# Children spontaneously compute scalar implicatures

Ebru Evcen[1] and David Barner[2]

[1]Department of Linguistics, UC San Diego

[2]Department of Psychology, UC San Diego

**Corresponding author:**

Ebru Evcen <eevcen@ucsd.edu>

Department of Linguistics, UC San Diego

9500 Gilman Drive, La Jolla, CA, 92092

**Abstract**

A large developmental literature finds that children often fail to compute scalar implicatures, accepting underinformative descriptions such as, “Some of the horses jumped” when all of them did, or, “The bear took a cup or a plate” when he took both. Here, we argue that past studies find such failures because they do not evaluate children’s spontaneous interpretation of utterances, but instead ask how they interpret language when provided proffered “worlds” that children would not consider otherwise. We show that when children are allowed the chance to act upon their own spontaneous interpretation of utterances, they compute scalar implicatures in an adult-like fashion. Specifically, we tested children in two conditions: a standard Truth Value Judgment (TVJ) condition, similar to past studies, and a novel, two-phase Act-Out+TVJ condition in which children first acted on the utterance by coloring pictures before judging a puppet's action. In the act-out task children almost never colored both disjuncts after hearing “or” (Experiment 1) and rarely colored all four objects after hearing “some” (Experiment 2). More importantly, after acting out these interpretations, children computed scalar implicatures in an adult-like fashion, and significantly more often than children in the TVJ-only condition. These findings suggest that preschoolers spontaneously access relevant alternative utterances and compute scalar implicatures when allowed to generate their own interpretations before evaluating experimenter-provided worlds, and that previous studies underestimate this ability by artificially proffering irrelevant worlds.

**Keywords:** inferential reasoning**,** scalar implicatures, pragmatic development, access to alternatives, question under discussion

Speakers often intend to communicate information that goes beyond the literal meanings of the utterances that they use. For example, upon hearing a sentence like, “Lucy fed some of the horses in the field,” adult listeners typically infer that Lucy did not feed all of them, even though the literal meaning of “some” is compatible with feeding every horse in the field. Similarly, upon hearing a sentence like, “Lucy fed the cow or the horse,” adult listeners typically derive an exclusive interpretation, inferring that Lucy fed one of the two animals, but not both, although the literal meaning of disjunction is compatible with a situation in which Lucy fed both animals. In both cases, listeners assume that a knowledgeable speaker would have used a stronger alternative, “Lucy fed all of the horses” or “Lucy fed the cow and the horse,” if either were true. These inferences, called scalar implicatures, pose an interesting problem, because they require children to infer the speaker’s intended meaning by reasoning about statements that were not actually uttered, to rule out alternative possible meanings and corresponding possible worlds.

Interestingly, a large developmental literature finds that children struggle to compute such pragmatic inferences well into their elementary school years. Often, these studies test participants using Truth Value Judgment or Felicity Judgment tasks, in which children are presented with a scene and an underinformative utterance, and are asked to judge if the utterance is a good description. For example, in one study Papafragou and Musolino (2003) presented children with a scene in which three horses jumped over a fence, and children were asked to evaluate the felicity of the description, “Some of the horses jumped over the fence.” They found that, unlike adults, who generally rejected such statements, children almost always accepted them (see also Gualmini et al., 2001; Noveck, 2001; Guasti et al., 2005; Barner et al., 2011; Huang & Snedeker, 2009; Foppolo et al., 2012; Papafragou & Tantalou, 2004; Noveck et al.,

2007). Similarly, studies of disjunction report that children accept underinformative disjunctive statements like, “The bear took a cup or a plate,” when both disjuncts are true, whereas adults generally do not (e.g., Chierchia et al., 2001; Hochstein et al., 2016; Skordos et al., 2020; Huang & Crain, 2020; Jasbi & Frank, 2021).

Multiple accounts of children’s failure to compute implicatures have been proposed, pointing to limited pragmatic reasoning abilities (Noveck, 2001), processing limits (Chierchia et al., 2001; Reinhart, 2006; Pouscoulous et al., 2007; Huang & Snedeker, 2009), difficulties understanding the communicative goals of speakers (Papafragou & Musolino 2003; Papafragou & Tantalou, 2004), a generalized tolerance for pragmatic infelicity on the part of children (Katsos & Bishop, 2011), incomplete knowledge of quantifier meanings (Horowitz et al., 2018). These accounts also differ in whether they treat implicatures as the product of reasoning about a speaker’s intentions or as derived in the grammar (Chierchia et al., 2012), by an operator that negates stronger alternatives. On all views, however, implicatures depend on which alternatives are considered relevant in context (Fox & Katzir, 2011). Related to this, many studies have argued that children may fail to compute implicatures because they struggle to access relevant alternative utterances, for example failing to spontaneously consider a statement containing “and” as an alternative to a disjunctive utterance, either because scales such as ⟨some, all⟩ are not yet lexicalized (Foppolo et al., 2012; Foppolo et al., 2021) or because alternatives are not recognized as relevant and retrieved in context (Barner & Bachrach, 2010; Barner et al., 2011; Barner et al., 2018; Bleotu et al., 2025; Chierchia et al., 2001; Gotzner et al., 2020; Horowitz et al., 2018; Pagliarini et al., 2018; Stiller et al., 2015; Tieu et al., 2016, 2017; Skordos & Papafragou, 2016). As evidence for this, children readily compute *ad hoc* implicatures, in which alternatives are contextually salient (Barner et al., 2011; Barner et al., 2018; Papafragou &

Tantalou, 2004; Rees et al., 2023; Yoon & Frank, 2019), including in picture-choice paradigms, where alternatives can be generated from visible stimuli (Foppolo et al., 2021; Kampa & Papafragou, 2020; Horowitz et al., 2018; Stiller et al., 2015), or in felicity and truth value tasks when the preceding context includes mention of relevant alternative statements (Skordos & Papafragou, 2016). Crucially, however, although children often compute adult-like inferences when relevant alternatives are contextually salient , no study has found comparable success in truth-value judgement tasks when relevant alternatives are not provided[1]. Given this, a natural explanation of past findings is that children fail to compute implicatures because they lack the ability to spontaneously generate relevant alternatives.

While the access-to-alternatives hypothesis provides a natural explanation of why children often fail to compute implicatures in felicity and truth value judgment tasks, and why contextual cues to alternatives help, these findings are also compatible with a somewhat opposite analysis. Specifically, rather than being neutral with respect to alternatives, judgment tasks may actually provide highly salient and highly misleading alternatives, which cause otherwise competent pragmatic reasoners to derive unintended inferences, and to ostensibly fail tests of implicature (see also Crain et al., 1996; Skordos & Papafragou, 2016). On this view, when children hear scalar expressions, they may spontaneously generate relevant alternatives and compute implicatures. However, when they are first shown a scene that proffers possible

[1] One possible exception is Katsos and Bishop (2011). In a binary judgment task, most children accepted underinformative statements, but a separate group of children rated such statements below optimal ones on a three-point scale. As Katsos and Bishop themselves note, however, these ratings need not reflect an implicature. When a speaker describes a situation the child has already seen, for example saying that a mouse picked up *some* of the carrots when it picked up all of them, the child can rate the statement lower simply because the speaker could have said *all*. This requires noticing that a stronger statement was available, but not inferring that it is false. Another possible exception is Papafragou and Musolino (2003), who trained children to judge whether statements were felicitous rather than merely true. With this support, children rejected underinformative *some* only about half the time (52.5%, up from 12.5%), compared with 92.5% among adults.

meanings that they would otherwise never consider – as is typical in truth value and felicity judgment tasks – this may impair children's ability to generate their own, spontaneous meanings, producing apparent failures.

In the present study, we explore this logic and present evidence that children may actually have adult-like abilities to compute implicatures when allowed to spontaneously generate their own meanings of utterances prior to a truth value judgment. In particular, we explore the possibility that children only fail in truth value judgment or felicity tasks when they are tested with scenes that proffer "worlds" that they would otherwise never consider when spontaneously interpreting language, that is, scenes that the experimenter presents before children have acted on their own interpretation of an utterance. In most past studies of implicature that find failures, including those described above, children are asked to evaluate the felicity of a statement given a scene that is created by the experimenter. For example, in a study of disjunction by Hochstein et al. (2016), children were asked to judge whether the statement, "The bear took a cup or a plate," was a good description of a proffered scene in which a bear had taken both objects (by choosing whether it was uttered by a silly puppet or a smart puppet). In order to ascribe this utterance to a "silly" speaker, participants were required to recognize a mismatch between the utterance and the proffered scene (see Gordon, 1996, for discussion). On this "proffered worlds" hypothesis, when an experimenter presents a scene that is incompatible with an implicature, this scene competes with the child's own preferred model of the utterance. This irrelevant scene not only interferes with the child's spontaneous meaning, but it also leads them to conclude that the experimenter must have an alternative "question" in mind, which is not compatible with an implicature.

The Proffered Worlds hypothesis takes inspiration from studies of quantifier spreading, where previous studies have also argued that proffering dispreferred worlds to children may

conceal adult-like competence. In this literature, multiple studies report that when preschool-aged children are presented with a scene in which a universally quantified expression like, “Every girl is riding an elephant,” is true, they nevertheless reject the utterance if, e.g., extra elephants are present (e.g., Aravind et al., 2017; Brooks & Sekerina, 2006; Philip, 1995; Skordos et al., 2022; Crain et al., 1996; Minai et al., 2012). On one analysis, provided by Crain et al. (1996), such cases may be pragmatically confusing to children both because the utterances are trivially true, but also because there are salient additional items that the children might assume are relevant, given that they have been proffered by the experimenter. According to Crain et al. (1996), to override these misleading contextual features, children require the ability to access “circumstances” in which it would be “appropriate to deny the test sentence” (p. 19). Compatible with this idea, Skordos et al. (2022) found that children readily accepted such statements if they were first presented a previous warm-up trial in which they saw 2 out of 3 girls riding elephants and were asked the identical question. In other words, when children were first tested with a scene that matched an adult-like understanding of the universally quantified sentence, they were no longer swayed by the presence of extra elephants in subsequent test trials.

