Journal of Experimental Child Psychology 183 (2019) 222–241

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Toddlers' understanding and use of verbal negation in inferential reasoning search tasks

Myrto Grigoroglou, Sharon Chan, Patricia A. Ganea *

University of Toronto, Toronto, Ontario M5S 1V6, Canada

ARTICLE INFO

Article history:
Received 11 April 2018
Revised 12 February 2019

Keywords:
Negation
Disjunctive syllogism
Reasoning by elimination
Inference
Logic
Language development

ABSTRACT

Past research has demonstrated that young children and nonhuman animals are able to reason by elimination ("If not A, then B") by relying on visual cues such as seeing that one container is empty. Other research has shown that young children can solve similar, simple inferential reasoning tasks where "emptiness" is conveyed verbally through negation (e.g., "The toy is not in the box"). However, it is unclear whether these tasks involved reasoning through the disjunctive syllogism, which requires the representation of logical negation (NOT A) and disjunction (A OR B) or simpler, nondeductive strategies. In Study 1, we extended this work by investigating whether 2-year-olds can infer the location of a toy in typical two-location elimination trials, when given both affirmative and negative sentences, and more complex three-location trials, when information about emptiness was conveyed verbally and visually. Younger 2-year-olds performed significantly better on the search task when hearing affirmative than negative sentences, whereas older 2-year-olds were equally successful with both types of sentences. Study 2 examined children's ability to use verbal negation to solve a more complex deductive task involving disjunctive syllogism. Results showed that, in this linguistic version of the disjunctive reasoning task, both 2.5- and 3-year-olds made accurate inferences about the location of a reward, unlike prior (nonlinguistic) evidence that demonstrated this ability in 3-year-olds but not in younger children. We conclude that by the end of their second year of life, children have a robust understanding of negation which they can apply in abstract reasoning.

* Corresponding author.
E-mail address: patricia.ganea@utoronto.ca (P.A. Ganea).

https://doi.org/10.1016/j.jecp.2019.02.004
0022-0965/

Introduction

An important question regarding human cognition is the extent to which prelinguistic infants and other nonlinguistic species share the same type of reasoning abilities as language-speaking adults. Some have argued that logic-like representations may be available to other species and human infants in a nonlinguistic format (Bräuer, Kaminski, Riedel, Call, & Tomasello, 2006; Call, 2004, 2006; Fodor, 1970). Others have argued that language is essential for logical thought and that it is only with the acquisition of language that individuals can begin to represent proposition-like arguments (Jackendoff, 1996, 2011). Different experimental paradigms have been used to explore these two theoretical positions. One line of research has investigated whether animals and human children have a representation of logical operators such as OR and NOT in nonlinguistic format in tasks that required them to reason by elimination (e.g., Call, 2004, 2006; Hill, Collier-Baker, & Suddendorf, 2012; Mody & Carey, 2016; Watson et al., 2001).

Reasoning by elimination is a logical process of becoming increasingly sure of an answer as you eliminate other possibilities. In its simplest form, when an outcome is constrained to A or B, knowing that the answer is not A leads to the deduction that it must be B. Several variations of elimination tasks are used in these studies, but in the basic paradigm participants infer the location of a target between a choice of two containers. In one version, participants may be acquainted with two different contents, shown the contents of one container, and must infer the contents of the remaining container. Alternatively, participants are tasked with finding a target between two containers, shown that one container is empty, and must infer that the desired contents reside in the remaining container. Research using such paradigms has demonstrated a basic ability to reason by elimination in a variety of animals, including great apes (Bräuer et al., 2006) and several species of birds such as ravens (Schloegl et al., 2009) and parrots (Pepperberg, Koepke, Livingston, Girard, & Hartsfield, 2013). Young children are also shown to pass similar reasoning tasks. In one study with 4-year-olds, one cup was baited with banana and another cup was baited with apple. After a brief interval in which the cups were hidden from view, the children watched as the baiter ate one of the treats, for example, banana. When given a chance to search for a treat, nearly all of the children in this study directly searched the cup that held apples, showing that they were able to piece together information to infer which cup still held its original contents and which did not (Premack & Premack, 1994). This finding has been replicated in children aged 3 and 4 years across visual and auditory modalities; for example, the absence of a rattling sound in one cup allowed children to infer that the target was in the alternative cup (Hill et al., 2012). The conclusion from this line of work is that both animals and young children have a representation of the concepts OR and NOT in nonlinguistic format and can use them in higher order logical operations.

However, more recent work has challenged this "rich" interpretation of the findings and suggested two other possible reasoning strategies that may underlie successful performance in these tasks (Mody & Carey, 2016). According to the simplest interpretation ("avoid empty"), participants see that A is empty, omit searching in the empty location, but search in B because it is the only other alternative, without deducing that the reward is *necessarily* in B. Evidently, if children and animals use this strategy, the representation of the operators OR and NOT is not required. On another interpretation ("maybe A, maybe B"), participants see that A is empty and search in B, merely as another *possible* location, without deducing that the reward is *necessarily* in B. This strategy requires the representation (or at least implementation) of the concept NOT but not of OR. To identify the exact strategy children use when solving search tasks, Mody and Carey (2016) used a modification of the classic two-cups task. In their paradigm, two pairs of cups were used, with a target hidden in each pair unseen by the children. The children were then shown an empty cup from one of the pairs, and given the opportunity to search. Children who employ lower level elimination strategies are equally likely to choose among the remaining three cups. On the other hand, children who use higher order reasoning of the form "A OR B, NOT A; THEREFORE B" (referred to as the *disjunctive syllogism*) should choose the cup that was paired with the empty cup because this is the only cup that they know *necessarily* holds a target, given that its partner was empty. The children have insufficient information to make the same deduction for the alternative pair of cups. Results of this study showed that 3-, 4-, and 5-year-olds searched the target cup significantly more often than chance, whereas 2.5-year-olds did not. Other research corroborates this finding by showing that 2.5-year-old children and nonhuman animals had difficulties in updating their inferences about the location of a reward hidden in one of three containers (Call & Carpenter, 2001; Watson et al., 2001; see Mody & Carey, 2016, for a discussion), whereas older children (4- to 6-year-olds) showed clear signs of inferential updating (Watson et al., 2001). Therefore, by their third birthday (but not before), children are using true logical representations to solve the task rather than the simplest alternative strategies described above.

Although the reason children under 3 years of age (as well as other nonhuman species) have problems in deriving logical inferences is currently unclear, it is possible that very young children are not capable of engaging in abstract combinatorial thinking and that logical operators such as NOT and OR are not available to them early on. Developmental patterns from children's acquisition of linguistic disjunction ("or") and negation ("no", "not") provide suggestive evidence in favor of this possibility. Children begin using "or" in their speech at around their third birthday (e.g., Bloom, Lahey, Hood, Lifter, & Fiess, 1980; French & Nelson, 1985; Lust & Mervis, 1980; Morris, 2008), and although they produce "no" at a fairly early age (15–18 months), it is not until 27 months that they begin using "not" as a fully developed truth-functional operator (i.e., as a word that changes the truth value of propositions; e.g., Bloom, 1970; Choi, 1988; Hummer, Wimmer, & Antes, 1993; Pea, 1980, 1982). This homology between children's language use and nonlinguistic reasoning can be informative for theories of language acquisition because it suggests that the immaturity of conceptual representations may be responsible for the (relatively) late acquisition of words with highly abstract logical meanings such as "or" and "not." The current set of studies explored the relation between young children's linguistic and logical representations by asking 2- and 3-year-olds to solve two inferential reasoning tasks where the cues for the location of the object were presented verbally instead of visually.

Only a few studies have examined children's use of language to reason by elimination, and most of these used negative utterances as linguistic cues for "emptiness" (e.g., "X is not in A"; Austin, Theakston, Lieven, & Tomasello, 2014; Feiman, Mody, Sanborn, & Carey, 2017; Mascaro & Sperber, 2009; but see Hill et al., 2012, for use of affirmative linguistic cues). In one study, Austin et al. (2014) looked at 20- to 30-month-olds' searching behavior following an exchange between two experimenters, one who hid a toy between two containers (e.g., a truck and a bucket) and one who provided an affirmative or negative sentence (e.g., "It's in the bucket" or "It's not in the bucket"). Results showed that 24- and 28-month-olds successfully searched for the toy after hearing a negative sentence. Perhaps counterintuitively, in this study comprehension of propositional negation seemed to precede that of affirmation, with children aged 24–26 months being more successful in their searches in the presence of negative cues as opposed to affirmative cues. Similarly, in a more recent study, Feiman et al. (2017) found that 27-month-olds (but not younger children) successfully used negative cues (e.g., "The ball is not in the bucket") to infer the location of a toy. Contrary to Austin et al. (2014), in this study, children's comprehension of negation followed that of affirmation, with even 20-month-olds showing robust comprehension of affirmative sentences. Together, these studies show that, by 27 months of age, children can reliably use negative sentences to reason by elimination.

