lenth, 2022) in R to get the estimated marginal mean for Experiment 2 from the model and compared it to chance performance (50%) and the performance predicted under the minimal representation hypothesis (75%). The average participant's performance (Probability Correct = 91%, 95% CI = [.83, .95]) was significantly above 50% (*z* = 6.33, *p* < .001), and significantly above 75% (*z* = 3.26, *p* = .001). We also provide converging evidence from a Wilcox test which compared participants' proportion correct across all six trials to 50% (*t*(35) = 526, *p* = < .001, *d* = .86), and in a separate test, to 75% (*t*(35) = 524, *p* = 0.0018, *d* = .52).

3.5.5 Experiment 2 Analysis: Slides Task

The mean performance on the slides task in Experiment 2 was 69%, whereas the mean performance on the slides task from Leahy (under review) was 61%. We compared the average participant's performance in the new slides task (Expt. 2) to participant's performance in the original slides task by Leahy (under review). A GLMM was fit predicting participants' responses from study (Leahy's slides task vs Experiment 2 slides task). There was no significant effect of study on participants' responses ($\hat{\beta}$ = -0.42, *SE* = 0.31, *z* = -1.37, *p* = 0.17). Thus, the performance on the slides task in Experiment 2 did not significantly differ from the performance in Leahy (under review). Completing the gumball task before completing the slides task did not significantly improve performance on the slides task.

3.5.6 Secondary Analysis: Correlation between children's performance on the gumball task and slides task

A Pearson correlation test found no significant correlation between participants' performance on the gumball task and the slides task (r(34) = .20, *p* = 0.22). We also provide converging evidence from a non-parametric Kendall rank correlation test (r_τ = .13, *p* = .38). Thus, children who performed well on the gumball task did not necessarily perform well on the slides task.

3.6 Discussion

Experiment 2 had three main findings. First, it replicated the results from Experiment 1 with a larger sample. Three-year-old children chose the gumball machine that *must* produce the desired gumball in the gumball task. They succeeded on this task without being primed with the probability trials involving large sets of marbles. Second, performing the gumball task prior to the slides task did not significantly improve children's performance on the slides task. Finally,

there was no correlation between children's performance on the gumball task and the slides task. Children who succeeded on the gumball task did not necessarily succeed on the slides task.

This pattern of results provides some support for the idea that quantifying over objects may be easier than quantifying over trajectories. However, these results are also open to other interpretations, which we will discuss in detail below.

4. General Discussion

Results from the gumball task in Experiments 1 and 2 provide preliminary evidence for the modal representation of possibility hypothesis, that three-year-old children have modal concepts and can reason about multiple incompatible possibilities. When presented with a gumball machine that *must* produce the desired gumball, and a gumball machine that *might* produce the desired gumball, children overwhelmingly chose the former. Children's success on the task speaks against the minimal representation of the possibility hypothesis. Children were able to distinguish between what *might* be and what *is*. The gumball task has a similar logical structure to previous tasks investigating children's modal concepts, but it does not involve occlusion or quantifying possibilities over locations or trajectories. It may be the case that quantifying over objects is easier than quantifying over locations/trajectories, and by using a simpler task, we revealed earlier competence in modal reasoning in children.

However, there is at least one alternative interpretation of our results. Perhaps three-year-old children succeeded in the gumball task by using a simple heuristic: avoid the gumball machine that contained an undesirable gumball. A previous infant study investigating probabilistic reasoning (Denison & Xu, 2014, Expt. 4) adopted a complex design with three sets of objects and showed that infants did not simply avoid a jar with more undesirable objects. This

suggests that children do not always adopt this simple heuristic but of course three-year-olds may behave differently from infants. To address this alternative interpretation, we will ask three-year-olds to choose between a gumball machine that contains 1 undesirable gumball and 1 desirable gumball, and a gumball machine that contains 1 undesirable gumball and 2 desirable gumballs. This study is now ongoing in our lab.