How are these results from quantifier spreading relevant to understanding children’s difficulties with scalar implicature? On the analysis provided by both Crain et al. (1996) and Skordos et al. (2022), when experimenters proffer worlds that render utterances trivially true, uninformative, or otherwise vacuous, children assume that the speaker must have intended some other, more informative meaning that corresponds to salient aspects of the proffered scene. According to Crain et al. (1996) children in previous studies of quantifier spreading were “led to suppose that another interpretation was intended” because the “circumstances were inappropriate” on the adult interpretation (p. 116). In other words, when a misleading world is

proffered, children may infer an unintended "Question Under Discussion" or QUD (Hamblin, 1976; Karttunen, 1977; Roberts, 2012, because listeners infer the QUD from context, and a scene presented by the experimenter serves as evidence about what the speaker is asking. The notion of QUD is central to implicature, because formally questions can be thought of as determining "the set of propositions which are the possible answers … to the question; these are the proffered alternatives" (p. 65). Thus, if the QUD is modified contextually, the set of linguistic alternatives that is considered to be relevant should also change accordingly. Consequently, children may fail to access alternatives that are required for scalar implicature when misleading worlds are proffered, because these worlds trigger unintended QUDs that activate distinct and irrelevant alternatives. Our contention is that otherwise children may not have difficulty accessing relevant alternatives, particularly when they are allowed to spontaneously compute their own meanings for scalar expressions before any such scene is presented, as they typically do in natural conversation.

Lending support to this idea, studies using act-out paradigms have found that when children are asked to act on "some" of the objects in a scene, they almost never act on all of them, and that when they are asked to act on one of the objects with "or", they seldom act on both (e.g., Pouscoulous et al., 2007; Miller et al., 2005; Bleotu et al., 2025). For example, in one early study Miller et al. (2005) asked preschoolers to "Make SOME faces happy", by drawing smiling mouths on pictures of faces, and found that they only drew smiles on all faces around 10% of the time. Using a similar paradigm, Bleotu et al. (2025) presented Romanian-speaking 5-year-olds with a puppet who said, "I would like either the triangle or the circle to have color," and were asked to respond by coloring objects, erasing color, or taking no action. In a 0-Object scenario, in which neither object was already colored, most children colored exactly one object

(86%) – suggesting that they rarely generated mental representations in which both disjuncts were true. Similarly, in a 1-Object scenario, just over half of the responses involved doing nothing (52.2%), rather than coloring the second object. Finally – and relevant to the present hypothesis – children's performance was worst when they were initially proffered a scene in which both disjuncts were true. Here, in the 2-Object scenario, children only removed one of the two objects on less than half of trials (44.1%), unlike adults who almost always did so.

These results from act-out studies indicate that when children hear a statement like, "The bear took the cup or the plate," they are unlikely to spontaneously generate a mental representation in which both disjuncts are true. Crucially, however, such evidence falls short of demonstrating that children therefore compute scalar implicatures. This is because while these studies reveal the nature of the world that children most strongly associate with the meaning of an utterance, they do not test children's beliefs about other, stronger utterances, and whether they infer that the speaker believes such alternatives to be false. For example, in an act-out task involving disjunction, a child who colors a single shape in response to, "I would like to color the triangle or the square," produces an action that is compatible with an exclusive interpretation, but also with an inclusive one, since they might also believe that coloring both shapes would be appropriate (but that only one response is possible). In this sense, act-out tasks capture the world that children most strongly associate with an expression, but not whether they strengthen the utterance by negating more informative alternatives to reach that interpretation. A useful analogy comes from work on typicality effects: when participants are asked to "name a bird" they spontaneously generate prototypical exemplars like "robin" but rarely mention atypical instances like "penguin," yet readily judge that penguins are birds when asked (Rosch, 1973). In the same

way, act-out tasks reveal which worlds children most readily associate with a sentence, but not whether they would reject other worlds as incompatible with the utterance.

In the present study, we hypothesized that children may only fail on tests of implicature when first provided with proffered worlds, which preclude their ability to spontaneously generate their own models of utterances. Based on this, we reasoned that the contrary may also be true: that children may successfully compute implicatures when first allowed to generate their own models of utterances, because the activation of these preferred worlds precludes the adoption of proffered worlds. In particular, we reasoned that if children are first tested in an act-out task, which allows them to generate their own model of an utterance, they may subsequently compute implicatures for that utterance when tested with a felicity judgment task, because they have already converged upon their own preferred meaning prior to being proffered irrelevant scenes.

To test this idea, we adopted a two-step design in which each trial contained two phases. In the first phase, children completed an act-out task, which allowed them to interpret the utterance and construct their own representation of its meaning. In the second phase, children completed a standard felicity judgment task. They saw a puppet, Mr. Monkey, who was given the same instructions as the child, and were asked to judge whether Mr. Monkey carried out the instruction correctly. In the act-out phase, children received no feedback, heard no statement containing a stronger alternative, and saw only a page of uncolored objects that was compatible with any response. The judgment phase did not require a new interpretation of the utterance. Instead, it tested whether the interpretation children had already formed excluded worlds in which the stronger alternative was true. In Experiment 1, children were provided with a disjunctive statement (e.g., “Lucy colored the cow or the horse”) and were asked to color a new page to match this behavior. We measured whether children spontaneously colored one or both of

the mentioned items. In the second phase of the trial, children then completed a standard felicity judgment task to test whether they computed a scalar implicature. Specifically, they saw a puppet, Mr. Monkey, who was given the same instructions as the child and was invited to color the page to match the disjunctive utterance. Children were then asked to judge whether Mr. Monkey's coloring matched the statement. Crucially, the way in which Mr. Monkey colored depended on the child's initial response. If the child colored only one of the mentioned objects, Mr. Monkey either colored one object or both, allowing us to test whether children treated the statement as exclusive. However, if the child colored both objects, Mr. Monkey always colored just one object, allowing us to test whether children interpreted disjunction conjunctively (by preferring to color both items but rejecting the coloring of only one). We also included a TVJ-only condition in which children did not color based on the statement themselves, but proceeded directly to evaluating Mr. Monkey's coloring. In this condition, we expected children to give literal responses and accept both interpretations.

To broaden the empirical scope of the paper, we sought converging evidence for children's ability to compute scalar implicatures with "some" in Experiment 2. Following previous act-out studies (e.g., Miller et al., 2005), children heard a sentence containing the quantifier "some" (e.g., "Lucy colored some of the cows") and were asked to color a new page with four cows to match this behavior. This phase measured whether children spontaneously colored one, two, three, or all four of the available items. After completing the act-out phase, children saw Mr. Monkey attempt the same task. Mr. Monkey was given the same instruction and colored the page himself. Children were then asked to judge whether Mr. Monkey's coloring matched the statement. Crucially, the way Mr. Monkey colored depended on the child's initial response. If the child colored a subset of the cows, Mr. Monkey either colored three cows or all

four cows. This allowed us to test whether children computed the implicature associated with "some." If children interpreted "some" pragmatically as excluding "all", they should reject the case in which Mr. Monkey colored all four cows. However, if the child initially colored all four cows, Mr. Monkey always colored three cows. This allowed us to test whether children interpreted "some" literally. In this case, children who treated "some" as compatible with "all" should reject Mr. Monkey's coloring of three cows. As in Experiment 1, we included a TVJ-only condition in which children did not color based on the statement themselves, but did a truth value judgement task. In this condition, we expected children to give literal responses and accept Mr. Monkey's coloring regardless of whether he colored some or all of the items, compatible with a literal interpretation of "some."

## Experiment 1

### Methods

All methods and analyses were preregistered prior to data collection. The preregistrations can be viewed at https://osf.io/93xqp/overview?view_only=dce6dcbec698432e8ae0c60df30ae3bc (Act-Out+TVJ condition) and https://osf.io/dpk5g/overview?view_only=a490893d43fd437297afcdb4b5538098 (TVJ-only condition). All data and analysis code for both experiments are available at https://osf.io/85f67/overview?view_only=d84eb974e75b4b258d706aa2e730d97c.

### Participants

We recruited a total of 84 children between the ages of 4;0 and 5;11 ($M$ = 4;11) from preschools and a children's museum in [redacted]. All participants spoke English as their primary language. As per our preregistered exclusion criteria, two children were excluded for failing at

least three attention check trials, and two were excluded due to consistent errors in repeating the critical sentence. Our final sample included 80 participants.

**Materials & Procedure**

Participants were tested individually in-person using an interactive touch-based tablet task framed as a coloring game. Participants were tested individually in person using a touch-based tablet task framed as a coloring game. The game featured three characters: Lucy, Percy, and Mr. Monkey. Children were told that Lucy loved coloring and had already finished coloring her own coloring book. Mr. Monkey had an uncolored copy of the same book, as did children in the Act-Out+TVJ condition, and the goal was to make each copy look like Lucy's. Because Lucy's book was not shown, a puppet, Percy, guessed what Lucy had colored on each page. Children and Mr. Monkey then colored their own copies according to Percy's guess. Each page showed uncolored objects against a colored background, which was part of every copy of the book. Children were asked whether Mr. Monkey got it right or wrong, that is, whether he colored his page the way Percy said. Percy's guesses were modeled on the guessing-game method (Skordos et al., 2020; Tieu et al., 2017), although in the present task children judged Mr. Monkey's coloring rather than Percy's guess. Children were assigned to one of two between-subjects conditions: an Act-Out+TVJ condition and a TVJ-only condition. The two conditions shared the same characters, trial structure, and stimuli; they differed in whether children actively colored the pictures themselves prior to completing the TVJ task. In the Act-Out+TVJ condition, the child was an active participant in the game and was asked to color based on Percy's statement prior to watching Mr. Monkey do the same, whereas in the TVJ-only condition children did not get the opportunity to color based on Percy's statements, but only watched Mr. Monkey do so.