Although this evidence was taken to suggest that children can understand truth-functional negation before age 3, children's comprehension of negation in contexts that do not involve reasoning by elimination seems to be particularly protracted. For instance, in one study where 3- to 5-year-olds were asked to evaluate affirmative and negative statements that were either true or false (e.g., "This is not an apple" in the presence of a picture depicting a banana), even 5-year-olds had difficulty in evaluating negative statements (Kim, 1985). Although children's failures in this task may be in part due to the pragmatic infelicity of the context and the metalinguistic demands involved in needing to explicitly reject or validate statements, children's difficulties persist even in tasks that do not require explicit judgments and provide more supportive context. In an eye-tracking study, 2- to 5-year-olds failed to establish an appropriate reference between two pictures (e.g., a boy with no apples and a boy with apples) when hearing a negative sentence (e.g., "Look at the boy with no apples"; Nordmeyer & Frank, 2014), whereas more recent eye-tracking evidence suggests that older 2-year-olds (shown to comprehend negation in reasoning by elimination tasks) still face semantic difficulties in processing negative statements (Reuter, Feiman, & Snedeker, 2018).

Given these persistent difficulties, the nature and robustness of toddlers' understanding of negation and their ability to use it in logical reasoning remains open. In particular, it is unclear whether 27-month-olds' success in reasoning by elimination tasks (Austin et al., 2014; Feiman et al., 2017) should be taken to reflect true comprehension of truth-functional negation (which assumes the representation of conceptual truth-functional negation) and disjunctive reasoning or can be explained by less sophisticated linguistic and inferential abilities. Because such studies have used elimination paradigms similar to the two-cups task, their findings are subject to the same criticism as the nonlinguistic tasks (see Mody & Carey, 2016; cf. Feiman et al., 2017, p. 446, for a relevant discussion). On one interpretation, children who succeeded in Austin et al. (2014) and Feiman et al. (2017), may have had no beliefs about the location of the toy and simply avoided the location they were told did not contain the toy (i.e., an equivalent of the "avoid empty" strategy). Alternatively, children may have not represented the dependent relation between the two possible hiding locations of the toy and, on hearing that the toy was not in A, searched in B (i.e., an equivalent of the "maybe A, maybe B" strategy). Crucially, both of these strategies presuppose some understanding of the negative verbal cue, but importantly they do not require a full-fledged truth-functional interpretation of negation (i.e., as an operator that flips the truth value of a proposition). In fact, the data of Austin et al. (2014) and Feiman et al. (2017) are entirely compatible with the possibility that children took sentences such as "The ball is not in the bucket" to express the concept of nonexistence negation (e.g., "the bucket is empty"), a function of negation that is conceptually less abstract than truth-functional negation (because it is a positive concept that can be learned through direct evidence of "emptiness") and is present in children's speech about a year earlier than truth-functional negation (see Bloom, 1970; Choi, 1988; Dale & Fenson, 1996; Hummer et al., 1993; Pea, 1980, 1982). Therefore, prior work does not elucidate the mechanisms that led to children's success in these verbal reasoning tasks and does not fully speak to the relation between children's linguistic representations of negation and the corresponding logical representations.

In the current research, we addressed these limitations in two studies. In Study 1, we tested younger and older 2-year-old children's ability to solve an inferential reasoning problem using verbal negation in the absence of any visual information that could indicate where the toy is or is not located. Given the reported inconsistency in the literature in terms of the timeline of children's understanding of propositional negation and affirmation, we compared toddlers' searching behavior with both negative and affirmative sentences. We hypothesized that children would be more successful in their searches when hearing affirmative as opposed to negative sentences for two reasons. First, processing of a negative sentence is more complex than processing an affirmative sentence, as processing negation involves understanding the semantics of the corresponding affirmative sentence in addition to the meaning of the negative operator; see Clark & Chase, 1972, 1974; Kaup, Lüdtke, & Zwaan, 2006; Orenes, Beltrán, & Santamaría, 2014). Second, in a search task, affirmative trials should be conceptually easier than negative trials given that they do not involve logical inferencing because the location of the toy is explicitly stated. Similarly to Austin et al. (2014) and Feiman et al. (2017), Study 1 included search trials where children were asked to reason about the location of a toy between two alternatives, as a linguistic equivalent of the standard two-cups task. However, our Study 1 also included more complex search trials in which children were asked to infer a toy's location across three containers by combining verbal negation (hearing that a toy is not in a particular location) and visual negation (in the form of seeing an empty container). We reasoned that the ability to combine information across modalities when making logical inferences about the location of a toy across three containers would suggest that children are not simply eliminating options but can flexibly combine concepts, independent of how these were conveyed, and use them in abstract deductive computations (cf. Call & Carpenter, 2001, and Watson et al., 2001, for evidence from three-location tasks).

In Study 2, we further tested children's application of this knowledge in a task that involved reasoning with the disjunctive syllogism, which requires attributing certainty to a single deduced possibility rather than simply avoiding eliminated options. Study 2 was a variation of Mody and Carey (2016) four-cups task, where the information about one cup not containing a prize was conveyed linguistically instead of visually. We reasoned that children's success in this more complex task, which requires a robust representation of the concepts of logical negation and disjunction, would demonstrate that children not only comprehend truth-functional negation but also are capable of using it in the disjunctive syllogism.

Study 1

Method

Participants

Children included in the study were required to be fluent in English, noted by parents as hearing and using English at least 75% of the time, cumulatively across contexts (e.g., home, daycare, school). Children were recruited from a database of parents who had voluntarily given their contact information to be approached for participation in research studies at recruitment venues as well as through ongoing flyer distributions. All children included in the sample were tested in a research lab. The final sample consisted of 25 younger 2-year-olds (M_{age} = 24.21 months, range = 23.15–26.1; 12 girls) and 21 older 2-year-olds (M_{age} = 35.39 months, range = 34.2–36.9; 9 girls). An additional 5 younger 2-year-olds were excluded from analysis due to failure on labeling during the familiarization phase (n = 3), spoiled trials (e.g., parent help or interference, n = 1), or experimenter error (n = 1). All 21 older 2-year-olds who were tested were included in the study. Ethnic background was reported for 36 of the 46 children, and the majority of the sample (75%) was identified as Caucasian White. Parent education level was reported by 41 parents, and the majority of the sample (95%) had completed a college education.

Materials

The task involved finding a hidden toy among several containers of similar size that were comfortably held by children. Two opaque purple- or green-lidded toy garbage bins and one transparent container with a clear lid of similar size (3.74 × 3.03 × 5.51 inches) were used in teaching and control trials. Four additional opaque containers were used for test trials, and these included a small purple cardboard box (3.94 × 3.94 × 2.20 inches), a yellow cup (3.15 × 3.94 inches), a plastic bag (3 × 5 inches), and a pink bowl (5.31 × 2.76 inches). All containers had a matching color lid to conceal the interior except for the bag, which had a Velcro closure. Pilot testing revealed that the majority (18 of 20) of young 2-year-olds were able to consistently identify the four containers.

The task was administered on a child-size table. A large white tablecloth spread over the table was used to cover the containers below when not in use and to conceal the transfer of the toy between containers. The child was seated approximately 1 m away from the experimenter, enough distance to ensure that the child could not easily reach for containers while hearing the verbal information. A small stuffed dog named George served as the target throughout the experiment. A soft toy that fit easily inside all containers was selected so that there would be no auditory cues that could be inadvertently used to indicate the correct hiding place. In all trials, the toy was always surreptitiously placed in the target location under a table. The table obscured the experimenter's arm movements while hiding the toy. One camera was built into the testing room facing the child to record the experimental session.

Procedure

Children were told that they would be playing a game with a silly dog named George who liked to hide in different places, and it was up to the children to find him. The study included three phases: a familiarization phase, a teaching phase, and a test phase.