If we are able to rule out the simple heuristic interpretation with our control study, we will be left with the question of why completing the gumball task prior to the slides task did not improve three-year-olds' performance on the slides task. We see two possible explanations. One explanation is that perhaps there is some task difficulty in the slides task (and the three-containers tasks) that prevents children from deploying representations of possibilities. We offer two suggestions. First, the two different color gumballs in one of the machines may serve as a reminder for children that there are two possible outcomes. In the gumball task, all gumballs are present throughout the test trials. If a child simulates pressing the handle and one of the gumballs falls out, the other gumball is still present in the gumball machine and may serve as a reminder for the child (in her mind's eye) that there is an alternative possible outcome. In contrast, in the three-containers task or the slides task, if a child simulates the reward being dropped into one of two containers or the marble coming out of the right exit of the Y-shaped slide, there is nothing left to serve as a visual reminder for the child that an alternative outcome is possible. Without a visual reminder of the alternative possibility, the child may decide to accept that one simulation and not simulate anymore, resulting in the behavior described under the minimal representation hypothesis. This interpretation is also supported by the many studies with infants investigating their concept of an object: young infants believe that an object can trace exactly one spatiotemporally connected path (e.g., Spelke et al. 1995; Xu & Carey, 1996). In the three-

containers task and the slides task, simulating an alternative outcome requires the child to take the reward or marble and imagine an alternative path for it.

A second explanation for children's success and failure on these tasks is that it may be easier for children to quantify over objects than locations/trajectories, as we discussed in Experiment 1. In the gumball task, the gumballs are laid out in front of the children and are visible throughout the test trials. The child may be able to succeed in the task by quantifying over the gumballs in the gumball machine. In the three-containers task or the slides task, success requires that the child tracks the reward as it goes behind the occluder or the marble as it goes down the slides. It is perhaps easier for young children to quantify over objects like gumballs than quantify over locations or trajectories. More generally, perhaps quantifying over individuals (e.g., objects, persons) may be easier than quantifying over events (e.g., locations, trajectories).

The gumball task was adapted from previous probability tasks with infants (see Denison & Xu, 2019 for a review), which raises the question of whether three-year-olds could have succeeded in our experiments by computing probabilities as opposed to possibilities. In other words, the lack of improvement on the slides task in Experiment 2 may indicate that children relied on different cognitive mechanisms in these tasks. Children may have succeeded on the gumball tasks using a form of probabilistic reasoning which does not require modal concepts. They can compute the probabilities of getting the desired gumball color in each of the machines by computing the ratios of desired gumballs over the total number of gumballs. This computational strategy underlies infants' success in various probability tasks with large sets (e.g., 40-80 Ping Pong balls or lollipops; Denison & Xu, 2010, 2014; Xu & Garcia, 2008). When infants are shown a box filled with many red balls and a few white ones, they can compute the

probability of drawing a red ball by estimating the number of red balls, using the approximate number system (ANS), and divide it by an estimate of the total number of balls, also using ANS.

Can children's success in the gumball task be achieved through computing ratios among objects rather than representing incompatible possibilities? We don't think so. In the gumball task, there are only one or two objects to attend to in each gumball machine, and many studies with children and adults suggest that when only a few objects are present (up to 3 or 4), the object tracking system is engaged to track each object individually (Feigenson, Dehaene & Spelke, 2004; Scholl & Pylyshyn, 1999; Xu, 2003; Carey & Xu, 2001). Thus, it is most likely that children use the object tracking system to enumerate the number of possibilities in each gumball machine (1 or 2).

Two recent studies, however, suggest that children and adults can use ANS on small sets under some circumstances (Hyde & Wood, 2011; Posid & Cordes, 2015). Specifically, ANS can be used to track a small number of objects when the attentional load is high or when work memory is taxed (i.e., when the object tracking system is unavailable) (Posid & Cordes, 2015). Given the circumstances when ANS is applied to small sets, it is again unlikely that the gumball task is prompting children to engage in ANS rather than object tracking.

An important theoretical issue that our study and other related studies raise is the relationship between possibility and probability. Téglás and colleagues (2007, 2015) proposed that with a small number of objects, infants and children employ intuitions of probability that are derived from enumerating possibilities. In one study, Téglás et al. (2007) showed 12-month-old infants a lottery machine that contained 1 blue and 3 yellow objects, and on the test trials one object exited the machine. They found that infants looked longer at the blue object exiting the

machine (the improbable outcome) than one of the yellow objects exiting the machine (the probable outcome). Téglás and colleagues proposed that infants enumerated the number of possible outcomes in the 1 vs. 3 experiment, a total of 4. Since three of these outcomes involved a yellow object and only one involved a blue object, the infants concluded that it was more probable for a yellow object to exit the machine than a blue object. However, Leahy and Carey (2020) offered a deflationary account of these findings: they pointed out that in the 1 vs. 3 experiment, if each infant fixates on one of the four objects and ignores the rest, then 75% of them will focus on a yellow object and 25% of them will focus on a blue object. Then when a blue object falls out, most infants will be surprised and look longer at it than when a yellow object falls out. On their account, infants may not have enumerated the number of possibilities and derived a probability distribution when they succeed in this task.[5]

