

1 Measuring Adults' and Children's Comprehension of Disjunction

2 Masoud Jasbi¹ & Michael C. Frank²

3 ¹ Harvard University

4 ² Stanford University

5 Author Note

6 All the experimental materials, data, randomization code, and analysis code for the
7 studies reported in this paper are available in the following online repository:
8 https://github.com/jasbi/disjunction_comprehension. The repository also includes
9 instructions for reproducing this research.

10 Correspondence concerning this article should be addressed to Masoud Jasbi, Boylston
11 Hall, Cambridge, MA. E-mail: masoud_jasbi@fas.harvard.edu

Abstract

Disjunction has had a key role in advancing theories of logic, language, and cognition. Previous research suggested that adults and preschool children might differ in their interpretation of linguistic disjunction in two ways. First, unlike adults, children might interpret a disjunction as conjunction. For example, children might consider a sentence like “*there is a cat or a dog*” only true if there is both a cat and a dog. Second, children might interpret *or* as inclusive disjunction when adults interpret it as exclusive. For example, adults may judge “*there is a cat or a dog*” as false when both animals are present, but children consider it true. We first review the long tradition of research on children’s development of disjunction. Then we present three studies that assess adults’ and children’s understanding of *and* and *or* in the same task, using three different measures: forced-choice binary truth value judgments, forced-choice ternary truth value judgments, and free-form verbal feedback. . . . The results suggest that forced-choice truth-value judgments may underestimate children’s pragmatic competence. In order to capture children’s semantic as well as pragmatic competence, we recommend complementing forced choice truth value judgment task with measures more sensitive to pragmatic infelicities.

Keywords: conjunction, disjunction, implicatures, semantics, pragmatics, logical connectives, language, acquisition, development, children

Word count: X

Measuring Adults' and Children's Comprehension of Disjunction

Introduction

When introducing disjunction to students of logic, Alfred Tarski (1941) complained about the complex factors that affect its comprehension in everyday language:

“The usage of the word *or* in everyday English is influenced by certain factors of a psychological character. Usually we affirm a disjunction of two sentences only if we believe that one of them is true but wonder which one. If, for example, we look upon a lawn in normal light, it will not enter our mind to say that the lawn is green or blue, since we are able to affirm something simpler, and at the same time, stronger, namely that the lawn is green. Sometimes even, we take the utterance of a disjunction as an admission by the speaker that he does not know which of the members of the disjunction is true. (Tarski, 1941, p. 21)”

In addition to this IGNORANCE implication – that the truth of neither disjunct is known – Tarski (1941) noted that a disjunction has at least two different interpretations: exclusive and inclusive. Suppose, “a child has asked to be taken on a hike in the morning and to a theater in the afternoon, and we reply: No, we shall go on a hike or we shall go to the theater” (Tarski, 1941, p. 20). Tarski explained that disjunction in this example is EXCLUSIVE because “we intend to comply with only one of the two requests” and not both. However, a disjunction may also have an INCLUSIVE interpretation like the following example: “Customers who are teachers or college students are entitled to a special reduction”. Tarski explained that *or* in this example is inclusive “since it is not intended to refuse reduction to a teacher who is at the same time a college student.”

Grice (1975) provided an pragmatic explanation for the complex set of interpretations that linguistic disjunction receives. He argued that the literal meaning of *or* (i.e. its

semantics) is captured by the truth conditions of logical inclusive disjunction. However, this literal meaning is enriched as speakers use a disjunction in context. Ignorance and exclusivity implications are inferences derived from our pragmatic reasoning on why the speaker used a disjunction. Grice (1975) generalized and systematized Tarski's intuition that we do not say "the lawn is green or blue" because we can say "something simpler and at the same time stronger"; namely "the lawn is green". He suggested a general communicative principle: speakers strive to be as truthful, informative, relevant, and brief as they can. Therefore, a disjunctive assertion commonly results in the inference that the speaker could not have uttered only one of the disjuncts, because s/he was uncertain about its truth (ignorance inference). Similarly, exclusivity of a disjunction is inferred by reasoning about the speaker's choice of the connective (*or* instead of *and*). Going back to Tarski's example, the child can reason that her dad could have said "we are going on a hike *and* we are going to the theater" if he intended to do both. He used *or* instead. Assuming he knew whether he wants to do both or not, his utterance must mean he wants to do one or the other (exclusivity inference). Within the Gricean framework, ignorance and exclusivity of *or* are secondary inferences, derived from the interaction of its literal (inclusive) meaning with conversational principles.