Familiarization phase. During the familiarization phase, the containers were presented under the pretext of showing the child all the different things that George liked to hide inside. The child needed to point or gesture to clearly identify the four containers used in test trials (cup, bowl, bag, and box) when the child heard the label (e.g., "Can you show me the cup? Which one is the cup?"). Labels that were spontaneously offered by the child were also acceptable. If the child's response was ambiguous, the items were scrambled and presented again or, alternatively, the child was asked to label the items again once the experiment was over. Children's data were excluded if the children were not able to label all four containers correctly.

Teaching phase. The teaching phase consisted of one trial with the purpose of familiarizing children with the toy being hidden in one of the containers presented on the table. Two toy lidded garbage bins were used. While under the table, the toy was placed in either the purple or green bin. The two bins were then simultaneously placed on the table. The lid of each bin was flipped open one at a time for the child to peer inside, followed immediately by the test question, "Where's George?" Prompts such as, "Show me where he is hiding!" and "Where is the puppy?" were used for encouragement if the child did not immediately search for the toy. Praise was given for correct searches. If the child chose incorrectly, the task was repeated until the child chose correctly (children typically chose correctly after the second time). If the child made a motion to search under the table when the toy was being moved to a new location, the experimenter would remind the child that there was no peeking. If the child appeared to consider that the toy might be hiding under the table, the experimenter directed the child's attention back to containers on top of the table and clearly stated that the puppy was hiding in one of the places on the table. All children passed this trial.

Test phase. Affirmative trials. Similarly to the teaching trial, in affirmative trials the toy was hidden inside one of two containers out of the child's view (i.e., under the table). Following this, the containers were brought out simultaneously and held in full view but out of the child's reach to ensure that the child heard the entire verbal sentence prior to searching. Once the child was paying attention, the experimenter delivered an affirmative sentence indicating the toy's location (e.g., "George is in the box"). This sentence was repeated once more, and then immediately afterward the test question "Where's George?" was asked before lowering both bins simultaneously onto the table. If the child chose correctly, the toy was revealed and the child was praised. If the child chose incorrectly, the experimenter revealed the correct location. The same containers were always paired together in affirmative trials: cup and box or bag and bowl. One pair was used for each trial.

Negative trials. The procedure for negative trials was identical to that for affirmative trials except that children heard a negative sentence and needed to infer the location of a toy from this sentence. For example, between the bowl and bag pair, if the toy was in the bowl, the child heard, "George is not in the bag." The statement was repeated once, and feedback was given at the end of the child's search by revealing the correct location of the toy.

Three-location trials. The purpose of these trials was to test whether children could integrate visual information about absence and verbal negation to infer the location of a toy. Three containers were always used for these trials (e.g., bowl, bag, and cup). The experimenter first engaged the child's attention by saying, "Look" and then brought up the first container (e.g., bowl), opened the lid, and revealed to the child that it was empty. The container was closed again and placed on the table out of reach of the child. This part of the trial provided children with the visual input for the toy's absence. Immediately afterward, the remaining two containers were simultaneously brought up from under the table, and the relevant negative sentence (e.g., "George is not in the cup") was given and repeated once. Thus, the child needed to infer that, given the prior visual input and the new verbal input, the toy's location must be in the third remaining location (bag). Following this presentation, all three containers were randomly scrambled and placed parallel to each other in front of the child, and the test question "Where's George?" was asked. Praise was given for a correct response. If the child chose incorrectly, the experimenter responded with "Hmm, he's not in there. Maybe here?" and revealed the correct location.

Control trials. Control trials provided a visual equivalent of negative trials and were used to get a baseline indicator of performance when no verbal negation was required. Thus, one of the containers was transparent and the other container was opaque, and the toy was always hidden in the opaque bin. The child was prompted to look up when the experimenter said "Look!" as both bins were brought out from under the table at the same time and held out of the child's reach. The test question "Where's George?" was then asked before lowering both bins at the same time for the child to choose.

Participants began with the familiarization phase, followed by the teaching phase and then the test phase. During the test phase, they completed two of each type of test trial—negative (N), affirmative

(A), and three-location (3L) trials—as well as two control (C) trials. The order of test trials was counterbalanced so that half of the blocks began with an affirmative trial and half began with a negative trial (e.g., N, C, A, 3L). Control trials were always placed in between trials. Thus, each child completed two blocks of these test trials (see Table 1 for an example test sequence). In total, each child completed eight test trials. The task took 15 to 20 min to complete.

Pairs of containers were alternated such that if the box and cup were used for one of the two location trials, the bowl and bag were used for the second one. In 3L trials, the bowl, bag, and cup always were used for the first trial and the bowl, cup, and box always were used for the second trial. In test trials that involved two containers (N, C, and A), the location of the toy was counterbalanced between left and right (in 3L trials, the location of the toy was randomly switched across the three containers). In teaching trials, the toy was counterbalanced between the green and purple bins. The experimental sessions were videotaped for reliability coding.

Results

Reliability coding

Two independent raters, blind to the study hypotheses, coded children's responses using video recordings of the sessions. To count a response as a search, the child was required to point clearly to one container or to pick or search for the toy in a container. Inter-rater reliability analyses using the kappa statistic were performed to determine consistency among raters on each of the test trials. Reliability of coders was high on all test trials of interest, including N trials (kappa = 0.89, $p < 0.001$), A trials (kappa = 0.95, $p < 0.001$), and 3L trials (kappa = 0.92, $p < 0.001$). Disagreements were resolved by a third research assistant.

Overall analysis

We measured participants' search accuracy for each trial. This measure was a binary outcome variable coded as 1 (correct search) or 0 (incorrect search). The data were analyzed using multilevel logistic mixed-effects modeling with crossed random intercepts for participants and items (Baayen, 2008; Baayen, Davidson, & Bates, 2008). This analytical approach allows for participants and items to be treated as random factors in a single model and for the appropriate treatment of categorical data (Fraundorf, Benjamin, & Watson, 2013; Jaeger, 2008; cf. Barr, 2008). All models were fit using the "glmer" function of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2018). Participants' correct searches are summarized in Fig. 1. Control trials were not included in the analysis because their purpose was to ensure that children performed at celling when there was no verbal information, which was found to be the case.

We first performed an overall analysis to test our main, hypothesis-driven predictions concerning the types of trials children would be more successful at and possible developmental differences in their performance. We predicted that children would be more successful in their searches when hearing an affirmative statement as opposed to a negative statement (in negative and three-location

Table 1
Example test sequence in Study 1.

Trial type	Materials	Information presented to child	Correct response
Teaching	Green bin and purple bin	Sees each container flipped open	Green bin
Negative	Cup and box	Hears "George is not in the cup"	Box
Control	Opaque bin and transparent bin	Sees both containers presented	Opaque
Affirmative	Bowl and bag	Hears "George is in the bowl"	Bowl
Three-location	Bowl, bag, and cup	Sees bowl empty; hears "George is not in the bag"	Cup
Affirmative	Cup and box	Hears "George is in the bowl"	Bowl
Control	Opaque bin and transparent bin	Sees both containers presented	Opaque
Negative	Bowl and bag	Hears "George is not in the bag"	Bowl
Three-location	Bowl, box, and cup	Sees box empty; hears "George is not in the cup"	Bowl

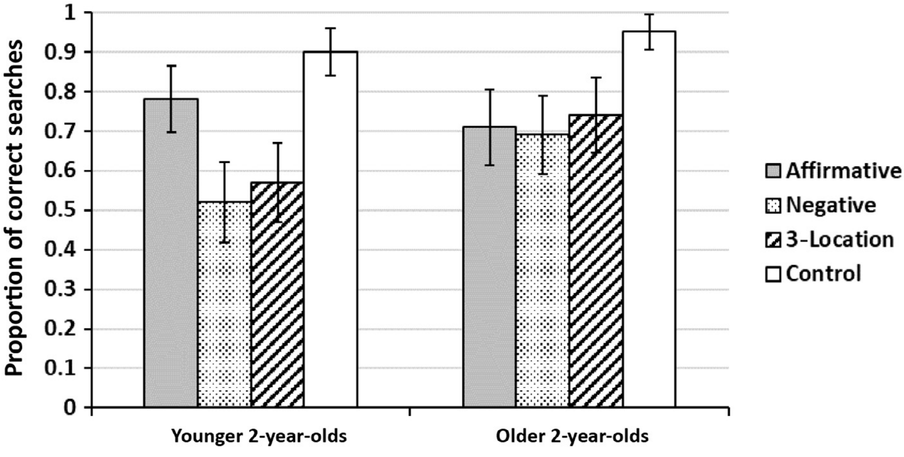

Fig. 1. Proportion of correct searches per age group and trial type in Study 1. Error bars represent standard errors.

trials); three-location trials, as cognitively more complex, might be more difficult than negative trials. We also predicted that older 2-year-olds would show better performance in N trials than younger 2-year-olds. The model included fixed effects of trial type (affirmative, negative, or three-location) as a first-level predictor, age (younger or older 2-year-olds) as a second-level predictor, and their interaction. The model also included random intercepts for participants and items. The fixed effect of age was coded with centered contrasts (−0.5, 0.5), and the fixed effect of trial type was analyzed with two simple contrasts comparing affirmative trials with all negative trials combined (c_1: −0.66, 0.33, 0.33) and comparing negative trials with three-location trials (c_2: 0, −0.5, 0.5). Parameter estimates for this model are summarized in Table 2. Results showed a significant interaction between trial type and age; younger 2-year-olds were more accurate in their searches in affirmative trials as opposed to negative and three-location trials combined (M_A = 0.78 vs. $M_{N\&3L}$ = 0.55), whereas older 2-year-olds were equally accurate in both types of trials (M_A = 0.71 vs. $M_{N\&3L}$ = 0.72); negative and three-location trials did not differ in any age group.