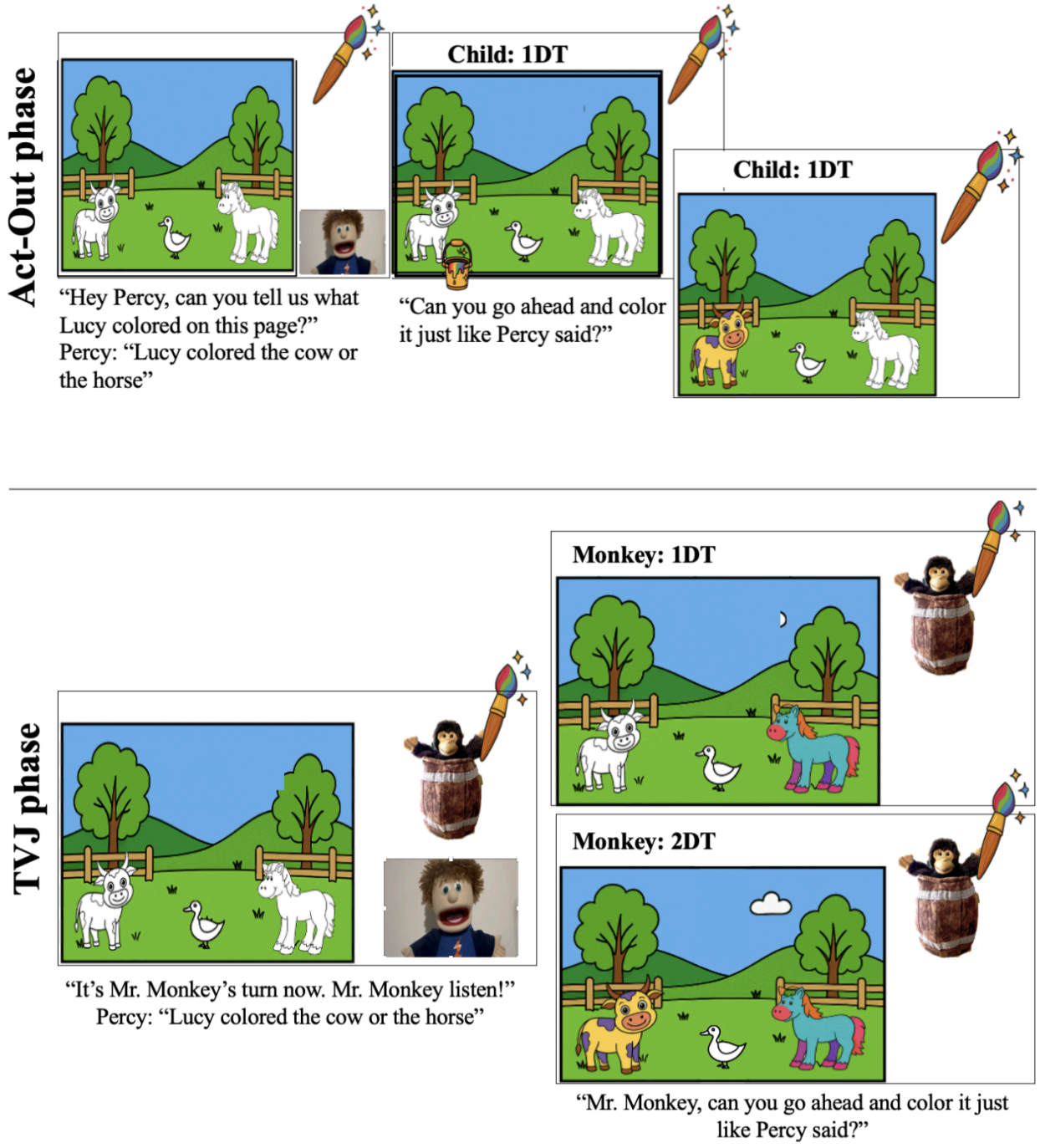



Figure 1. Sample trial structure for the Act-Out + TVJ condition. In the Act-Out phase (top), Percy the puppet describes a coloring action ("Lucy colored the cow or the horse") and the child is asked to color a new scene the same way. In the TVJ phase (bottom), Mr. Monkey colors the scene and children judge whether he colored it correctly. Monkey's coloring results in either a 1DT trials (Monkey colors the other mentioned object) or a 2DT trials (Monkey colors both mentioned objects)

In the Act-Out+TVJ condition, on each page children saw three uncolored objects of different types (e.g., a cow, a horse, and a duck) which they were given the opportunity to color (see Figure 4). Children first completed a short familiarization phase to learn the game mechanics. They were told that to color objects, they first had to tap the items they wished to paint (which caused a paint bucket to appear next to each selected item), and then tap a paintbrush to apply color to their selections. This two-step procedure ensured that children had clearly committed to their intended selection before any coloring happened.

After the familiarization phase, children completed two practice trials before the experiment began. In these trials, Percy's statement did not contain disjunction. For example, in a scene showing a seahorse alongside other sea animals, Percy said, "Lucy colored the seahorse." Children first colored the picture according to Percy's statement and then judged Mr. Monkey's coloring. On one trial, Mr. Monkey colored the named object (e.g., seahorse) and was therefore correct. On the other, he colored an unnamed object and was therefore incorrect. During these practice trials, children received explicit feedback on their judgments and repeated the trial if necessary.

Following the practice trials, children completed 10 experimental trials: six critical disjunction trials and four attention check trials. For descriptive purposes, we refer to trials in which only one of the two mentioned objects was colored as 1-Disjunct-True (1DT) trials, and trials in which both mentioned objects were colored as 2-Disjunct-True (2DT) trials. Because Mr. Monkey's response was contingent on the child's earlier action, children who consistently colored both objects did not contribute to trials in which both mentioned objects were colored in the judgment phase, whereas children who colored a single object on at least some trials contributed both types of trials.

In the Act-Out+TVJ condition experimental trials consisted of two parts: an Act-Out phase followed by a Truth Value Judgment phase. In the Act-Out phase, the experimenter addressed Percy and asked, “Percy, what did Lucy color on this page?”, inviting him to make a guess about the current page. Percy then appeared in a video window and, following the logic of previous studies using a guessing-game method (e.g., Tieu et al., 2017; Skordos et al., 2020), made a guess about what Lucy had colored using a disjunctive statement — for example, “Lucy colored the cow or the horse.” Before coloring, children were asked to repeat Percy's statement back to the experimenter. Children were then asked to color the picture just like Percy said (“Can you color it just like Percy said?”).[2] In the Truth Value Judgment judgment phase, which was present in both the Act-Out+TVJ and TVJ-Only conditions, children heard Percy’s disjunctive statement once more, after which Mr. Monkey colored the picture. Children were then asked whether Mr. Monkey got it right or got it wrong. In the Act-Out+TVJ condition children always acted on Percy’s statement themselves before making a truth value judgment. Crucially, Mr. Monkey’s coloring differed depending on the child’s earlier response in the act-out phase. If a child colored only one of the two mentioned objects (e.g., the cow), Mr. Monkey colored the other mentioned object on half of the trials (e.g., only the horse) and both mentioned objects on the other half (e.g., the cow and the horse). This allowed us to test whether children treated the statement as exclusive. If a child colored both mentioned objects in the act-out phase, Mr.

---

[2] To make sure that children had heard Percy’s sentence, they were asked to repeat it before coloring. If a child repeated the sentence correctly, no further repetitions were requested on subsequent trials. If a child instead replaced “or” with “and,” the experimenter did not correct it but replayed Percy’s utterance and asked the child to repeat it once more before continuing. No children repeated the sentence with an incorrect object. Three children initially repeated the first trial using “and,” corrected their response after hearing the sentence again, and continued with the task. As per our preregistered criteria, two children who consistently repeated Percy’s disjunctive statements using “and” were excluded from the final sample.

Monkey colored only one of the mentioned objects in the judgment phase, allowing us to test whether children interpreted disjunction conjunctively (by preferring to color both items but rejecting the coloring of only one) (see Figure 1).

There were also four attention check trials, two of which were single-target trials that resembled practice trials and did not include disjunction. In these trials, Percy made a simple statement (e.g., "Lucy colored the robot"), and Mr. Monkey colored the mentioned object on one trial and an unmentioned object on the other. The remaining two were two-target attention check trials, which appeared at the end of the experiment. In these trials, children saw displays containing two objects from the same category (e.g., two animals or two fruits) and one unrelated object (e.g., a tractor), and Percy said that Lucy colored the two matching objects. Mr. Monkey colored one matching object on one trial and both matching objects on the other. These trials confirmed that children were willing and able to select two objects when the statement explicitly required it, ruling out the possibility that selecting a single object in the critical disjunction trials reflected a general bias toward minimal responding or inattention.

In both conditions, trials were presented in one of two pseudo-randomized orders. Trials were presented in one of two pseudo-randomized orders, counterbalanced across participants. In one list, early critical trials were arranged such that, contingent on the child producing a 1DT response in the act-out phase, Mr. Monkey produced two consecutive 2DT trials before alternating with 1DT trials. In the other list, children instead first encountered two consecutive 1DT trials before alternation began. This counterbalancing ensures that any potential order effects (e.g., children anchoring on the first response type they observe) do not systematically bias the results. The spatial positions of the mentioned objects were also counterbalanced across

trials, such that the two disjuncts appeared in different pairings (left-middle, middle-right, left-right). The entire session lasted approximately 10 minutes.

**Results**

Our primary analyses followed our pre-registered plan. For children in the Act-Out+TVJ condition, we distinguished trials on which children produced 1DT responses from those on which they produced 2DT responses. This was important because children's initial act-out responses determined which Mr. Monkey outcome they saw in the subsequent judgment phase. In particular, following a 2DT response in the Act-Out phase, children only ever saw a 1DT outcome in the TVJ phase from Mr. Monkey. The resulting data were therefore not directly comparable across trial types and were analyzed separately. When there are too few observations of a given trial type to support a mixed-effects analysis, we report those data descriptively rather than fitting a full model.

To assess whether children understood the task, we examined performance on the attention check trials. In the Act-Out+TVJ condition, these trials required both carrying out the coloring step correctly and judging Mr. Monkey's response accurately, whereas in the TVJ-Only condition they required only correct judgment of Mr. Monkey's responses. On the single-target attention check trials of the Act-Out+TVJ condition, children always colored the mentioned object and were highly accurate in judging Mr. Monkey's response on these trials ($M$ = .96, $SD$ = .19). Children in the TVJ-only condition were similarly accurate ($M$ = .92, $SD$ = .26). On the two-target attention check trials in the Act-Out+TVJ condition, children typically selected both target objects on these trials ($M$ = .85, $SD$ = .35), showing that they were able to carry out instructions that explicitly required coloring two objects. Across conditions, children were also highly accurate in judging Mr. Monkey's response on these trials ($M$ = .88, $SD$ = .32). Taken

together, these findings show that children understood the task and responded as expected on the attention check trials, including correctly selecting one or two target objects when required.

In the Act-Out phase of the Act-Out+TVJ condition, children almost always produced 1DT responses (94.6% of trials, 227/240; *SD* = .221); 2DT responses were rare, occurring on only 5.4% of trials (13/240). Thus, when interpreting disjunctive statements, children almost never spontaneously represented scenes in which both disjuncts were true. Because 2DT responses were too infrequent to support inferential statistics, we report those responses descriptively only. Briefly, in our sample of 40 children, one child produced a 2DT response and accepted Mr. Monkey's subsequent 1DT response, and one child produced a 2DT response and rejected Mr. Monkey's 1DT response. Additionally, one child produced a 2DT response on a single trial (out of six) and accepted Mr. Monkey's 1DT response on that trial (see Fig. 2). In the remainder of the Results, we focus on trials on which children produced 1DT responses, as only these provided sufficient data for inferential analyses.