Complexities involved in the interpretation of disjunction have consequences for developmental theories. How does this intricate knowledge develop in humans? When do children begin to interpret a disjunction? What are their early interpretations like? Do they differ significantly from adult interpretations? Previous studies have suggested that preschool children (age 3-5 years) may differ in their interpretations of disjunction in two ways. First, they are more likely to interpret a disjunction word such as *or* as a conjunction (Braine & Romain, 1981; Neimark, 1970; Singh, Wexler, Astle-Rahim, Kamawar, & Fox, 2016; Tieu et al., 2016). Second, children are more likely to interpret a disjunction as inclusive. In other words, unlike adults, children do not compute exclusivity inferences, and therefore consider a disjunction as completely felicitous when both disjuncts are true (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Chierchia et al., 2004; Crain, 2008).

In the present study, we tested adults' and children's comprehension of disjunction using two- and three-alternative forced-choice judgment tasks. We also collected children's spontaneous verbal responses in the same tasks. The results of our experiments suggested that children (and adults) differentiated linguistic disjunction and conjunction: overall, they did not interpret *or* the way they interpreted *and*. In the context of predicting an outcome, both adults and children considered a true disjunction to be of "intermediate felicity": not wrong, but not completely right either. In adults, the infelicity of disjunction was better reflected in the judgment task with three alternatives, than the one with only two. In children, this infelicity was better reflected in their spontaneous verbal corrections, than their forced-choice judgments with either two or three alternatives. These results suggest that previously observed differences between adults and children's comprehension of disjunction may be at least partly due to the type of measurements used.

In the next section, we provide a broad review of the literature on children's acquisition of disjunction. We show that previous studies provide considerable support for the role of task design and measurement in shaping our conclusions regarding the development of disjunction. Experiment 1 tested adults in two- and three-alternative forced-choice tasks. Experiment 2 tested preschool children with a three-alternative forced choice judgment task. It also reports their spontaneous verbal feedback and corrections in the same task. Experiment 3 tested children's comprehension of disjunction in a two-alternative forced-choice task and reported their spontaneous feedback and corrections as well. In General Discussion, we report our conclusions regarding these experiments and their consequences for theories of semantic and pragmatic development.

Previous Research

Children's comprehension of logical connectives *and* and *or* have been studied within two research programs. The first program, starting in 1960s, was inspired by Piaget's

developmental theory (Inhelder & Piaget, 1958) and focused on the emergence of logical concepts in humans. The second research program started in late 1990s and was inspired by Grice's theory of meaning. Rather than conceptual development, it focused on linguistic development, separating the roles of semantics and pragmatics in language acquisition. In this section, we briefly outline some of the main findings in these two research programs, summarizing how task design and measurement affected their conclusions.

Within the Piagetian program, researchers hypothesized that the abstract and logical notion of disjunction (i.e. inclusive disjunction) is constructed from the more concrete concept of "choice between two options". The prediction was that until the age of 11 (concrete operational stage), children understand a disjunction like "A or B" as "one of the two options". This is similar to an exclusive meaning for disjunction. After age 11 (formal operational stage), children start to form abstract logical concepts and interpret "A or B" as inclusive. To examine this hypothesis, researchers conducted large scale in-class tests of school children and college students (Neimark & Slotnick, 1970; Nitta & Nagano, 1966). Participants were presented with pictures of objects and asked to circle those described by a statement such as "not bird", "bird and white", "bird or white". These studies concluded that the majority of the participants understood negation and conjunction, but only college students correctly answered statements with disjunction. They reported that participants made two types of "errors". First across all ages, some participants interpreted disjunction as conjunction. Second, some participants interpreted disjunction as exclusive. Based on these results Neimark (1970) concluded that a "correct" (inclusive) understanding of disjunction only develops in the high school years and depends on the attainment of formal operations as defined in the Piagetian theory¹.