Learning, order, location, and container effects

To assess the possibility that children's performance was affected by factors beyond our main (hypothesis-driven) manipulations, we performed an exploratory analysis to examine whether including additional fixed effects in the analysis would improve the fit of the model to our data. In particular, we tested for a learning effect across trials (i.e., whether children performed better in later trials as opposed to earlier trials), an order effect (i.e., whether children performed better if provided with an affirmative trial first as opposed to a negative trial first), and a container effect (i.e., whether children performed better in trials where the toy was hidden in the same container as before). An assess-

Table 2
Parameter estimates for correct searches in test trials in Study 1.

Effects	Estimate	*SE*	*z*
Intercept	1.01	0.29	3.46***
Trial type (affirmative vs. all negative)	−0.70	0.43	−1.64
Trial type (negative vs. three-location)	−0.29	0.50	−0.59
Age (younger vs. older)	0.61	0.50	1.23
Trial Type (affirmative vs. all negative): Age (younger vs. older)	1.39	0.68	2.04*
Trial Type (negative vs. three-location): Age (younger vs. older)	−0.14	0.75	−0.19

* $p < 0.05$.
*** $p < 0.001$.

ment of model fit based on chi-square tests of the change in −2 restricted log likelihood showed that including these factors did not significantly improve model fit (model with Learning: $\chi^2(6) = 8.09$, $p = 0.23$; model with Order: $\chi^2(6) = 7.04$, $p = 0.32$; model with Container $\chi^2(6) = 3.58$, $p = 0.73$).

Binomial tests of robust performance in negative trials

Next, we compared the two age groups on the number of children who specifically achieved scores of 2 of 2 on the negative type trials (N and 3L) to get a clearer picture of robust performance on the task (see Table 3). We examined the number of children who scored 2 of 2 on the negative trials only because these were our test trials of interest. On negative trials, 8 of 25 (32%) of the younger 2-year-olds and 12 of 21 (54.1%) of the older 2-year-olds obtained perfect scores. Binomial tests (two-tailed) showed that the proportion of children who scored 2 of 2 correct on the negative trials was significantly above chance (0.25) in the older group ($p = 0.002$), but not in the younger group ($p = 0.49$). Similarly, on the three-location trials, the proportion of children who scored 2 of 2 correct was significantly higher than expected by chance for the older 2-year-olds ($p = 0.002$), but not for the younger 2-year-olds ($p = 0.25$).

Discussion

The goal of Study 1 was to examine 2-year-olds' use of affirmative and negative sentences in a search by elimination task in the absence of visual cues. We found that younger 2-year-olds performed significantly better when searching based on affirmative sentences compared with negative sentences, whereas older 2-year-olds performed the same for both types of sentences. Thus, when younger children in our sample needed to rely on their understanding of propositional negation (on negative and three-location trials) to infer the location of a toy, they were less likely to search correctly compared with when hearing affirmative sentences. This result is in accordance with our predictions and findings by Feiman et al. (2017), who also found that 20- to 24-month-olds performed significantly better when hearing affirmative sentences compared with negative sentences, but it is in contrast to findings by Austin et al. (2014), who found better performance in 24-month-olds with negative sentences.

Regarding children's performance on the trials involving verbal negation, older 2-year-olds performed better than younger 2-year-olds on both types of negative trials (N and 3L). Furthermore, the proportion of children achieving perfect scores on negative trials (N and 3L) was higher than what was expected by chance only in the older 2-year-olds. This set of findings is consistent with our hypotheses based on prior research, reporting robust comprehension of negation by 27 months of age, and indicates that children's ability to make inferences when information is given using negation (e.g., "The ball is not in the bucket") undergoes significant development during the second year of life (Austin et al., 2014; Feiman et al., 2017).

Our more complex three-location trials tested whether children can integrate visual and verbal negation to make an inference about the location of a toy (children were shown that one location was empty and then were told that another location did not contain the toy). The finding that children in both age groups performed similarly in negative and three-location trials suggests that these more

Table 3
Correct responses out of two by trial type and age in Study 1.

	0/2	%	1/2	%	2/2	%	*N*
Younger 2-year-olds (M_{age} = 24.21 months)							
Control	0	0	5	20	20	80	25
Affirmative	1	4	9	36	15	60	25
Negative	7	28	10	40	8	32	25
Three-location	6	24	10	40	9	42.9	25
Older 2-year-olds (M_{age} = 35.39 months)							
Control	1	4.76	0	0.0	20	95.2	21
Affirmative	4	19	4	19	13	61.9	21
Negative	4	19	5	23.8	12	57.1	21
Three-location	2	9.5	7	33.3	12	57.1	21

complex negative trials did not impose additional cognitive demands on children, in contrast to our hypothesis. However, it should be noted that only older 2-year-olds performed significantly better than chance on three-location trials, thereby suggesting that it was only by the end of their second year that children were able to hold in mind different propositional arguments (''The toy is not in A" and "The toy is not in B") and successfully infer the conclusion based on their logical combination (''Therefore, the toy must be in C") across visual and verbal modalities.

Were children in our task engaging in deductive reasoning, or were they simply eliminating options? Our findings on negative trials (involving search in two locations) are compatible with both explanations because they show that the majority of older 2-year-olds successfully comprehended the negative verbal cue and searched in the alternative location, but they do not provide evidence as to whether children represented the dependent relation between the two possible locations (i.e., ''The toy is in A OR in B") and interpreted verbal negation as truth functional (i.e., ''The toy is NOT in A") in order to make inferences (i.e., ''THEREFORE, the toy is in B"). However, older 2-year-olds' success with three-location trials provides suggestive evidence that these children were not merely avoiding empty locations but were effectively using the cross-modal information provided to them to update their belief that the toy *necessarily* needed to be located in the third container. Study 2 directly tested older 2-year-olds' ability to apply their knowledge of verbal negation in a more complex search task that required them to reason with the disjunctive syllogism.

Study 2

Study 2 was a variation of Mody and Carey (2016) four-cups paradigm, where children were required to use a negative statement about the location of a toy instead of being shown visually that one cup is empty. The goal of the task was for children to find as many coins as possible in order to win a prize at the end of the game. Similar to Mody and Carey's study, our second study included training and test trials. In training trials, children were presented with three different-colored cups, with a pair of cups being placed on one side of the table and a single cup being placed on the other side of the table (see Fig. 2). Children watched as two coins were hidden (behind an opaque barrier) in the single cup and in the pair of cups. Children needed to then make a choice between the single cup on one side, which certainly held a coin, and the two uncertain cups on the other side. If children are sensitive to certainty versus uncertainty, they were expected to search in the single cup consistently, ignoring the uncertain cups in the pair. However, if toddlers search among the three cups randomly, we might conclude that the toddlers are unable to distinguish between the certainty that comes with having only one option and the uncertainty of having more than one option with no further information.