The relationship between possibility and probability is, however, more complex than the Téglás and colleagues model suggests. When we encounter multiple incompatible possibilities (e.g., when the handle is pressed for the two-gumball machine, one *or* the other gumball will fall out), it is not always the case that each possibility is equally probable. When young children are presented with a small number of objects, they can track them with their attention system (Scholl & Pylyshyn, 1999; Carey & Xu, 2001). Each of these objects may be considered as a possibility for some future outcome. The default assumption of the learner may be that each object has an equal probability as the outcome of some future event. Under these circumstances, the number of

[5] For the gumball task, if each child fixated on one of the two gumballs in the gumball machine, they would choose the target gumball machine 75% of the time. This is the same prediction made by the minimal representation of possibility hypothesis. As our results show, children's performance was significantly higher than 75%. By three years of age, children can enumerate the number of possibilities, simulate what happens when a handle is pressed, and use the simulated outcome to guide their action.

possibilities (e.g., blue, yellow, yellow, yellow) naturally gives rise to a probability distribution that says that the yellow outcome is 3 times as probable as the blue outcome. However, in many cases, the number of possibilities is dissociable from how a probability distribution is arrived at. For example, if the blue object is for some reason always closer to the exit of the lottery machine than the three yellow objects (Téglás et al. 2011), then the learner may update their probability distribution to include that information, and adjusts their predictions as to which object, blue or yellow, is more likely to exit the machine. In general, alternative possibilities are often not all equally probable, thus dissociating our representations of possibility from our representations of probability distributions. It is the probability of each possibility, not just the number of possibilities, that guides our predictions about which future outcome is more or less likely.

The recent discussion of possibility and modal concepts is inspired by work in logic and formal semantics (e.g., Leahy & Carey, 2020; Kratzer, 2012; Phillips & Knobe, 2018), whereas much of the research on probabilistic reasoning in infants and children is inspired by the development of Bayesian probabilistic models (e.g., Denison & Xu, 2019; Gopnik & Wellman, 2012; Xu, 2019; Xu & Tenenbaum, 2007). These two communities of researchers have not traditionally been in dialogue with each other. A notable exception is the work by Teglas, Bonatti, and colleagues where they discuss the idea that enumerating possibilities is how infant learners derive a probability distribution. Furthermore, this probability distribution allows the learner to predict the more likely outcome (Teglas et al., 2007, 2015). We hope that our study will spark future conversations between scholars in logic, semantics, and computational modeling.

In conclusion, our study provides preliminary evidence for the hypothesis that three-year-old children can represent multiple incompatible possibilities and have some rudimentary form

of modal concepts. The current study also raises many new questions for future research. Given three-year-old children's success on the gumball task, one might ask whether children much younger, perhaps even infants, can succeed on the task. Additionally, future research needs to adjudicate the task differences that could be causing differences in performance on the gumball task, and the other modal reasoning tasks (Leahy, under review; Leahy et al.,2022; Mody & Carey, 2016). Finally, more theoretical work is needed to explicate the relationship between representations of possibility and representations of probability, and how these representations are deployed in inferential reasoning.

Acknowledgments: [redacted]

References

Carey, S. & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, *80*, 179-213. https://doi.org/10.1016/S0010-0277(00)00154-2

Denison, S., & Xu, F. (2010). Twelve- to fourteen-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, 13, 798-803. https://doi.org/10.1111/j.1467-7687.2009.00943.x

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition,* 130(3), 335-347.https://doi.org/10.1016/j.cognition.2013.12.001

Denison, S., & Xu, F. (2019). Infant Statisticians: The Origins of Reasoning Under Uncertainty. *Perspectives on Psychological Science, 14*(4), 499-509. https://doi.org/10.1177/17456916198472

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Engelmann, J.M., Völter, C.J., O'Madagain, C., Proft, M., Haun, D.B.M., Rakoczy, H. & Herrmann, E. (2021) Chimpanzees consider alternative possibilities. *Current Biology*, 31(20), R1377-R1378.https://doi.org/10.1016/j.cub.2021.09.012