¹The term "error" has different definitions in the literature on the comprehension of disjunction. Early studies considered any response other than an inclusive interpretation as erroneous. More importantly, what counted as an error was decided by researchers. Today, however, both exclusive and inclusive interpretations are considered correct and the conjunctive interpretation is more likely to be considered erroneous. Researchers

Further investigations suggested that the conjunctive errors may be due to the task design of in-class tests. Paris (1973) reported that in his in-class truth-judgment task, even a fifth of college students did not differentiate *or* from *and*, interpreting both as conjunction. He attributed these conjunctive interpretations of *or* to the application of non-linguistic strategies when the task is difficult or confusing (See Clark, 1973 for a discussion of nonlinguistic strategies in child language acquisition). He explained that children in his task (as well as some adults) were probably “comparing visual and auditory information with little regard for the implied logical relationship in the verbal description.” In a disjunction such as “A or B”, participants responded with “true” if the individual disjuncts (A, B) matched the pictures and false otherwise. Such a non-linguistic matching strategy would yield correct answers for conjunction but incorrect (conjunctive) answers for disjunction. This account also explains why in Paris (1973)’s study, conjunctive readings reduced with age and why using the word *either* along with *or* helped reduce conjunctive interpretations further.

Further evidence for the task-dependent nature of conjunctive readings or “errors” come from “give-item” tasks. Suppes and Feldman (1969) provided children with wooden blocks of different colors and shapes and used commands such as “give the things that are round or green.” They found that depending on the exact phrasing of the command, preschool children can interpret a disjunction as exclusive or conjunctive. However, using a similar “give-item” task, Johansson and Sjolín (1975)’s did not find considerable conjunctive interpretations. They tested Swedish-speaking children’s comprehension of disjunction in present tense sentences such as “Richard wants to drink lemonade or milk. Show me what he drank!” and imperative sentences such as “Put up [the picture of] the car or the doll!”. They reported that children, as young as four years of age, interpreted a disjunction as exclusive. Based on these findings, Johansson and Sjolín (1975) argued that the linguistic *or* should be

also focus more on adult-like vs. non-adult-like behavior in children rather than “errorneous” behavior. Depending on the context, a disjunction may be interpreted as exclusive, inclusive, or even conjunctive, and adults set the benchmark interpretation for children’s performance in experimental tasks.

kept separate from the logical notion of (inclusive) disjunction. While linguistic understanding of *or* develops early as exclusive disjunction, the logical understanding of it (as inclusive disjunction) develops late.

Braine and Romain (1981) tested participants with both a simplified replication of Suppes and Feldman (1969)'s "give-item" task and a version of what is today known as the truth value judgment task. For their replication of Suppes and Feldman (1969), they reported that both children and adults provided a "choose-one" (i.e. exclusive) interpretation of disjunction. They did not find any conjunctive interpretations, providing even further support for the role of task design. However, this was not the case in the truth value judgment task. In this task, a puppet described the contents of four boxes, each containing four animal toys. For example, the puppet said "Either there is a horse or a duck in the box." The first box had both animals, the second had only a horse, the third only a duck, and the last had neither. Participants were asked if the puppet was right. The results showed that adults were split between an inclusive and an exclusive interpretation of disjunction. The 7 to 10 year-olds were more likely to consider the disjunction as inclusive. However, the youngest group (5-6 years old) was most likely to interpret a disjunction similar to a conjunction: they said the puppet was right when both animals were in the box and not right or partly right if only one of the animals was in the box. Following Paris (1973), Braine and Romain (1981) argued that in this task, younger children do not take the contribution of the connective *or* into account. Instead, they use a non-linguistic strategy in which the disjunction is right if both propositions are true, partly right if only one is true, and wrong if neither is true. Braine and Romain (1981) concluded that children's ability to interpret a disjunction in a command develops earlier than their ability to judge its truth values.