Test trials were similar to training trials but included four cups placed in pairs on either side of the table. As before, two coins were hidden inside the cups (one in each pair). Unlike Mody and Carey (2016), who demonstrated visually that one of the cups was empty, we used a negative sentence to convey the premise ''NOT A" (i.e., ''There is no coin in the red cup"). If children comprehend the negative sentence and engage in the disjunctive syllogism, they should represent the truth-functional negative argument (NOT A) and combine it with their belief about the location of the object (A OR B) in order to deduce which location contains the toy (THEREFORE B). Furthermore, if children engage in certainty monitoring, they are expected to reason that the two pairs of cups are independent from each other (whereas the individual cups are dependent: A OR B, C OR D), and after learning that the coin is not in Cup A, they are expected to consider that Cup B is more certain to contain the coin compared with the other two (dependent) locations. If, however, children are using the ''maybe A, maybe B" strategy, they may represent the negative argument (as either truth-functional or nonexistential), but they would have no representation of the dependent relation between the cups in each pair and, thus, would consider all the remaining alternatives as equally likely to contain a coin. If children are using the ''avoid empty" strategy, they should interpret the negative cue as non-existential (e.g., ''A is empty") and choose one of the remaining cups at random. Notice that ''maybe A, maybe B" and ''avoid empty" do not make distinct predictions in terms of children's behavior but may differ in terms of how negation is represented. Finally, because our task involves verbal negation, if children do not comprehend the negative cue, they may completely disregard negation and simply choose the named container.

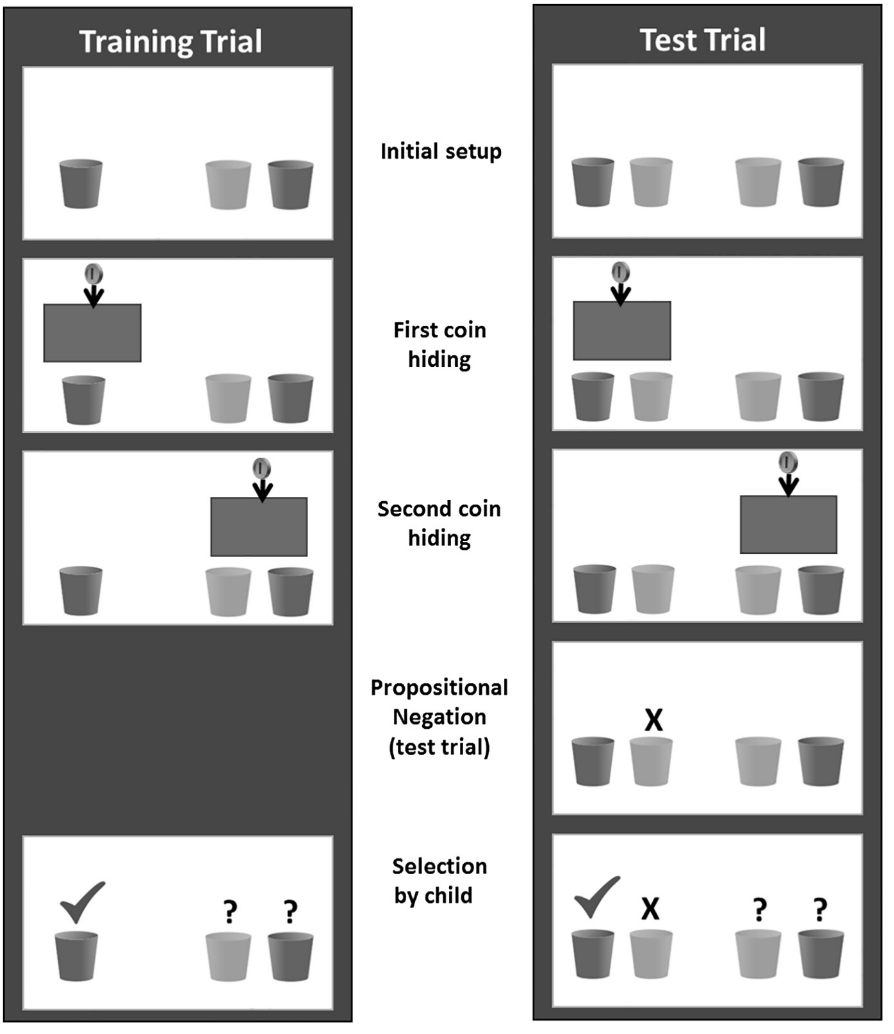

Fig. 2. Structure of training trials (left) and test trials (right) in Study 2. Children watch a coin hidden behind a screen for each side of cups. Children hear a negative sentence for the cup with the cross (e.g., "There is no coin in the blue cup"). The cup with the checkmark is certain to contain a coin, and the cups with question marks may or may not contain a coin.

We chose our group of 2-year-olds in this more complex search task based on the age when children are shown to pass the simpler two-cups verbal task from Study 1. Given that Feiman et al. (2017) found that toddlers understood propositional negation at 27 months of age, and we found that toddlers in our older age group showed robust understanding of verbal negation at a mean age of 35 months, these two ages served as the upper and lower bounds of our first age group of 2-year-olds. These were compared with a group of 3-year-olds. Based on prior research (Mody & Carey, 2016), we expected that 3-year-olds would perform better than 2.5-year-olds in this more complex search task. We also expected that children might perform worse in test trials, as opposed to training trials, because of the added demands of reasoning over uncertainty and needing to process linguistic information in order to draw inferences.

Method

Participants

All children included in the sample were recruited through and tested in the same research lab as in Study 1, and were of comparable socioeconomic status. The final sample was composed of 51 children: 24 2.5-year-olds (M_{age} = 32.3 months, range = 27.3–35.6; 10 girls) and 27 3-year-olds (M_{age} = 41.5 months, range = 36.0–47.5; 16 girls). An additional 3 3-year-olds were excluded from analysis due to failure on pretest involving color labeling (*n* = 1), fussiness or failure to follow/complete task instructions (*n* = 1), or experimenter error (*n* = 1). An additional 10 2.5-year-olds were excluded from analysis due to failure on color labeling (*n* = 6), or fussiness or failure to follow/complete task instructions (*n* = 4).

Materials

Four paper cups (green, blue, yellow, and red) were used as hiding locations. In each trial, two plastic golden coins were used as targets to be retrieved, and these coins were later exchanged by children for a prize. A white screen with a standing leg was used to obscure the cups from view as coins were lowered into a target cup. The experimenter sat across the table from children.

Procedure

Children were told that they would be playing a game in which the goal was to collect as many golden coins as possible, which they could later exchange for a prize. The rules of the game were that children were not allowed at any point to look inside the cups but needed to listen to clues and indicate which cup they thought contained a coin. The minimum number of coins required for a prize was not explicitly stated in order to tailor this information according to individual performance and to ensure that children were not unduly stressed but rather sufficiently motivated to try their best.

Pretest. Color swatches including the four target colors (red, blue, yellow, and green) and two distractor colors (purple and pink) were presented in an array to the child. Subsequent data were analyzed only from children who were able to correctly identify all four target colors by correctly pointing to the correct color swatch or, alternatively, spontaneously offering the name of the color. This step was taken to ensure that children's failures in our verbal search task were not due to inability to establish reference for the appropriate cup in the verbal cues provided in test trials (e.g., "There is no coin in the yellow cup").

Training trials. The three training trials were similar to Mody and Carey (2016) procedure; however, children were prompted to make the searches themselves instead of observing a second experimenter perform the search because the role of a second experimenter (E2) was eliminated from our paradigm. Three cups were used in training trials, with the unused color varying across the three trials. A pair of cups was placed on one side of the table, and a single cup was placed on the other side (see Fig. 2). The location of the single cup and the pair of cups was counterbalanced between left and right across trials. Following Mody and Carey's paradigm, the distance between each cup in a pair was approximately 4 inches and the distance between the sets of cups was approximately 12 inches. These distances were marked with pale tape to ensure reliable spacing across experiments.

As illustrated in Fig. 2, the white screen was placed over each set of cups in turn, as a coin was lowered from the center of the screen using both hands. The correct choice was always the single cup, as opposed to either cup in the pair, because the single cup was the only location that was certain to hold a coin. Prior to lowering the coin, the experimenter ensured that the child was paying attention and watching the coin. Once both coins were lowered, the screen was removed and children were then prompted to make a choice with the statement "Which color do you choose?" The training trials were fixed such that in the uncertain pair the experimenter made the appearance of lowering a coin behind the screen but, unbeknownst to the child, held onto the coin such that no coin was actually placed in either of the uncertain cups. This was done in order to avoid reinforcing an incorrect search toward an uncertain cup by allowing a child to obtain a coin from these searches.

Two demonstration trials were placed in between the training trials in the following order: demonstration, training, demonstration, training, training. The demonstration trial involved the same structure as the training trial except that the experimenter offered an explicit suggestion to the child, for example, "On this side, the coin could be in the red cup or the green cup, but you can't be sure. On this side, the coin has to be in the yellow cup, so you should choose the yellow cup." As in Mody and Carey (2016) paradigm, the demonstration trials were included to cue the children that one cup necessarily contained the coin, and a response based on certainty was required for both the training and test trials.