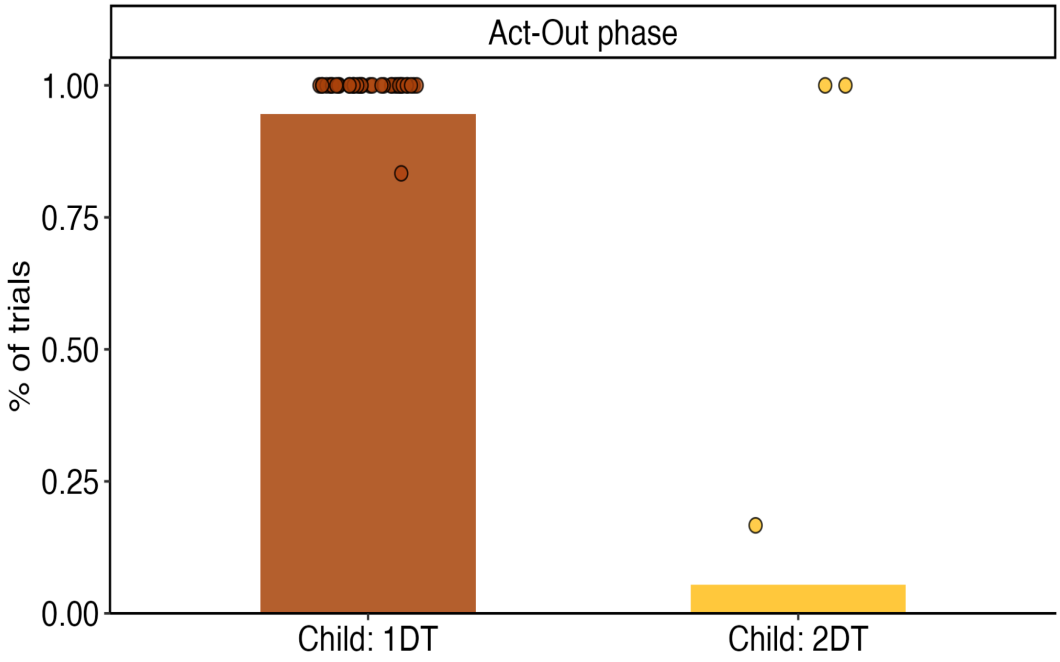



**Figure 2.** Proportion of trials on which children colored one object (Child: 1DT) or both objects (Child: 2DT) during the act-out phase, among all Act-Out+TVJ participants. Each dot represents one child's proportion.

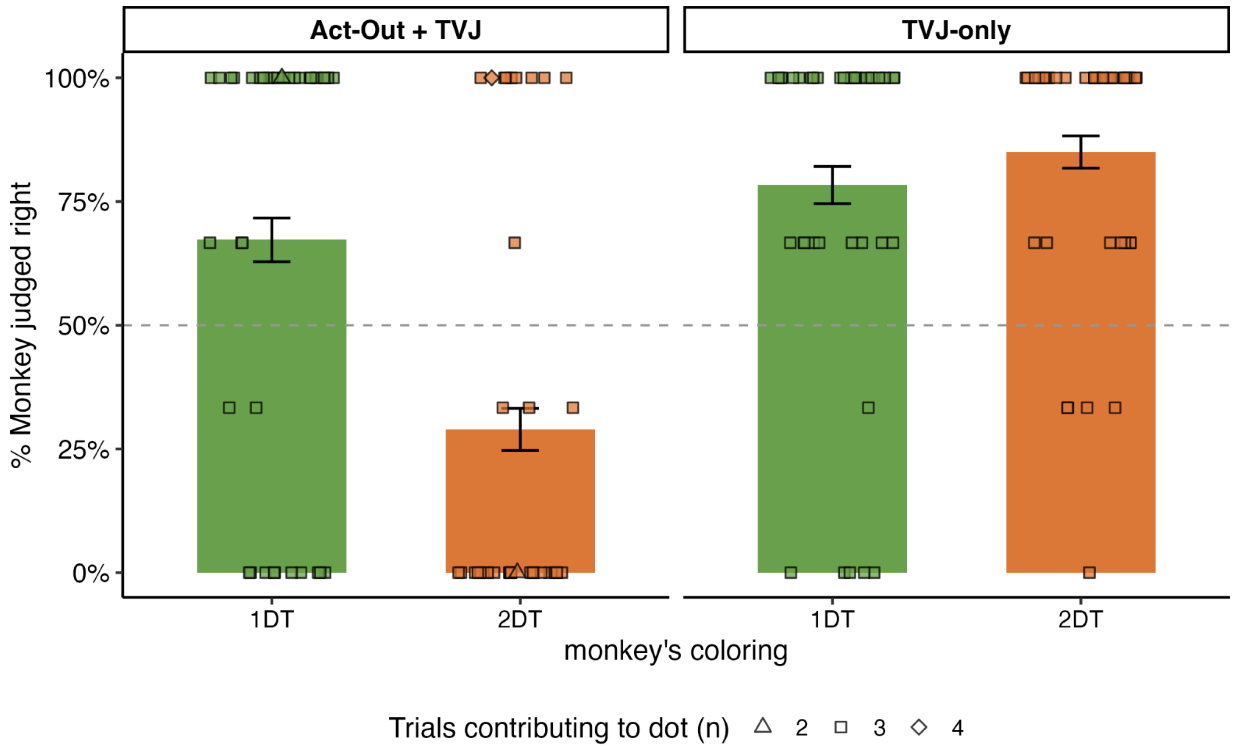



**Figure 3.** Proportion of trials on which children judged Mr. Monkey's coloring as correct, by Trial Type (1DT vs. 2DT) and Condition (Act-Out+TVJ vs. TVJ-only). Only children who colored one object in the act-out phase are shown for brevity; children who colored both objects (5.4% of Act-Out+TVJ trials) were too few. Bars represent trial-level means with standard error. Individual dots represent per-child proportions, with shapes indicating the number of trials contributing to each data point. The dashed line marks chance.

Our primary confirmatory analysis tested whether children's truth-value judgments differed as a function of whether they had completed the act-out task. Because only 1DT responses provided enough data and allowed comparison across 1DT and 2DT trial types, we focused on trials in which children produced a 1DT response in the act-out phase. We fitted a mixed-effects logistic regression model predicting whether children judged Mr. Monkey's response as correct, with fixed effects of Trial Type (1DT vs. 2DT), Condition (Act-Out+TVJ vs. TVJ-only), and their interaction, by-participant random intercepts, and by-item random intercepts. The model revealed a main effect of Condition ($\beta = -2.70$, SE = 0.64, $z = -4.21$, $p <$

.001), such that children were overall less likely to judge Mr. Monkey as correct in the Act-Out+TVJ condition than in the TVJ-only condition, and a main effect of Trial Type ($\beta = -1.29$, SE = 0.33, $z = -3.93$, $p < .001$), such that children were less likely to judge Mr. Monkey as correct on 2DT trials than on 1DT trials, both qualified by a significant Trial Type × Condition interaction ($\beta = -3.74$, SE = 0.67, $z = -5.57$, $p < .001$). To unpack this interaction, we conducted Bonferroni-corrected pairwise comparisons using the *emmeans* package (Lenth et al., 2018). Within the Act-Out+TVJ condition, children were significantly more likely to accept Mr. Monkey on 1DT trials than on 2DT trials ($\beta = 3.16$, $z = 6.00$, $p < .001$), consistent with an exclusive interpretation of disjunction. In contrast, within the TVJ-only condition there was no significant difference between 1DT and 2DT trials ($\beta = -0.58$, $z = -1.42$, $p = .94$). Crucially, children in the Act-Out+TVJ condition accepted Mr. Monkey's coloring on 2DT trials significantly less often than children in the TVJ-only condition (28.9% vs. 85.0%; $\beta = 4.57$, $z = 5.89$, $p < .001$), whereas acceptance on 1DT trials did not differ across conditions (67.3% vs. 78.3%; $\beta = 0.83$, $z = 1.25$, $p > .05$; see Figure 3).

**Discussion**

In Experiment 1, we found evidence that children compute implicatures for disjunction when first allowed to act on their own spontaneous interpretation of utterances. When children were asked to act-out their interpretation of disjunctive statements by coloring images, children almost always produced 1DT responses, suggesting that they almost never spontaneously represented scenes in which both disjuncts were true. However, because act-out data are potentially compatible with either an exclusive or an inclusive interpretation, evidence from children's truth-value judgments was crucial. After producing a 1DT response in the act-out

phase, children in the Act-Out+TVJ condition were significantly less likely to accept Mr. Monkey's coloring on 2DT trials than on 1DT trials, and were also significantly less likely to accept 2DT trials than children in the TVJ-only condition. In contrast, children in the two conditions accepted 1DT trials at similar rates, indicating that acting out the utterance did not lead children to reject Mr. Monkey's coloring in general, but specifically when both disjuncts were true. Children in the TVJ-only condition accepted 1DT and 2DT trials at similarly high rates, a pattern compatible with an inclusive interpretation of disjunction. These findings provide strong evidence that children can compute the exclusive interpretation of disjunction, and that this ability is revealed when they first have the opportunity to construct the relevant scene for themselves, but not in standard TVJ tasks.

## Experiment 2

In Experiment 2, we asked whether the pattern observed for disjunction would generalize to "some". As in Experiment 1, our goal was to compare a condition in which children first constructed their own interpretation of the utterance (Act-Out+TVJ) with one in which they judged an experimenter-provided scene without first doing so (TVJ-Only). To do so, we presented children with the same tasks as in Experiment 1, but in which children were asked to either color "some" of 4 items presented on a page (in the Act-Out task), or to judge how Mr. Monkey colored items when he colored only some (i.e. 1, 2, or 3 items) or all of them (i.e., 4).

### Methods

All methods and analyses were preregistered prior to data collection. The preregistrations can be viewed at https://osf.io/2tbej/overview?view_only=8ca725c6b4034d6c9f997ade9aeabf2d.

### Participants

We recruited a new set of 84 children between the ages of 4;0 and 5;11 ($M$ = 4;11) from preschools and a children's museum in [redacted]. All participants spoke English as their primary language. As per our preregistered exclusion criteria, four children were excluded for failing at least three attention check trials. Our final sample included 80 participants.