In Braine and Romain (1981)'s judgment task, the puppet uttered a disjunction even though the content of the box was known to both the puppet and the participant (i.e. speaker lacked ignorance). As Tarski (1941) noted, such uses of disjunction sound odd

and infelicitous. This may have contributed to the application of a non-linguistic strategy and resulted in conjunctive readings. Later truth value judgment studies such as Chierchia, Crain, Guasti, and Thornton (1998) controlled for this effect of disjunction by making the puppet utter disjunction as a prediction of an unknown event, and let participants judge the prediction after they see the outcome. Furthermore, Chierchia et al. (1998) argued that in order to truly capture children's semantic competence with *or*, experiments need to test its comprehension in contexts that do not invite exclusivity implicatures. These contexts include embedding *or* under linguistic operators such as negation or conditionals.

Since Chierchia et al. (1998)'s arguments, numerous studies within the Gricean program have tested preschool children's comprehension of disjunction in embedded contexts as varied as negative sentences (Crain, Gualmini, & Meroni, 2000), conditional sentences (Gualmini, Crain, & Meroni, 2000), restriction and nuclear scope of the universal quantifier *every* (Chierchia et al., 2001, 2004), nuclear scope of the negative quantifier *none* (Gualmini & Crain, 2002), restriction and nuclear scope of *not every* (Notley et al., 2012a), and prepositional phrases headed by *before* (Notley et al., 2012b), as well as similar environments in other languages such as Mandarin Chinese and Japanese (Goro & Akiba, 2004; Su, 2014; Su & Crain, 2013). These studies almost unanimously support the hypothesis that the inclusive interpretation emerges earlier than the exclusive interpretation. This conclusion stands in sharp contrast to the earlier conclusions from the give-item tasks, however. Since under the Gricean account, the exclusive interpretation of disjunction is the result of pragmatic (scalar) implicatures, the earlier emergence of inclusive interpretations is considered consistent with evidence from development of quantifier implicatures. (Barner, Brooks, & Bale, 2011; Noveck, 2001; Papafragou & Musolino, 2003).

Methodological issues qualify this seemingly strong conclusion, however. As mentioned earlier, Braine and Romain (1981) found that the same children were more likely to interpret a disjunction as exclusive in a give-item task and inclusive/conjunctive in a truth value

judgment task. Therefore, truth value judgment tasks may not reveal the full picture regarding children's knowledge of exclusivity implicatures. Furthermore, several studies listed above test children's knowledge of disjunction in environments that largely collapse the distinction between *and* and *or*. For example, in the restriction of *every*, a conjunction and a disjunction can result in the same interpretation (e.g. *Every man or woman is happy* vs. *Every man and woman is happy*). Therefore, successful interpretation in such studies can also be achieved by a nonlinguistic strategy such as ignoring the contribution of *or* and independently checking the truth of each proposition, as discussed by earlier studies (Braine & Romain, 1981; Paris, 1973).

More recently, two truth value judgment studies reported that the majority of preschool children in their sample interpreted a disjunction similar to a conjunction (Singh et al., 2016, and Tieu et al. (2016)). To control for ignorance, Tieu et al. (2016) used the "prediction mode" of the Truth Value Judgment Task, in which the puppet provides a prediction or guess. Then an event occurs and participants are asked if the prediction was right. For example, there was a chicken on the screen and two toy objects, for example a bus and a plane. The puppet appeared on the screen and predicted that "the chicken pushed the bus or the plane". Then the chicken pushed either one or both of the objects. Participants stamped on a happy face or a sad face to show whether the puppet's guess was right or wrong. They reported that unlike adults, preschool children were more likely to consider a disjunction as "right" when both disjuncts were true, rather than only one. They concluded that children - the majority of them in their sample - interpreted disjunction as conjunction. They hypothesized that this conjunctive interpretation of disjunction is due to children's non-adult-like pragmatic enrichments.