Feigenson, L., Dehaene, S., & Spelke, E.S. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(10), 307-314.https://doi.org/10.1016/j.tics.2004.05.002

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138, 1085–1108. http://dx.doi.org/10.1037/a0028044

Grigoroglou, Chan, S., & Ganea, P. A. (2019). Toddlers' understanding and use of verbal negation in inferential reasoning search tasks. *Journal of Experimental Child Psychology*, 183, 222–241. https://doi.org/10.1016/j.jecp.2019.02.004

Hyde, D., & Wood, J.N. (2011). Spatial attention determines the nature of non-verbal numerical cognition. *Journal of Cognitive Neuroscience*, 23(9),2336-2351. https://doi.org/10.1162/jocn.2010.21581

Kratzer, A. (2012). Modals and Conditionals: New and Revised Perspectives. *Oxford Studies in Theoretical Linguistics*. https://doi.org/10.1093/acprof:oso/9780199234684.001.0001

Leahy, B.P. (under review). Don't you see the possibilities? Evidence that preschoolers lack modal concepts.

Leahy, B. P., Huemer, M., Steele, M., Alderete, S., & Carey, S., (2022). Minimal representations of possibilities at Age 3. *Proceedings of the National Academy of Sciences - PNAS.* https://doi.org/10.1073/pnas.2207499119

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*,*24*(1), 65-78.https://doi.org/10.1016/j.tics.2019.11.004

Leahy, B. P. & Zalnieriunas, E. (2021). Might and might not: Children's conceptual development and the acquisition of modal verbs. *Proceedings from Semantics and Linguistic Theory*, *31*, 426. https://doi.org/10.3765/salt.v31i0.5082

Lenth R (2022). _emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.8.2 https://CRAN.R-project.org/package=emmeans

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40-48.https://doi.org/10.1016/j.cognition.2016.05.012

Posid, T. & Cordes, S. (2015).The small-large divide: A case of incompatible numerical representations in infancy. In D. Geary, D. Berch, & K. Mann-Koepke (Eds.), *Evolutionary Origins and Early Development of Basic Number Processing (pp. 253-276) Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-420133-0.00010-7*

Phillips, J. and Knobe, J. (2018) The psychological representation of modality. Mind Lang. 33, 65–94. doi: 10.1111/mila.12165

Redshaw, J. & Suddendorf, T. (2020). Temporal Junctures in the Mind. *Trends in Cognitive Sciences*, *24*(1), 52–64. https://doi.org/10.1016/j.tics.2019.10.009

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood. *Cognitive Psychology*, 38(2), 259

Shtulman, A. & Phillips, J. (2018) Differentiating "could" from "should": developmental. changes in modal cognition. *Journal of Experimental Child Psychology, 165,* 161-182.

Spelke, E.S., Kestenbaum, R., Simons, D.J., and Wein, D. (1995) Spatiotemporal continuity, smoothness of motion and object identity in infancy. British Journal of Developmental Psychology, 13, 113–142.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, 104(48), 19156-19159.https://doi.org/10.1073/pnas.0700271104

Téglás, E., Ibanez-Lillo, A., Costa, A., & Bonatti, L. L. (2015). Numerical representations and intuitions of probabilities at 12 months. *Developmental Science*, 18(2), 183-193. https://doi.org/10.1111/desc.12196

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332(6033),*1054-1059. https://doi.org/10.1126/science.1196404

U.S. Census Bureau (n.d.) Quick Facts, Alameda County, California. Retrieved from https://www.census.gov/quickfacts/alamedacountycalifornia#

Vaidya, A. (2009). The epistemology of modality. *Stanford Encyclopedia of Philosophy*.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. Cognitive Psychology, 30(2), 111–153. https://doi.org/10.1006/cogp.1996.0005

Xu, F. & Tenenbaum, J.B. (2007) Word learning as Bayesian inference. *Psychological Review, 114,* 245-272. doi: 10.1037/0033-295X.114.2.245

Xu, F. (2019) Towards a rational constructivist theory of cognitive development. *Psychological Review, 126, 841-864.* http://dx.doi.org/10.1037/rev0000153

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89(1), B15-B25.https://doi.org/10.1016/S0010-0277(03)00050-7

Xu, F., & Garcia, V. (2008). Intuitive Statistics by 8-Month-Old Infants. *Proceedings of the National Academy of Sciences - PNAS, 105*(13), 5012-5015. https://doi.org/10.1073/pnas.0704450105