**Materials & Procedure**

We used the same two-step design as Experiment 1, but tested implicatures associated with the quantifier "some". Children again completed either an Act-Out+TVJ condition or a TVJ-only condition in the same tablet coloring game with Lucy, Percy, and Mr. Monkey. Unless noted below, the procedure was the same as in Experiment 1. In the Act-Out task, on each "page" of the electronic coloring book children saw four uncolored objects of the exact same type (e.g., four cows) which they were given the opportunity to color (see Figure 1).

Children first completed the same familiarization procedure to learn game logistics, and then completed two practice trials. Similar to Experiment 1, these trials did not contain the scalar expression, and children received explicit feedback on their judgments. To match the visual structure of the critical trials as closely as possible, the practice trials again involved multiple similar objects, but only one target matched Percy's description. For example, children might see two bears that differed by a salient property, such as whether they were wearing hats, and hear Percy say, "Lucy colored the bear with a hat." Children in the Act-Out+TVJ condition first colored the target object themselves and then judged Mr. Monkey's response. On one practice trial, Mr. Monkey colored the correct one (e.g., bear with a hat); on the other, he colored the incorrect one (e.g., bear without a hat).

Following practice, children completed 10 experimental trials: six critical "some" trials and four attention check trials. On each critical trial in the Act-Out+TVJ condition, Percy

described what Lucy had colored using "some" (e.g., "Lucy colored some of the cows"), and children were asked to color the picture just as Percy said ("Can you color it just like Percy said?"). In the subsequent Truth Value Judgment phase, Percy's statement was repeated, Mr. Monkey colored the picture, and children judged whether he got it right or wrong. In the act-out phase, children could color one, two, three, or all four objects, all of which are compatible with the standard analysis of the logical meaning of "some" (Barwise & Cooper, 1981). As in Experiment 1, Mr. Monkey's response depended on the child's earlier act-out response. If a child colored one, two, or three of the four objects in the Act-Out phase, Mr. Monkey colored three objects on half of the trials and all four objects on the other half in the Truth Value Judgement phase.[3] This allowed us to test whether children treated "some" as excluding "all": children who computed the implicature should reject trials on which Mr. Monkey colored all four objects, whereas children who interpreted "some" literally should accept both response types. If a child colored all four objects in the Act-Out phase, Mr. Monkey colored only three in the Truth Value Judgment phase. For descriptive purposes, we refer to trials in which fewer than four objects were colored as "Some-True" trials (satisfying the implicature), and trials in which all four objects were colored as "Some-False" trials (violating the implicature) (see Figure 4).

[3] On trials where children themselves had colored three objects, Mr. Monkey's three-object response was presented in a different configuration. For example, if a child colored the first three cows, Mr. Monkey's three-object response would color a different set of three cows rather than the same three. This ensured that children could not accept his response simply because it exactly matched their own earlier coloring.

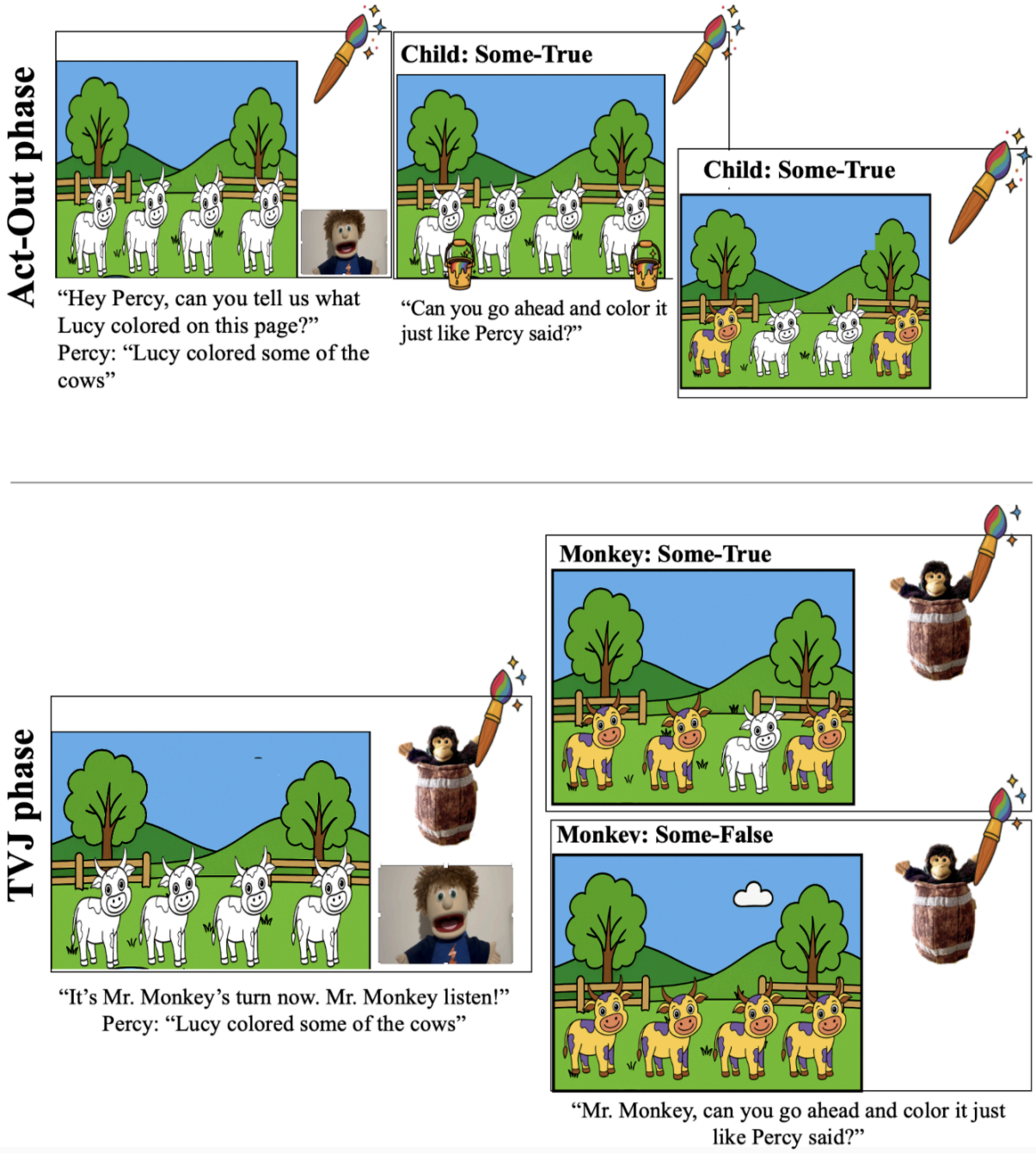



Figure 4. Sample trial structure for the Act-Out + TVJ condition. In the Act-Out phase (top), Percy the puppet describes a coloring action ("Lucy colored some of the cows") and the child is asked to color a new scene the same way. In the TVJ phase (bottom), Mr. Monkey colors the scene and children judge whether he colored it correctly. Monkey's coloring results in either a Some-True trials (Monkey colored three of four cows) or a Some-False trials (Monkey colored all four cows)

As in Experiment 1, attention check trials were designed to parallel the structure of the critical trials. Two were single-target attention check trials that resembled the practice trials. In these trials, children again saw multiple similar objects, but only one matched Percy's description. For example, children might see two pigs, only one of which was wearing boots, and hear Percy say, "Lucy colored the pig with boots." Or they might see two bears, only one of which was wearing a hat, and hear Percy say, "Lucy colored the bear with a hat." On these trials, Mr. Monkey colored either the correct object or the contrasting one. The remaining two attention check trials contained "all" and appeared at the end of the experiment, since previous studies find that access to statements including "all" may increase the likelihood of implicature in young children (e.g., Skordos & Papafragou, 2016). In these trials, Percy said that Lucy had colored all four objects (e.g., "Lucy colored all the dolphins", "Lucy colored all the aliens"), and Mr. Monkey colored either all four or only three. These trials ensured that children were willing and able to color all four objects when the utterance explicitly required it.

Similar to Experiment 1, trials were presented in one of two pseudo-randomized orders, counterbalanced across participants. Half the children first encountered critical trials where Mr. Monkey produced "Some-False" responses (coloring all four times), while the other half first encountered "Some-True" responses (coloring three items). This counterbalancing prevents children from simply imitating Mr. Monkey's early response pattern. The entire session lasted approximately 10 minutes.

**Results**

Our primary analyses were based on our pre-registered plan. We first report children's performance on the attention check trials. In the Act-Out+TVJ condition, these trials required children both to carry out the coloring step correctly and to evaluate Mr. Monkey's coloring,

whereas in the TVJ-only condition they involved only the judgment. On the single-target attention check trials, Percy mentioned one of the objects in the scene (e.g., the pig with boots) and children judged Mr. Monkey's coloring accurately (Act-Out+TVJ: $M$ = .94, $SD$ = .24; TVJ-only: $M$ = .96, $SD$ = .19). The other attention check trials contained "all" (e.g., "Lucy colored all the dolphins"), and children were expected to color all four objects in the scene in the Act-Out+TVJ condition, and they did so at high rates ($M$ = .96, $SD$ = .19), indicating that they were able to carry out instructions that explicitly required coloring every object. Children were also highly accurate in evaluating Mr. Monkey's response on these trials across both conditions (Act-Out+TVJ: $M$ = .99, $SD$ = .11; TVJ-only: $M$ = .96, $SD$ = .19). Together, these results indicate that children understood the task and responded as expected on the attention check trials, including correctly coloring all four objects when required.

We next consider how children interpreted "some" in the act-out phase of the Act-Out+TVJ task. On nearly all trials, children produced Some-True responses ($M$ = .97, $SD$ = .16), whereas Some-False responses were rare ($M$ = .02, $SD$ = .16). Among Some-True trials, children colored one object on some trials ($M$ = .32, $SD$ = .46), two objects on others ($M$ = .26, $SD$ = .44), and three objects on others ($M$ = .37, $SD$ = .48; see Figure 5). Thus, children almost never spontaneously colored all four objects in response to some. Because Some-False trials were too infrequent to support inferential analyses, we report those trials descriptively only. Only two children produced such responses: one child colored all four objects on all six trials, and another did so on one trial. On each of these trials, Mr. Monkey colored three objects, and both children rejected his coloring on every trial. In the remainder of the Results, we focus on trials on which children produced some-subset responses.