However, a recent replication of Tieu et al. (2016) by Skordos, Feiman, Bale, and Barner (n.d.) suggests that the high rate of conjunctive interpretations were most likely due to the experimental context's lack of plausible dissent: the experiment did not provide

232 conditions under which utterances could be deemed false plausibly. They tested preschoolers
233 in two conditions: replication (two-alternatives) and three-alternatives. The first condition
234 was a direct replication of Tieu et al. (2016). The three-alternatives condition provided three
235 objects; for example a plane, a bus, and a bicycle. The reasoning was that if there are only
236 two objects, a disjunction is trivially true, and consequently children may consider that
237 unacceptable. The results replicated Tieu et al. (2016)’s findings in the replication condition,
238 but showed that conjunctive interpretations of disjunction disappeared almost completely in
239 the three-alternatives condition. Skordos et al. (n.d.) concluded that children’s conjunctive
240 interpretations are most likely due to non-linguistic strategies applied when they are
241 uncertain about some aspect of the experimental task. This conclusion is similar to the
242 conclusions of Paris (1973) and Braine and Romain (1981) in early studies of disjunction.

Table 1

Summary of tasks used in previous studies and their conclusions

Task	Conclusion
School test (Imperative) <i>Circle all that are bird or black!</i>	Comprehension of <i>or</i> develops in high school. Before that children often interpret it as conjunction.
School test (Truth Value Judgment) <i>The bird is in the nest or the shoe is on the foot.</i>	Children (8-14 years) as well as some adults interpret <i>or</i> as a conjunction. This is likely due to task demands and application of nonlinguistic strategies.
Give-item <i>Give me all the green things or give me all the round things.</i>	Children (4-7 years) interpret disjunction as exclusive (choose-one). The inclusive (logical) concept of disjunction develops later.

Task	Conclusion
Give-item + Truth Value Judgment <i>Give me all the green things or give me all the round things.</i> <i>Either there is X or there is Y in the box.</i>	Children (5-6 years) interpret <i>or</i> as exclusive in commands but ignore its contribution in truth value judgments and interpret it as a conjunction. Interpretation of disjunction in commands develops earlier than the knowledge of its truth conditions.
Truth Value Judgment (controlling for Speaker Ignorance) <i>A troll ate a piece of pizza or an ice cream.</i>	Children (4-6 years) understand <i>or</i> as inclusive disjunction. Two studies report majority conjunctive interpretations too.
Truth Value Judgment (controlling for Speaker ignorance and Number of alternatives)	Children (4-6 years) understand the truth conditions of <i>or</i> similar to inclusive disjunction. No evidence for conjunctive interpretations.

To summarize, our review of previous literature suggests that the design of an experimental task can have a big impact on our conclusions about children's comprehension of disjunction (Table 1). First, different tasks may be more or less suitable for capturing different interpretations of disjunction. For example, the "Give-item" task can successfully capture exclusive interpretations, while the TVJT task is more successful in capturing inclusive interpretations. Second, regardless of task type, increased task demands or infelicitous use of disjunction can result in increased conjunctive interpretations of disjunction. With the give-item task, Suppes and Feldman (1969) found a considerable rate of conjunctive interpretations, but these interpretations disappeared in Braine and Romain (1981)'s more simplified replication. Similarly, Tieu et al. (2016) reported that a large

number of children interpreted *or* as *and*, but these conjunctive readings also disappeared when Skordos et al. (n.d.)’s replication controlled for the number of alternatives in the task. Therefore, previous studies highlight the role of task design and measurement in studying children’s comprehension of disjunction. More specifically and with respect to conjunctive interpretations of disjunction, previous studies provide substantial evidence linking them to task design. Therefore, while it is plausible to consider non-adult-like pragmatic computations as a cause of conjunctive readings of disjunction in children, it is important to first conclusively rule out the influence of task design.

Present Study

The goal of this study was to further simplify task design and measure children’s comprehension of disjunction in multiple ways. We used simple existential sentences (e.g. *there is a cat or a dog*) and tested the interpretation of participants in a card game. The game controlled for the role of speaker ignorance, by making the speaker guess what was on a card without seeing it. It used trials with the conjunction word *and*, as well as adult participants as controls. Children’s interpretations were measured in three different ways: a two-alternative forced choice task (2AFC), a three-alternative forced choice task, and the analysis of children’s free-form verbal responses in each task. Table 2 provides the summary of methods used in Experiments 1, 2, and 3.