Two negative trials followed the training trials. In these trials, the child observed a coin being hidden in one of two cups and received a verbal cue, for instance, "There is no coin in the yellow cup." The goal of these trials was to familiarize children with receiving verbal cues to infer the location of the coin in a simple two-cups paradigm. As expected based on results from Study 1, both younger and older children had high success rates in these trials (.86 and .87, respectively).

Test trials. The four test trials were identical to the training trials except that four cups were used in each trial (see Fig. 2). The child observed one coin hidden in each pair behind the white screen, but just as in the training trials, a coin was lowered only in the target pair, unbeknownst to the child, in order to avoid reinforcing an incorrect choice. Thus, the child obtained a coin only for a correct choice and obtained no coin for any of the remaining three choices. Following the hiding phase, the experimenter gave the child a clue, always indicating one of the four cups as empty, for example, "There is no coin in the green cup." As in the training trials, the child needed to choose between searching for a coin either in the uncertain pair or in the remaining cup paired with the empty cup. The child received four test trials, with the location of the coin counterbalanced across cup colors and across left and right sides.

Results

Reliability coding

Two independent raters, blind to the study hypotheses, coded children's responses using video recordings of the sessions. To count a response as a search, the child was required to point clearly to one of the cups, label the color of the cup, or pick up the cup. Inter-rater reliability analyses using the kappa statistic were performed to determine consistency among raters on training and test trials. Reliability of coders was high for both training trials (kappa = 0.88, $p < 0.001$) and test trials (kappa = 0.93, $p < 0.001$). Discrepancies were examined and resolved by a third research assistant.

Training trials

To test for children's performance in training and test trials, we computed percentages of successful searches for each child (see Fig. 3). We followed this data-analytic strategy for ease of comparison between our study and Mody and Carey (2016). In training trials, the 2.5-year-olds selected the correct cup on 44.4% of trials and the 3-year-olds did so on 62.3% of trials. These results indicate that our training was similar to that used by Mody and Carey. In their study, the 2.5-year-olds scored correctly on 47% of trials and the 3-year-olds did so on 60% of trials. Using similar above-chance tests as Mody and Carey, the results showed that the 3-year-olds' performance was significantly greater than chance (0.33), $t(26) = 4.68$, $p < 0.001$, whereas the performance of 2.5-year-olds was at chance, $t(23) = 2.21$, $p = 0.037$. A Bonferroni correction for multiple comparisons was applied for an adjusted alpha of 0.025.

Test trials

In test trials, the 2.5-year-olds chose the target cup on 54.2% of trials and the 3-year-olds did so on 57.1% of trials. Comparisons with chance (0.33) showed that both 2.5-year-olds' and 3-year-olds' performance was significantly above chance [2.5-year-olds: $t(23) = 2.95$, $p = 0.007$; 3-year-olds: $t(26) = 3.72$, $p < 0.001$]. A Bonferroni correction for multiple comparisons was applied for an adjusted alpha of 0.025. Mody and Carey (2016), whose paradigm showed an empty cup instead of using propositional negation, reported average passing rates of 36% and 58% for 2.5- and 3-year-olds respectively.

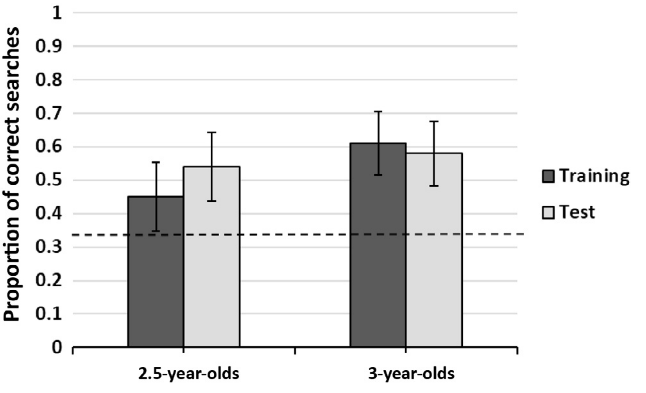

Fig. 3. Proportion of correct searches in training and test trials per age group in Study 2. Error bars represent standard errors, and the dotted line represents chance level (0.33).

Overall analysis

Whereas training trials looked only at children's ability to weigh certainty between two sets of cups (i.e., a single certain choice vs. an uncertain pair), the test trials also required reasoning by disjunction (i.e., between two pairs, one pair is certain and the other remains uncertain because an alternative has been eliminated). To test our predictions about children's performance in test and training trials and age differences, we analyzed their search accuracy (a binary dependent variable) across all trials. The data were analyzed using multilevel logistic mixed-effects modeling with crossed random intercepts for participants and items and a random slope of trial type. The model included fixed effects of trial type (training or test) as a first-level predictor, age (2.5- or 3-year-olds) as a second-level predictor, and their interaction. Both fixed effects of trial type and age were coded with centered contrasts (−.5, .5). Results showed no significant effects or interactions (see Table 4).

Learning effect

To explore the possibility that non-theoretically driven predictors affected children's performance, we examined whether a learning effect occurred in later trials as opposed to earlier trials for both training and test trials. Tables 5 and 6 show the breakdown for the number of children who passed on each training and test trial in terms of overall group as well as separate age groups. Beginning with training trials, to test for a learning effect, we used a mixed-effects logistic regression model with crossed random intercepts for participants and items and fixed effects of trial, age, and their interaction. The main effect of trial (Trials 1 and 2 or Trial 3) was analyzed with two simple contrasts comparing Trial 3 (i.e., the latest training trial) with the earliest two trials combined (c_1: −0.33, −0.33, 0.66) and comparing Trial 1 with Trial 2 (c_2: −0.5, 0.5, 0). Results showed a significant effect of trial, with children across age groups being more successful in their searches in Trial 3 as opposed to Trials 1 and 2 combined (β = 0.74, SE = 0.37, z = 1.98, p = 0.048); Trials 1 and 2 did not differ (p = 0.425). Furthermore, the analysis showed a marginally significant effect of age, with 3-year-olds tending to be more successful in their searches compared with 2.5-year-olds (β = 0.67, SE = 0.35, z = 1.91,

Table 4
Parameter estimates for correct searches across trials and age groups in Study 2.

Effects	Estimate	*SE*	*z*
Intercept	0.24	0.19	1.25
Trial type (training vs. test)	−0.22	0.37	−0.59
Age (2.5- vs. 3-year-olds)	0.42	0.31	1.37
Trial Type (training vs. test): Age (2- vs. 3-year-olds)	0.47	0.61	0.76

Table 5
Trial-by-trial passing frequencies and percentages for each training trial in Study 2.

	Trial 1	Trial 2	Trial 3
All children (*N* = 51)			
n	22	24	33
%	43.1	47.1	64.7
2-year-olds (*n* = 24)			
n	7	12	13
%	29.2	50	54.2
3-year-olds (*n* = 27)			
n	15	12	20
%	55.5	44.4	74.1

Table 6
Trial-by-trial passing frequencies and percentages for each test trial in Study 2.

	Trial 1	Trial 2	Trial 3	Trial 4
All children (*N* = 51)				
n	31	26	33	23
%	60.8	60	64.7	45
2-year-olds (*n* = 24)				
n	15	11	14	12
%	62.5	45.8	58.3	50
3-year-olds (*n* = 27)				
n	16	15	19	11
%	59.3	55.6	70.4	40.7

$p = 0.057$). A similar analysis was conducted for test trials. Because there were four test trials, the main effect of trial was analyzed with centered contrasts (−0.5, 0.5) between earlier trials (Trials 1 and 2) and later trials (Trials 3 and 4). Results showed that there was no effect of learning in any age group and no interaction between them (all *p*s > 0.05).