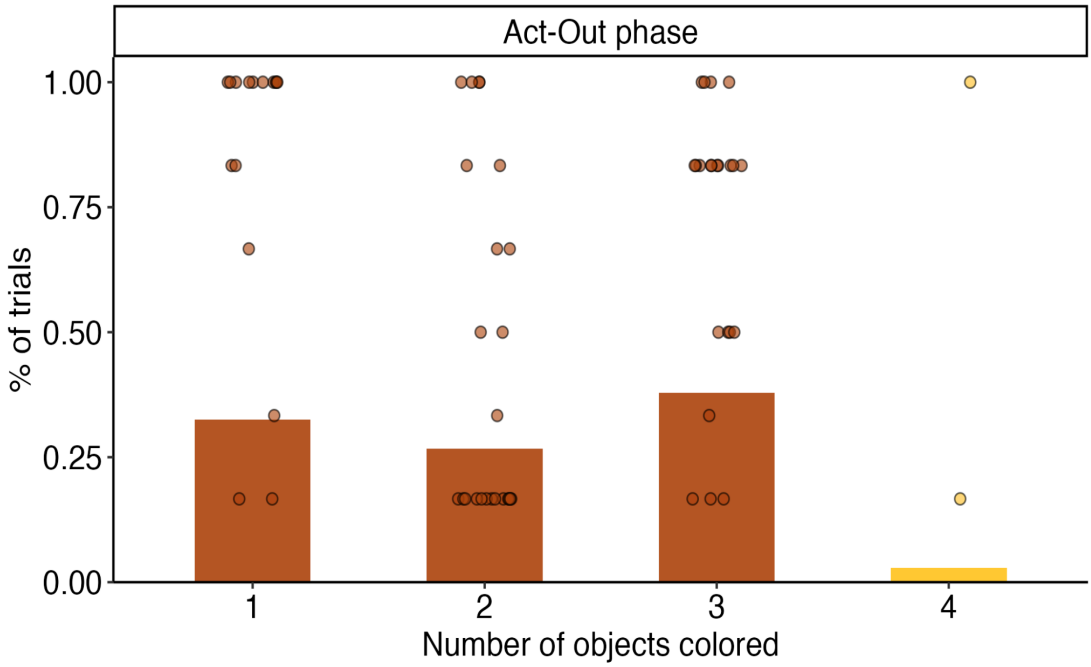



Figure 5. Proportion of trials on which children colored one, two, three (Some-True), or four (Some-False) objects during the act-out phase, among all Act-Out+TVJ participants. Each dot represents one child's proportion of their own trials falling in that category, with children who switched strategies across trials appearing as dots in multiple bars.

Our primary confirmatory analysis focused on trials in which children produced a Some-True response in the act-out phase (i.e., colored fewer than four objects). We fitted a mixed-effects logistic regression model predicting whether children judged Mr. Monkey's response as correct, with fixed effects of Trial Type (Some-True vs. Some-False), Condition (Act-Out+TVJ vs. TVJ-only), and their interaction, and by-participant random intercepts. The model revealed a main effect of Trial Type ($\beta = -4.85$, SE = 0.62, $z = -7.88$, $p < .001$), such that children were less likely to judge Mr. Monkey as correct on Some-False trials than on Some-True trials. There was also a main effect of Condition ($\beta = 1.70$, SE = 0.65, $z = 2.61$, $p = .009$), such that children were overall more likely to accept Mr. Monkey in the TVJ-only condition than in the Act-Out+TVJ condition, both qualified by a significant Trial Type ×

Condition interaction (β = 3.07, SE = 1.03, z = 2.97, p = .003). To unpack this interaction, we conducted Bonferroni-corrected pairwise comparisons using the *emmeans* package (Lenth et al., 2018). . In both the Act-Out+TVJ and TVJ-only conditions, children were significantly more likely to accept Mr. Monkey on Some-True trials than on Some-False trials (Act-Out+TVJ: β = 6.38, z = 6.22, p < .001; TVJ-only: β = 3.32, z = 6.83, p < .001). Crucially, rejection of Some-False trials was significantly higher in the Act-Out+TVJ condition than in the TVJ-only condition (82.1% vs. 50.0%; β = 3.23, z = 3.82, p < .001), whereas acceptance of Some-True trials did not differ across conditions (83.6% vs. 90.8%; p > .05; see Figure 6). Thus, while children in both conditions distinguished three-colored from four-colored trials, the coloring manipulation strengthened children's rejection of the infelicitous outcome.

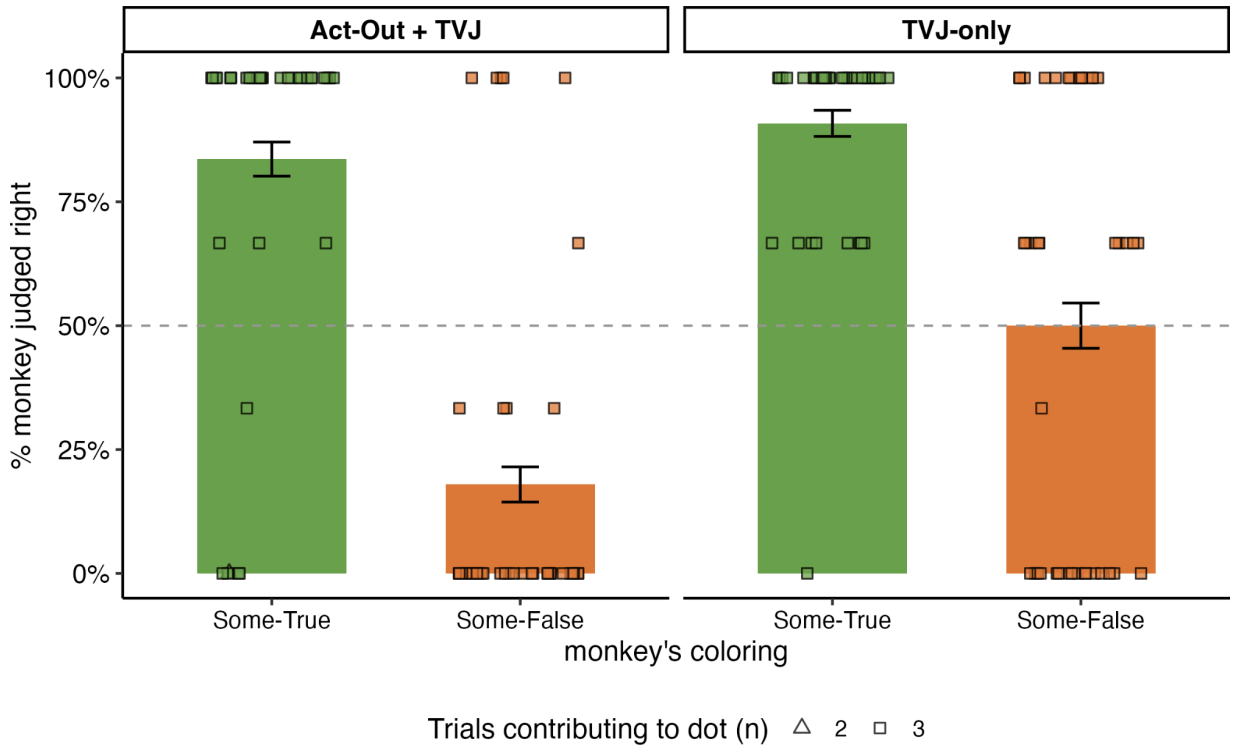



Figure 6. Proportion of trials on which children judged Mr. Monkey's coloring as correct, by monkey's coloring (Some-True vs. Some-False objects) and Condition (Act-Out + TVJ vs.

TVJ-only). Only children who produced subset responses in the act-out phase are shown for brevity. Individual dots represent per-child proportions, with shapes indicating the number of trials contributing to each data point. The dashed line marks chance.

**Discussion**

In Experiment 2, we found that children reliably computed scalar implicatures in a truth value judgment task when this was preceded by an act-out phase that allowed them to express their own spontaneous interpretation of "some". In the act-out phase, children colored a subset of the four available objects on almost all trials (one, two, or three objects), replicating previous act-out findings (Pouscoulous et al., 2007; Miller et al., 2005). Critically, children who first acted out the utterance were much more likely to subsequently reject Mr. Monkey when he colored all four objects relative to children in the TVJ-only condition. These results are consistent with the hypothesis that when children are able to construct and act on a spontaneous interpretation of an utterance, they readily compute implicatures when subsequently presented with alternative scenes. By contrast, children in the TVJ-only condition were much less likely to compute implicatures, at rates consistent with past studies that used TVJ or comparable tasks (e.g., Foppolo et al., 2012; Foppolo et al., 2021). These results are compatible with those of Experiment 1, and suggest that (1) children almost never spontaneously interpret "some" as compatible with "all," and (2) compute scalar implicatures at adult-like levels in a truth value judgment task if they are first allowed to express their own, spontaneous meaning prior to being proffered alternative worlds.

## General Discussion

We asked whether preschoolers spontaneously compute scalar implicatures when they are first allowed to construct their own models of an utterance before being proffered alternative worlds by the experimenter – as they might in naturalistic conversation. Across two experiments,

we tested children in a condition in which they first acted out their interpretation before making a truth value judgment and compared this to a condition in which they only made truth value judgments for experimenter proffered scenes. We found two main results. First, children almost never colored both disjuncts after hearing "or" (Experiment 1) and rarely colored all four objects after hearing "some" (Experiment 2). Second, when children made truth value judgments after first acting out their spontaneous interpretations, they rejected outcomes incompatible with strengthened meanings at significantly higher rates than children who made truth value judgments without first acting out their own meanings. This was true for both "some" and disjunction. Taken together, these results provide evidence that when children are allowed to generate their own models of an utterance before evaluating experimenter-provided worlds, they readily arrive at adult-like strengthened meanings and compute scalar implicatures.

In the Introduction, we raised the possibility that children only fail to compute implicatures when they are tested with stimuli that proffer "worlds" that they would otherwise never spontaneously consider when interpreting language. On this "proffered worlds" hypothesis, children fail to compute implicatures when they are shown scenes that do not correspond to their own spontaneous interpretation of an utterance. In particular, proffering worlds interferes with the child's ability to spontaneous generate their own model of how the world should look, and, as argued by Crain et al. (2001) and Skordos et al. (2022), may lead children to assume that the speaker has some other meaning in mind that corresponds to salient aspects of the provided scene. Adults may not struggle in this way because, unlike young children, they are better able to simultaneously represent multiple competing representations of the world and flexibly move between them – a capacity which is slow to develop in childhood, and impacts a range of behaviors including counterfactual reasoning (Rafetseder et al., 2010),

theory of mind (Perner & Lang, 1999), rule switching (Doebel & Zelazo, 2015), and the representation of mutually exclusive possible worlds (Gauffroy & Barrouillet, 2011; Sophian & Somerville, 1988).