Table 2

Summary of Experiment 1, 2 and 3 methods

Study	N	Age	Mode	Response Options
Experiment 1	57	Adults	Online (Mturk)	Wrong, Right
	52	Adults	Online (Mturk)	Wrong, Kinda Right, Right
Experiment 2	42	3;1-5;2, M=4;3	Study Room	Circle (Wrong), Little Star (Little Right), Big Star (Right)

Study	N	Age	Mode	Response Options
Experiment 3	50	3;6-5;9, M=4;7	Study Room	Yes (Right), No (Wrong) + Open-ended Feedback

Experiment 1: Adult’s 2AFC and 3AFC Judgments

This study examines adults’ comprehension of *or*, and uses it as a benchmark for children’s comprehension in Experiments 2 and 3. We tested adults in both two-alternative and three-alternative forced choice tasks (2AFC and 3AFC).

Methods

Participants. 109 English speaking adults participated via Amazon Mechanical Turk (MTurk). 57 of them were assigned to a 2AFC judgment task and 52 to a 3AFC judgment task. In the 2AFC task, participants had to judge using the options “wrong” and “right”. In the 3AFC task they had to choose between “wrong”, “kinda right”, and “right”². The two conditions were otherwise identical. The task took about 5 minutes on average to complete. At the end of the study, participants received \$0.4 as compensation.

Stimuli. We used six cards, each with one or two cartoon animals on them. Three cards had one animal and three cards had two (Figure 17 in Appendix). We represent these six cards with animal names in small caps: CAT, DOG, ELE, CAT+DOG, CAT+ELE, DOG+ELE (ELE stands for elephant). In each trial, a card was shown to the participant and

²There are many possible labels for the middle option on a scale, including “kinda right”, “kinda wrong”, or “neither”. A later experiment, tested different intermediate labels and found that adults consider “kinda right” to be a more suitable option for capturing pragmatic infelicities (see Jasbi, Waldon, & Degen, 2019). We expect similar behavior from labels that refer to non-maximal degrees of being “right” such as “a bit right” or “a little right”.

a blindfolded cartoon character guessed what animal was on the card. The guess was either a simple existential sentence (e.g. *There is a cat*), one with a conjunction (e.g. *There is a cat and a dog*), or one with a disjunction (e.g. *There is a cat or a dog*). In this paper we use the short forms “*cat*”, “*cat and dog*”, and “*cat or dog*” to represent these guesses. Crossing the different types of cards and guesses results in 12 different possible trial types. We chose 8 trial types, balancing the number of one-animal vs. two-animal cards, simple vs. connective guesses, and (expected) true vs. false trials.

Figure 1 shows our trial types using example cards as rows and example utterances as columns. Control trials consisted of simple guesses (e.g. *elephant, cat*) with cards that had one animal (e.g. CAT) or two animals (e.g. CAT+DOG). In half of these trials the description was true and in half it was false. When two animals are on the card (e.g. CAT+DOG) and one is guessed (e.g. *cat*), the guess could be false if interpreted exhaustively (e.g. **only** *cat*). In addition to acting as a control, such trials could show how often children make exhaustive inferences. Conjunction trials (e.g. *cat and dog*) were controls for disjunction trials. Half of conjunction trials were false and half true (false when only one animal was on the card and true when both were.) Finally, disjunction trials constituted the critical trials of our experiments. When two animals were on the card (e.g. CAT+DOG), a disjunction guess (e.g. *cat or dog*) could be true if interpreted inclusively (e.g. *cat or dog, or both*), and false if interpreted exclusively (e.g. *cat or dog, not both*).