Discussion

The goal of Study 2 was to examine children's ability to use the disjunctive syllogism in a verbal task using propositional negation. We used a task modeled after Mody and Carey (2016) four-cups disjunctive syllogism paradigm using verbal information about the absence of a toy. In our study, the information was given to the child verbally ("There is no coin in the blue cup"). The main finding of this study was that, in test trials that required children to reason with the disjunctive syllogism, both 3- and 2.5-year-olds chose the target cup significantly above chance. Specifically, 3-year-olds chose the cup that was certain to contain a coin 57.1% of the time and 2.5-year-olds did so 54.2% of the time. Importantly, as the absence of a learning effect indicated, this tendency was consistent across trials, with both age groups being more likely to choose the target cup from the very first trial. Furthermore, contrary to our predictions, the two age groups did not differ in the likelihood of choosing the target. Thus, beginning at about 2.5 years of age, children in our sample understood negation and used it to make a choice between an option that definitely contained a prize and a less certain option (one that might have contained a prize), suggesting that they effectively engaged in disjunctive syllogism. This finding is strikingly different from the results of Mody and Carey, who found evidence for disjunctive reasoning only in children older than 3 years (who chose the target cup 58% of the time) but not in 2.5-year-olds (who chose the target cup 36% of the time).

A comparison between test and training trials showed that children's performance in the two types of trials did not differ, but comparisons with chance showed that, in training trials, only 3-year-olds, but not 2.5-year-olds, chose the target cup significantly above chance. In particular, 3-year-olds chose the target 62.3% of the time and 2.5-year-olds did so 44.4% of the time. At first glance, this finding seems puzzling given 2.5-year-olds' success in the more complex test trials. However, children's performance was not uniform across training trials, with successful search rates being higher in the last training trial compared with the earlier two trials. Therefore, 2.5-year-olds' overall performance in training trials may have been affected by some early failures before they were familiarized with the task. Comparing our overall passing rates with those of Mody and Carey (2016), we found that these were nearly identical for 3-year-olds (61%) and were comparable for 2.5-year-olds (47%).

Despite 2.5-year-olds' success with the disjunctive syllogism in our verbal version of the four-cups task, children's performance was not at ceiling, given that even 3-year-olds demonstrated errors in their searches. Thus, although children showed signs of highly abstract reasoning before age 3, and earlier than previously assumed, this ability continues to develop during the preschool years. Interestingly, our task showed an additional type of search error that was not present in Mody and Carey (2016) nonlinguistic paradigm. Although in their study children's searches toward a cup that was shown to be empty were about 1%, in our study this rate was 8.8%. This suggests that children were not as accurate in eliminating search options based on verbal negation as they were based on visual negation. This finding is consistent with a strategy whereby children did not understand negation and simply searched in the named container. Interestingly, the language demands of our task presented additional challenges to some children, but not enough to affect the overall performance of the group on the disjunctive syllogism task.

General discussion

The current research examined young children's understanding of propositional negation and their ability to reason through elimination and disjunction. Study 1 showed that, by the end of their second year of life, children can understand propositional negation and, in the absence of visual cues, use this information to reason by elimination. Study 2 examined whether children could apply their understanding of propositional negation in a task requiring disjunctive reasoning, a complex logical inference of the type "A OR B, NOT A; THEREFORE B," where multiple possibilities are represented at the same time. The findings showed that both 2.5- and 3-year-olds can make disjunctive inferences when hearing a sentence involving propositional negation (e.g., "There is no coin in the green cup"), unlike prior nonlinguistic search tasks that demonstrated this ability only in 3-year-olds but not in younger children (Mody & Carey, 2016; cf. Call & Carpenter, 2001).

A first important aspect of our findings concerns the developmental trajectory of children's comprehension of propositional negation. Our data from Study 1 demonstrated that older 2-year-olds, but not younger 2-year-olds, showed robust understanding of negation in a reasoning by elimination task, in accordance with prior findings (Austin et al., 2014; Feiman et al., 2017). Furthermore, our results showed that children across age groups performed significantly better when searching with affirmative sentences compared with negative sentences, consistent with the finding that understanding of affirmation arrives prior to understanding of negation (Feiman et al., 2017). At first glance, this finding seems to be at odds with Austin et al. (2014), who found better comprehension of negatives than of affirmatives. However, a closer inspection of the success rates across studies reveals more similarities than differences; our younger 2-year-olds succeeded in affirmative trials 78% of the time and older 2-year-olds did so 72% of the time, whereas the respective rates in Austin et al. (2014) study were 74% and 78% and in Feiman et al. (2017) study were 78% and 78% (see also Feiman et al., 2017, p. 441, for a discussion).

A second important aspect of our findings concerns children's ability to use verbal negation in logical reasoning. Older 2-year-olds in our studies were able to use negation to infer the location of a toy not only in simple reasoning by elimination (i.e., negative trials in Study 1) but also in more complex reasoning that involved the combination of verbal and visual information (i.e., three-location trials in Study 1) as well as in disjunctive syllogism (e.g., "There is no coin in the green cup" in Study 2). Together, these findings suggest that 2.5-year-olds have a highly abstract understanding of negation that can be combined with other abstract arguments to reach logical conclusions. Importantly, this abstract understanding of negation needed to be truth-functional in nature; although children could have succeeded in simple two-cups reasoning by elimination tasks by interpreting negation as nonexistential (i.e., as reflecting the positive thought "A is empty"), success in the disjunctive syllogism in the more complex four-cups paradigm required truth-functional understanding of negation (i.e., the representation of the negative premise "NOT A"). Therefore, our results suggest that 2.5-year-olds' success in verbal reasoning by elimination in both our Study 1 and prior research (Austin et al., 2014; Feiman et al., 2017) could have relied on true comprehension of the truth-functional meaning of negation. This conclusion resonates with findings from children's language production showing that children produce truth-functional negation at about the same age as they are found to use negation in reasoning by elimination tasks (e.g., Choi, 1988; Hummer et al., 1993; Pea, 1980, 1982; cf. Feiman et al., 2017, for a discussion). Although our conclusions are constrained by the fact that they are based on children's performance in search tasks that only implicitly measure language comprehension, our study offers more conclusive evidence with respect to the nature of children's linguistic representations of negation compared with prior research.

A final point worth discussing is the finding that 2.5-year-olds in our verbal version of Mody and Carey (2016) four-cups task showed signs of engaging in disjunctive syllogism, whereas in the nonlinguistic version of the task only children older than 3 years succeeded. Several possibilities can explain this discrepancy. First, it is conceivable that the discrepancy is due to sampling differences between our study and Mody and Carey (2016). For instance, our study had a higher exclusion rate than Mody and Carey's study because we excluded children who did not know the labels for colors in order to ensure comprehension of the linguistic cues. This may have resulted in our sample being cognitively more advanced. Although this possibility cannot be excluded, it seems unlikely given that the results from the training phase (which was similar across studies) revealed very similar success rates between the two studies and, if anything, slightly better performance in Mody and Carey's 2.5-year-olds. Thus, it does not seem to be the case that 2.5-year-olds in our study were exceptionally good at logical reasoning. Second, it is possible that children may have been motivated differently in the two studies. For instance, children may have found collecting tokens to exchange for a prize, as in our paradigm, more motivating than winning against a competitor who always seemed to lose, as in Mody and Carey's paradigm. However, this explanation faces the problem that only 2.5-year-olds, but not 3-year-olds, showed better performance in our task and did so only in test trials, but not in training trials. Thus, the motivating effect was not evident across the board.

A third, more likely possibility is that the presence of verbal negation may have facilitated children's logical inferences. It is possible that offering children a negative proposition verbally gave them more direct access to the relevant premise "NOT A" compared with when the same premise needed to be constructed from visual evidence of "emptiness." Thus, this "head start" in the syllogism may have facilitated the subsequent steps of the inferential process. Importantly, an additional modification in our paradigm may have also contributed to the greater success rates in 2.5-year-olds. Before test trials, children were presented with two negative trials where they were offered a cue in the form of verbal negation, similar to test trials, but they were asked to reason by exclusion between only two alternatives. Both 2.5- and 3-year-olds' performance in these trials was particularly high (86% and 87%, respectively). These trials may have helped our participants, especially the youngest ones, realize how the information provided by the negative cue could be combined with thoughts about the dependent relation of the paired alternatives to infer the location of the prize. Future research could test this possibility empirically by "training" children in reasoning by exclusion using verbal negation and then having them complete the nonlinguistic version of the four-cups task.