In most naturalistic conversations, by contrast, children are free to generate their own spontaneous models of sentence meanings, without being led to consider worlds that might otherwise be strongly dispreferred. The results of Experiments 1 and 2 are compatible with this idea. For example, whereas most past studies of disjunction find that children interpret "or" inclusively in truth value judgment tasks – often endorsing scenes in which both disjuncts are true – we found that children almost never preferred scenes where both disjuncts were true when first allowed to act on their own spontaneous interpretation. We also found no evidence for so-called "conjunctive" responding (e.g., Singh et al., 2016), in which children only accept disjunction when both disjuncts are true. Similarly, whereas many past studies find that children often accept "some" when "all" is true (with rates typically ranging from 45-50% of the time, e.g., Foppolo et al., 2012; Foppolo et al., 2021 to 80-90%, e.g., Noveck, 2001), we found that when children first acted out their preferred meaning, they almost never colored all available objects, and were subsequently much more likely to compute implicatures and reject scenes where "all" was true.

In the Introduction, we noted that the proffered worlds hypothesis might explain not only children's difficulties with scalar implicature, but also other phenomena where children fail to make adult-like judgements, like Piagetian class inclusion tasks. In studies of these phenomena, children are also proffered worlds that they would otherwise be highly unlikely to spontaneously generate themselves. For example, in classic studies, Inhelder and Piaget (1964) presented children with questions like, "Are there more flowers or roses?" in displays with four roses and

two tulips. In such cases, a child might answer "roses" not because they fail to understand class inclusion, but instead because the question asked is pragmatically infelicitous, and the scene that is proffered is more compatible with a more pragmatically sensible question – i.e., "Are there more roses or tulips?" Compatible with this, even adults show significant processing costs when presented such items (Borst et al., 2013), suggesting that they also struggle to accommodate the provided scene with the test sentence. Similarly, children famously judge that the sentence, "Every rose is a flower," to be false when presented with a mix of roses and tulips. Here, too, children may infer that, given the scene presented, the experimenter's intended question must be whether every flower is a rose, both because this contrast is salient in the presented scene, but also because this is the only sensible and potentially informative question that could be intended in the context.

As alluded to in the Introduction, the Proffered Worlds hypothesis suggests a modest revision of theories which argue that children struggle with implicature due to difficulties accessing alternatives. On such accounts, children fail to compute implicatures because they do not spontaneously access relevant alternatives, and can be made to succeed when alternatives are explicitly provided, made salient by manipulation of the Question Under Discussion (Skordos & Papafragou, 2016), or via visual cues (e.g., as in *ad hoc* implicature; Barner et al., 2011; Barner et al., 2018; Foppolo et al., 2021; Franchin et al., 2024; Porrini et al., 2026; Stiller et al., 2015; Yoon & Frank, 2019). Common to all of these accounts, however, is that without these cues, children should fail to spontaneously identify the QUD and generate stronger alternatives. In contrast, the present results suggest that children may be able to access stronger lexical alternatives without them being directly provided. In the Act-Out+TVJ condition, no statement containing "and" or "all" preceded the critical trials, and the experimenter did not make these

alternatives salient in any other way, yet children rejected outcomes in which the stronger alternative was true. This suggests that children may not actually struggle to access relevant alternative utterances in naturalistic conversations, but may only struggle to do so after they have been presented with irrelevant worlds, which trigger unintended QUDs (as in standard truth value judgment tasks). Because formally a question can be defined as the set of alternative propositions that express answers to it, the act of contextually triggering an unintended QUD should also trigger irrelevant alternatives, while simultaneously eliminating relevant alternatives from consideration. For example, as noted by Skordos and Papafragou (2016), many past studies of implicature present children with sentences like, "Some of the people have shovels," in scenes where "all" is obviously true, but where the identity of objects varies systematically across trials, such that sometimes people have shovels, sometimes scarves, etc. Children may therefore assume that the QUD relates to the kinds of things some people have, rather than to the number of people that have shovels, such that, "Some of the people have scarves," is a relevant alternative, whereas, "All of the people have shovels," is not. Under these conditions, explicitly providing them with relevant alternatives involving "all" may be important because it directly activates the speaker's intended meaning, allowing them to "repair" their understanding of the intended QUD. By contrast, such repair – and being explicitly provided with linguistic alternatives – should not be necessary if the child is not presented with irrelevant worlds, and they spontaneously converge on the speaker's intended QUD and its concomitant alternatives.

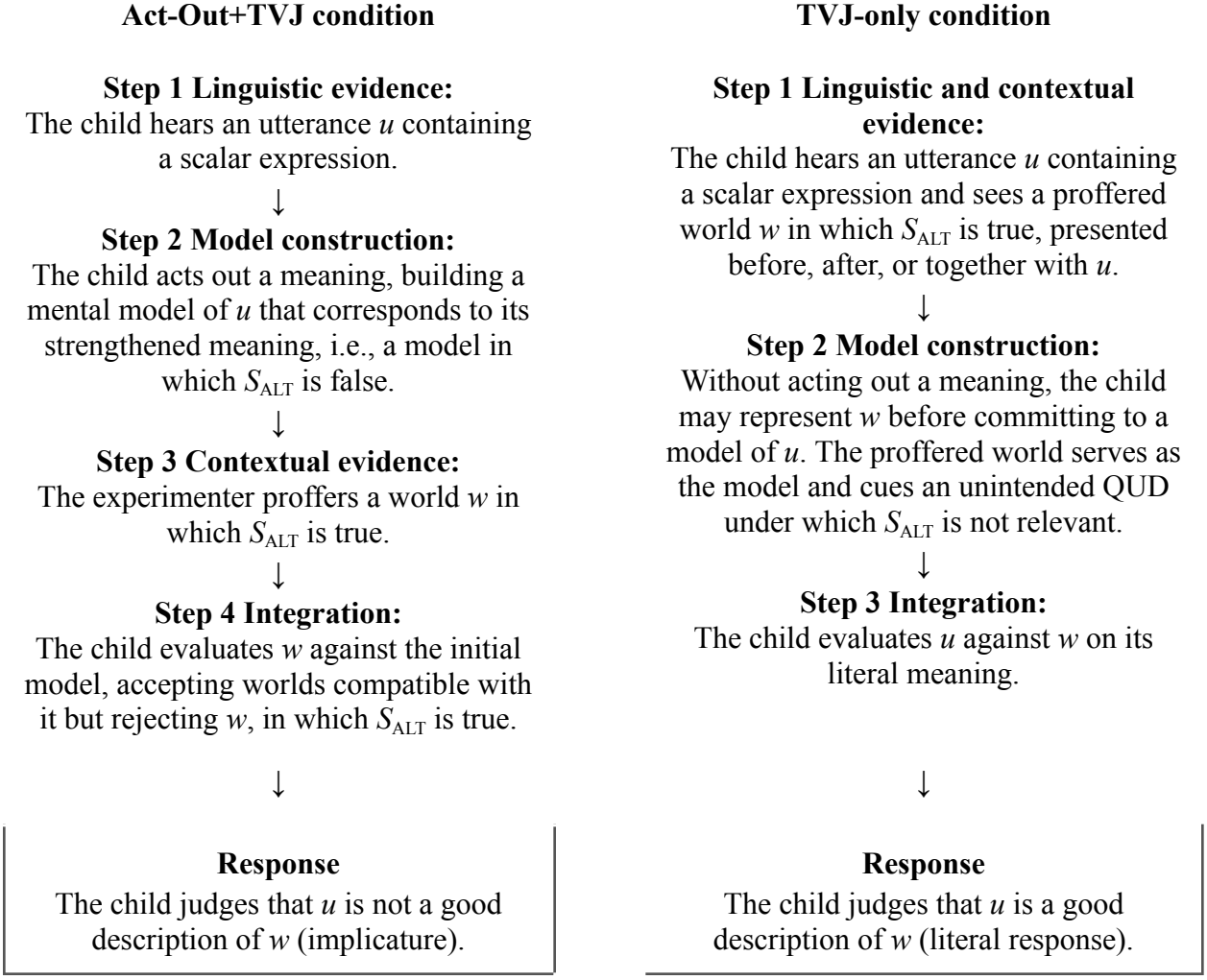



**Figure 7.** Possible impact of acting out a meaning on the integration of children's mental models and experimenter-proffered worlds. *u* = utterance, $S_{ALT}$ = stronger alternative, *w* = proffered world.

Our analysis assumes that acting out a meaning allows children to build a mental model of an utterance, which allows them to use this initial model as a basis for informing subsequent judgments in the TVJ task (see Figure 7). Still, alternative interpretations for our findings might be possible. For example, one possibility is that performing the Act-Out task made it easier for children to infer the speaker's goals, improving their performance relative to children tested only with the TVJ. Children who had just colored their own page may have inferred this goal more

readily. This possibility is compatible with the Proffered Worlds hypothesis. Inferring the speaker's goal can be understood as identifying the intended QUD, and what may make this harder in standard tasks is an irrelevant world presented before children interpret the utterance. Acting out the utterance first allowed children to establish their own interpretation before evaluating Mr. Monkey's coloring, so they did not need to infer the speaker's goal from his coloring and were less likely to consider alternative meanings it suggested. Second, coloring first may have made children less tolerant of pragmatic violations, in the sense of Katsos and Bishop (2011), for example because they were less willing to accept a coloring that differed from their own. Our data do not support this possibility. Children who colored first accepted colorings unlike their own as long as they were compatible with the strengthened meaning. When they had colored one object (e.g., the cow), they accepted Mr. Monkey coloring the other object (e.g., the horse) in Experiment 1, and when many had colored only one or two cows, they accepted Mr. Monkey coloring three in Experiment 2, in both cases as often as TVJ-only children did. They rejected only colorings on which the stronger alternative was true, that is, colorings incompatible with the implicature. Finally, children may have succeeded simply because coloring was engaging, improving their attention, motivation, or time to process each story, which could matter because strengthened interpretations are more costly to compute. However, being active and engaged does not seem to be sufficient. For example, Bleotu et al. (2025) tested children's interpretation of disjunction in a coloring task that began either from an empty scene or from a pre-set scene, in which one or both objects were already colored. Children's responses to pre-set scenes were inconsistent. When one object was already colored, nearly half colored the second as well, producing a scene in which both objects were colored. Yet when both objects were already colored, nearly half removed the color from one. Like the scenes in judgment tasks, pre-set

scenes proffer a world before children have interpreted the utterance, and children appear to be misled by them (see also Pouscoulous et al., 2007).