Procedure. The experiment had three phases: introduction, instruction, and test. In the introduction, participants saw the six cards and read that they would play a guessing game. Then a blindfolded cartoon character named Bob appeared on the screen. Participants were told that in each round of the game, they would see a card and Bob was going to guess what animal was on the card. The study emphasized that Bob could not see anything. Participants were asked to judge whether Bob’s guess was right. In the instruction phase, participants saw an example trial where a card with the image of a dog was shown



elephant	cat	cat and dog	cat or dog	
simple control false	simple control true	<i>connective</i> control false	<i>connective</i> critical true	
simple control false	simple control true exhaustivity implicature	<i>connective</i> control true	<i>connective</i> critical true scalar implicature	

Figure 1. Control and critical trial types in the connective guessing game. Rows show example cards and columns example utterances (guesses).

with the following sentence written above Bob’s head: *There is a cat on the card*. All participants correctly responded with “wrong” and proceeded to the test phase. In the test phase, participants saw one trial per trial type. Within each trial type, the specific card and guess were chosen at random. The order of trial types was also randomized. Figure 17 in the appendix shows an example test trial.

Results

Figure 2 shows the results for the adult 2AFC task. The two left columns show the simple guesses and serve as controls. The results show that if the guessed animal (e.g. *elephant*) was not on the card (e.g. CAT/CAT+DOG), participants judged the guess as “wrong”. If the guessed animal was on the card (e.g. *cat*), participants judged the guess to be “right”. The next two columns of Figure 2 show the results for the test conditions, namely conjunction and disjunction. A conjunction (e.g. *cat and dog*) was considered “wrong” if only

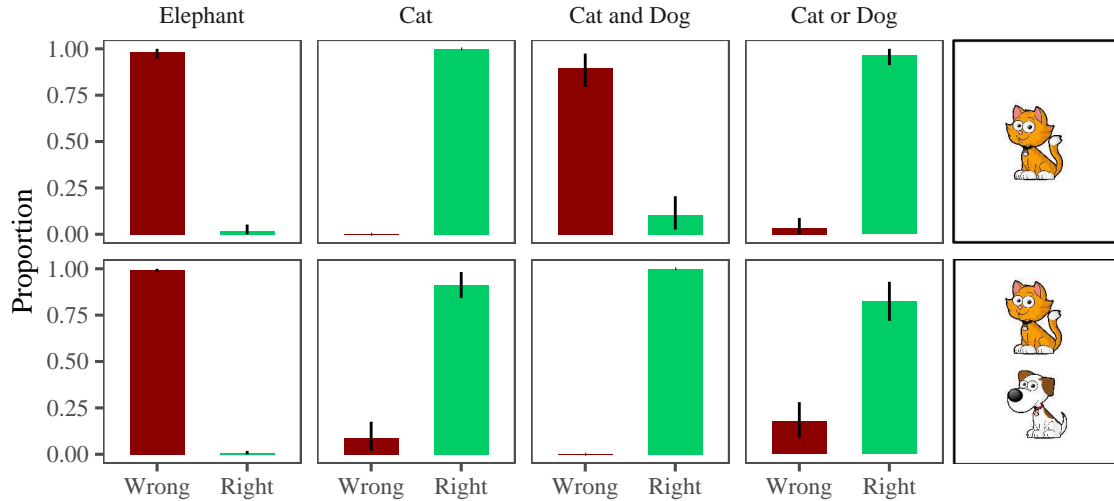


Figure 2. Adults' two-alternative forced choice judgments in Experiment 1.

one of the animals was on the card (e.g. CAT), and “right” if both were (e.g. CAT+DOG). A disjunction (e.g. *cat or dog*) was “right” whether one or both animals were on the card (e.g. CAT/CAT+DOG). The patterns of “right” and “wrong” responses in the two-alternative task match the expectations for truth and falsehood of conjunction and inclusive disjunction.

Figure 3 shows the results for the 3AFC judgment task. The addition of an intermediate response option affected the results in some trial types but not others. If the animal mentioned (e.g. *elephant*) was not on the card (e.g. CAT/CAT+DOG), participants judged the guess as “wrong”, regardless of the number of response options. Similarly, if the animal(s) mentioned (e.g. *cat*, *cat and dog*) perfectly matched the one(s) on the card (e.g. CAT, CAT+DOG), participants judged the guess as “right” in both 2AFC and 3AFC tasks. Judgments for such trial types had an extreme nature (completely wrong/right) and not an intermediate one.