How exactly did the presence of language facilitate 2.5-year-olds' reasoning about the location of a reward? So far, following Mody and Carey (2016), we have assumed that success in the four-cups task required children to use deductive reasoning, whereby participants are absolutely certain of both the premises of the syllogism and the logical conclusion (i.e., that the reward certainly needed to be in A or B, and because it was certainly not in A, it necessarily needed to be in B). However, as Mody and Carey pointed out, there is no clear evidence that children in the four-cups task did not engage in *probabilistic* reasoning (p. 46), a type of reasoning whereby participants are never entirely certain of either the premises or the conclusion but engage in Bayesian redistribution of coupled probabilities (i.e., as the probability of A decreases, the probability of B increases; Rescorla, 2009; see also Srećković, 2018, for a detailed discussion of the two reasoning mechanisms). Although we cannot know how certain children were of their propositions, our results may suggest that children's reasoning was not based on absolute certainty; the 2.5-year-olds may have succeeded in our verbal version of the four-cups task because the linguistic evidence based on which they needed to construct the premise "NOT A" was considered a highly reliable source of evidence that increased their certainty about the premise and subsequently the conclusion. By contrast, 2.5-year-olds' failures in the nonlinguistic version of the task may have been due to the fact that the visual information based on which they needed to construct the same premise was somewhat uncertain,[1] leading to incorrect predictions about the location of the reward. This possibility is supported by research demonstrating that young children are more likely to trust information about the location of an object when this is conveyed verbally compared with when the same information is conveyed visually, especially if the verbal information is provided by a visible engaging speaker, as was our experimenter in Study 2 (see Jaswal, Croft, Setia, & Cole, 2010; Ma & Ganea, 2010).

Currently, it is difficult to adjudicate between the deductive and probabilistic reasoning mechanisms because their details are theoretically underdeveloped (Srećković, 2018). Future theoretical research could work out exactly what types of representations and logical structures these two mechanisms involve in order to develop specific predictions for each one that could be tested empirically. In particular, it would be interesting to see whether logical reasoning, either deductive or probabilistic, is language-based or defined over nonlinguistic representations. Such theoretical advancements would be particularly pertinent for developmental debates on the ontogeny of logical thought and its relation to language.

Although we still lack a convincing theory about the nature of logical representations, the development of syllogistic reasoning in children and the relation between language and logical thought, our study provides evidence for the presence of logical reasoning in very young children and is in accordance with recent work showing that prelinguistic infants already have the ability to use some elementary form of the disjunctive syllogism (Cesana-Arlotti et al., 2018). Given the evidence that children are capable of logical reasoning already at 12 months of age, and their behavioral (i.e., oculomotor) responses when they (presumably) draw logical deductions resemble those of adults (Cesana-Arlotti et al., 2018), it seems rather unlikely that children's ability to reason with the disjunctive syllogism passes through a stage where children use simple heuristics (such as "avoid empty") that gradually develop into a fully deductive mental process. Instead, it seems more likely that children simply fall back to simple (task-dependent) heuristics in situations where solving inferential problems is cognitively demanding (e.g., in situations where children are asked to simultaneously represent multiple alternatives and reason over uncertainty, as in Mody & Carey, 2016). This type of processing difficulty may explain why, despite their success in our task, 2.5- and 3-year-olds' performance was not perfect, with approximately one third of our participants consistently failing to use the disjunctive syllogism. Certainly, children's ability to use logical reasoning undergoes significant change over development, well beyond the preschool years (Mody & Carey, 2016), in order to reach an adult level of maturity. Exploring the mechanisms that underlie this change is an exciting avenue for future research.

Acknowledgments

This work was supported by funds from Natural Sciences and Engineering Research Council of Canada (2016-05603) awarded to P. A. Ganea. We thank members of the Language and Learning Lab for their contributions and the families and children who participated in this research. We also thank Angela Nyhout and Brian Leahy for their comments on a previous draft of the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2019.02.004.

Supplementary data 1 The supplementary data to this article is available on the Open Science Framework: https://osf.io/n5sah/?view_only=9cb866158cf54f59b8055b57e7196e85.

[1] The uncertainty of knowledge based on perceptual (visual) evidence may be due to difficulties in drawing a link between *seeing* and *knowing*, which has been used to explain why 2.5-year-old children and apes do not show signs of inferential updating of their beliefs about the location of a reward (Call & Carpenter, 2001).

References

Austin, K., Theakston, A., Lieven, E., & Tomasello, M. (2014). Young children's understanding of denial. *Developmental Psychology, 50*, 2061–2070.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language, 7*, 235–261.

Bräuer, J., Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2006). Making inferences about the location of hidden food: Social dog, causal ape. *Journal of Comparative Psychology, 120*, 38–47.

Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla,* and *Pongo pygmaeus*). *Journal of Comparative Psychology, 118*, 232–241.

Call, J. (2006). Inferences by exclusion in the great apes: The effect of age and species. *Animal Cognition, 9*, 393–403.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 3*, 207–220.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science, 359*, 1263–1266.

Choi, S. (1988). The semantic development of negation: A cross-linguistic longitudinal study. *Journal of Child Language, 15*, 517–531.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Clark, H. H., & Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory & Cognition, 2*, 101–111.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*, 125–127.

Feiman, R., Mody, S., Sanborn, S., & Carey, S. (2017). What do you mean, no? Toddlers' comprehension of logical "no" and "not". *Language Learning and Development, 13*, 430–450.

Fodor, J. D. (1970). Formal linguistics and formal logic. In W. J. Savitch, E. Bach, W. E. Marsh, & G. Safran-Naveh (Eds.), *The formal complexity of natural language* (pp. 24–40). Amsterdam: Springer.

Fraundorf, S. H., Benjamin, A. S., & Watson, D. G. (2013). What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language, 69*, 196–227.

French, L. A., & Nelson, K. (1985). *Young children's knowledge of relational terms: Some ifs, ors, and buts*. New York: Springer-Verlag.

Hill, A., Collier-Baker, E., & Suddendorf, T. (2012). Inferential reasoning by exclusion in children (*Homo sapiens*). *Journal of Comparative Psychology, 126*, 243–254.

Hummer, P., Wimmer, H., & Antes, G. (1993). On the origins of denial negation. *Journal of Child Language, 20*, 607–618.

Jackendoff, R. (1996). How language helps us think. *Pragmatics and Cognition, 4*, 1–34.

Jackendoff, R. (2011). What is the human language faculty? Two views. *Language, 87*, 586–624.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21*, 1541–1547.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38*, 1033–1050.

Kim, K. (1985). Development of the concept of truth-functional negation. *Developmental Psychology, 21*, 462–472.

Lust, B., & Mervis, C. A. (1980). Development of coordination in the natural speech of young children. *Journal of Child Language, 7*, 279–304.

Ma, L., & Ganea, P. A. (2010). Dealing with conflicting information: Young children's reliance on what they see versus what they are told. *Developmental Science, 13*, 151–160.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*, 367–380.

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

Morris, B. J. (2008). Logically speaking: Evidence for item-based acquisition of the connectives AND & OR. *Journal of Cognition and Development, 9*, 67–88.

Nordmeyer, A. E., & Frank, M. C. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language, 77*, 25–39.

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language, 74*, 36–45.

Pea, R. D. (1980). The development of negation in early child language. In D. R. Olson (Ed.), *The social foundations of language and thought: Essays in honor of Jerome S. Bruner* (pp. 156–186). New York: Norton.

Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two- and three-year-olds. *Journal of Child Language, 9*, 597–626.

Pepperberg, I. M., Koepke, A., Livingston, P., Girard, M., & Hartsfield, L. A. (2013). Reasoning by inference: Further studies on exclusion in grey parrots (*Psittacus erithacus*). *Journal of Comparative Psychology, 127*, 272–281.

Premack, D., & Premack, A. J. (1994). Levels of causal understanding in chimpanzees and children. *Cognition, 50*, 347–362.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rescorla, M. (2009). Chrysippus' dog as a case study in non-linguistic cognition. In R. W. Lurz (Ed.), *The philosophy of animal minds* (pp. 52–71). Cambridge, UK: Cambridge University Press.
Reuter, T., Feiman, R., & Snedeker, J. (2018). Getting to no: Pragmatic and semantic factors in two- and three-year-olds' understanding of negation. *Child Development, 89*, e364–e381.
Schloegl, C., Dierks, A., Gajdon, G. K., Huber, L., Kotrschal, K., & Bugnyar, T. (2009). What you see is what you get? Exclusion performances in ravens and keas. *PLoS One, 4*(8). e6368.
Srećković, S. (2018). Reasoning of non- and pre-linguistic creatures: How much do the experiments tell us? *Belgrade Philosophical Annual, 31*, 115–126.
Watson, J. S., Gergely, G., Csanyi, V., Topal, J., Gacsi, M., & Sarkozi, Z. (2001). Distinguishing logic from association in the solution of an invisible displacement task by children (*Homo sapiens*) and dogs (*Canis familiaris*): Using negation of disjunction. *Journal of Comparative Psychology, 115*, 219–252.