Our study adds to this literature in three ways. First, it shows that preschoolers can access stronger alternatives without being given them. When children interpreted an utterance before seeing an experimenter-provided world, they rejected outcomes in which the stronger alternative was true, for both "or" and "some", even though no stronger alternative had been mentioned. Second, it suggests why children fail in previous studies: worlds proffered by the experimenter before children interpret an utterance can cue an unintended QUD, so that children evaluate the utterance relative to the wrong question. Third, it introduces a new, two-phase method in which the same child first constructs the world they prefer for an utterance and then judges whether a different world is compatible with it. Although we used this method to study scalar implicature, it can be applied to other phenomena in which children's apparent failures may depend on the scenes they are shown.

One such domain may be children's broader logical reasoning. An open question is what role, if any, difficulties restricting the QUD play in children's broader logical reasoning, and how much the ability to mentally represent and entertain alternative possibilities matters for reasoning beyond the case of scalar implicature. Drawing a disjunctive inference, for instance, requires representing two states of affairs as mutually exclusive possibilities (e.g., "A or B", A is ruled out, therefore B), and similar demands on representing and selecting among alternatives arise in reasoning about modals, counterfactuals and possibility reasoning more broadly. Although infants and toddlers appear to succeed in simple reasoning-by-elimination tasks where they infer the location of a hidden object after seeing that one of the two possible locations is empty, preschoolers often fail in related tasks (see Cesana-Arlotti et al., 2018; Cesana-Arlotti &

Halberda, 2024; Feiman et al., 2022; Gautam et al., 2021; Mody & Carey, 2016). Notably, 2.5-year-olds who fail this kind of task succeed when the relevant premise (i.e., not "A") is conveyed linguistically (e.g., "There is no ball in the red cup"; Grigoroglou et al., 2019), perhaps because the linguistic cue makes clearer what the experimenter intends in showing the empty cup, and thus which possibility is meant to be ruled out. Preschoolers also overaccept necessity modals such as "have to" when an alternative option is available (e.g., accepting "Cora has to take the yellow road" when another road is also possible; Öztürk & Papafragou, 2015; Cournane et al., 2025; Dieuleveut et al., 2022), perhaps because they answer whether the mentioned route is a viable option rather than the only one, treating the experimenter's question as one about possibility rather than necessity. They also have difficulty reasoning counterfactually about alternative antecedents, often failing the "muddy boots" task when asked whether the floor would be clean if one character had taken his shoes off, even though another character had also tracked in mud (Rafetseder et al., 2013), though they perform better when the two muddy events are clearly separated in time and motive (Nyhout et al., 2019), perhaps because each antecedent then becomes more salient as a distinct alternative the question could be about. Future work should build on the findings we report here by exploring whether children's struggles across these domains reflect difficulties in recovering what is being asked when contextual features trigger an unintended QUD and activate irrelevant alternatives, and reasoning about possibilities more broadly (e.g., Leahy & Carey, 2020; Turan-Küçük & Kibbe, 2024).

In summary, our results suggest that when children are allowed to construct their own models of utterances, they successfully identify the speaker's intended Question Under Discussion, and compute strengthened interpretations that closely approximate adult performance. By contrast, when experimenter-proffered worlds precede interpretation, children

fail to compute implicatures.. Once children have identified the intended QUD – whether through their own spontaneous model construction or through explicit priming of appropriate alternatives – they readily make adult-like judgments. The critical challenge for children, then, may not be computing implicatures or accessing lexical alternatives *per se*, but recovering the intended interpretation when visual context has already cued them toward an unintended discourse structure.

## References

Aravind, A., de Villiers, J., de Villiers, P., Lonigan, C., Phillips, B., Clancy, J., Landry, S., Swank, P., Assel, M., Taylor, H., Eisenberg, N., Spinrad, T., & Valiente, C. (2017). Children's quantification with *every* over time. *Glossa: A Journal of General Linguistics*, *2*(1), 43.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, *60*, 40–62.

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

Barner, D., Hochstein, L., Rubenson, M., & Bale, A. (2018). Four-year-old children compute scalar implicatures in absence of epistemic reasoning. In K. Syrett & S. Arunachalam (Eds.), *Semantics in language acquisition* (pp. 325–349). John Benjamins.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Philosophy, language, and artificial intelligence: Resources for processing natural language* (pp. 241–301). Springer Netherlands.

Bleotu, A. C., Panaitescu, M., Benz, A., Nicolae, A., Bîlbîie, G., & Tieu, L. (2025). Coloring disjunction in child Romanian. *Experiments in Linguistic Meaning*, *3*, 65–74.

Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2013). Inhibitory control efficiency in a Piaget-like class-inclusion task in school-age children and adults: A developmental negative priming study. *Developmental Psychology*, *49*(7), 1366–1374.

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

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

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Annual Boston University Conference on Language Development* (Vol. 25, pp. 157–168). Cascadilla Press.

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

Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., & Woodams, E. (1996). Quantification without qualification. *Language Acquisition*, *5*(2), 83–153.

Dieuleveut, A., Van Dooren, A., Cournane, A., & Hacquard, V. (2022). Finding the force: How children discern possibility and necessity modals. *Natural Language Semantics*, *30*(3), 269-310.

Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, *38*, 241–268.

Feiman, R., Mody, S., & Carey, S. (2022). The development of reasoning by exclusion in infancy. *Cognitive Psychology*, *135*, 101473.

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*, *8*(4), 365–394.

Foppolo, F., Mazzaggio, G., Panzeri, F., & Surian, L. (2021). Scalar and ad-hoc pragmatic inferences in children: guess which one is easier. *Journal of Child Language, 48*(2), 350-372.

Franchin, L., Porrini, A. T., & Surian, L. (2024). Ad-hoc conversational implicatures in two-year-olds. *Language Learning and Development*, *20*(3), 219–231.

Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology*, *47*(4), 1000–1011.

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

Gordon, P. (1996). The truth-value judgment task. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 211–231). MIT Press.

Gotzner, N., Barner, D., & Crain, S. (2020). Disjunction triggers exhaustivity implicatures in 4- to 5-year-olds: Investigating the role of access to alternatives. *Journal of Semantics*, *37*(2), 219–245.

Grigoroglou, M., 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.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. In *Proceedings of Semantics and Linguistic Theory* (pp. 231–247).

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696.

Hamblin, C. L. (1976). Questions in Montague English. In B. Partee (Ed.), *Montague grammar* (pp. 247–259). Academic Press.

Hochstein, L., Bale, A., Fox, D., & Barner, D. (2016). Ignorance and inference: Do problems with Gricean epistemic reasoning explain children's difficulty with scalar implicature? *Journal of Semantics*, *33*(1), 107–135.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*, *89*(6), e572–e593.

Huang, H., & Crain, S. (2020). When OR is assigned a conjunctive inference in child language. *Language Acquisition*, *27*(1), 74–97.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation*. Routledge and Kegan Paul.

Jasbi, M., & Frank, M. C. (2021). Adults' and children's comprehension of linguistic disjunction. *Collabra: Psychology*, *7*(1), 27702.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*(1), 3–44.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

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

Lenth, R. V., Lenth, M. R. V., & Vdgraph, S. (2018). Package 'rsm'. R Foundation for Statistical Computing.

Miller, K., Schmitt, C., Chang, H. H., & Munn, A. (2005). Young children understand some implicatures. In *Proceedings of the 29th Annual Boston University Conference on Language Development* (Vol. 2, pp. 389–400). Cascadilla Press.

Minai, U., Jincho, N., Yamane, N., & Mazuka, R. (2012). What hinders child semantic computation: Children's universal quantification and the development of cognitive control. *Journal of Child Language*, *39*(5), 919–956.

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

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Noveck, I. A., Guelminger, R., Georgieff, N., & Labruyere, N. (2007). What autism can reveal about *every … not*sentences. *Journal of Semantics*, *24*(1), 73–90.

Nyhout, A., Henke, L., & Ganea, P. A. (2019). Children's counterfactual reasoning about causally overdetermined events. *Child Development*, *90*(2), 610–622.

Öztürk, Ö., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, *11*(3), 191–214.

Pagliarini, E., Crain, S., & Guasti, M. T. (2018). The compositionality of logical connectives in child Italian. *Journal of Psycholinguistic Research*, *47*(6), 1243–1277.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*(1), 71–82.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*(9), 337–344.

Philip, W. (1995). *Event quantification in the acquisition of universal quantification* [Doctoral dissertation, University of Massachusetts, Amherst].

Porrini, A. T., Franchin, L., & Surian, L. (2026). Role of Theory of Mind in children's derivation of ad hoc conversational implicatures. *Language Acquisition*, *33*(2), 181–192.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*(4), 347–375.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world." *Child Development*, *81*(1), 376–389.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404.

Reinhart, T. (2006). The acquisition of scalar implicatures. In T. Reinhart (Ed.), *Interface strategies: Optimal and costly computations* (pp. 272–292). MIT Press.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 6–1.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). Academic Press.

Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., & Fox, D. (2016). Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics*, *24*, 305–352.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18.

Skordos, D., Feiman, R., Bale, A., & Barner, D. (2020). Do children interpret 'or' conjunctively? *Journal of Semantics*, *37*(2), 247–267.

Skordos, D., Myers, A., & Barner, D. (2022). Quantifier spreading and the question under discussion. *Cognition*, *226*, 105059.

Sophian, C., & Somerville, S. C. (1988). Early developments in logical reasoning: Considering alternative possibilities. *Cognitive Development*, *3*(2), 183–222.

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*(2), 176–190.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2016). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, *33*(2), 269–298.

Tieu, L., Yatsushiro, K., Cremers, A., Romoli, J., Sauerland, U., & Chemla, E. (2017). On the role of alternatives in the acquisition of simple and complex disjunctions in French and Japanese. *Journal of Semantics*, *34*(1), 127–152.

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

Yoon, E. J., & Frank, M. C. (2019). The role of salience in young children's processing of ad hoc implicatures. *Journal of Experimental Child Psychology*, *186*, 99–116.