The four remaining trial types showed different patterns of judgments in the 2AFC and the 3AFC tasks. If one animal was mentioned (e.g. *cat*) but two animals were on the card (e.g. CAT+DOG), participant judgments were divided between “right” and “kinda right”. When only one animal was on the card (e.g. CAT) and the guess was a conjunction (e.g. *cat*

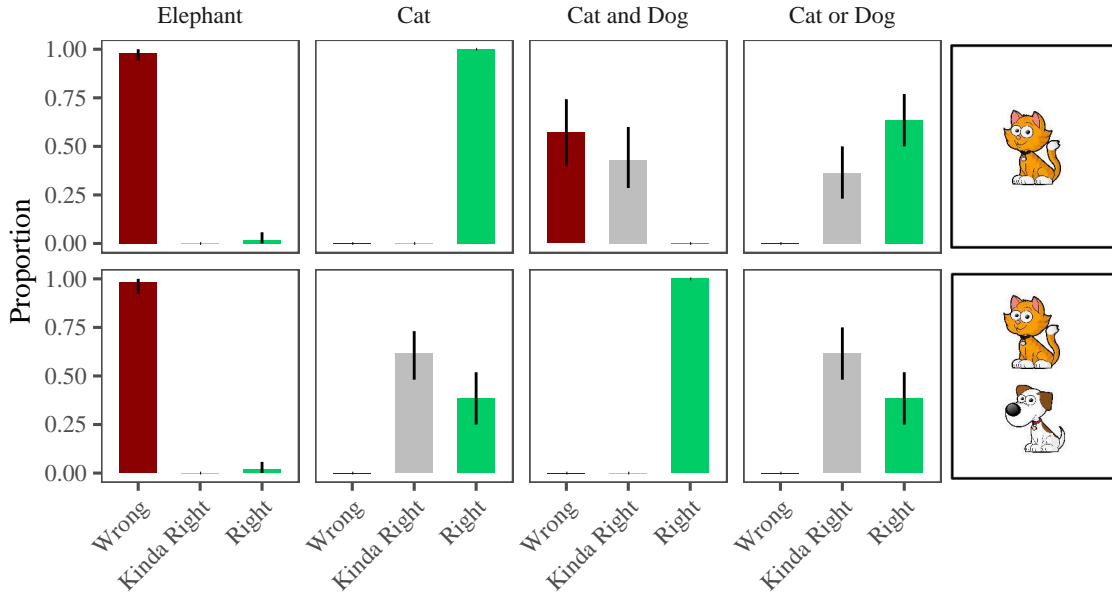


Figure 3. Adults' three-alternative forced choice judgments in Experiment 1.

and dog), most adults considered the guess “wrong” but some chose “kinda right”, possibly suggesting that the intermediate option was used to express partial truth of a guess. With disjunction guesses (e.g. *cat or dog*), responses were split between “kinda right” and “right”.

It is likely that for the different disjunction trial-types, participants had different reasons to consider them not completely right. When only one animal was on the card (e.g. cat), participants may have considered a simple guess (e.g. *there is a cat*) as completely right. When two animals were on the card (e.g. cat and dog), participants may have expected the connective *and* instead of *or*. As we shall see in the next two experiments, children explicitly mention these alternatives in their open-ended (free-form) responses.

Discussion

Table 3

Truth conditions of conjunction, inclusive, disjunction, and exclusive disjunction. “cat” and “dog” represent the propositions that “there is a cat” and “there is a dog” on the card respectively.

cat	dog	$\text{cat} \wedge \text{dog}$	$\text{cat} \vee \text{dog}$	$\text{cat} \oplus \text{dog}$
T	T	T	T	F
T	F	F	T	T
F	T	F	T	T
F	F	F	F	F

Consider the truth conditions for conjunction and disjunction shown in Table 3. A conjunction is true when both conjuncts are true and false otherwise. An inclusive disjunction is true when at least one disjunct is true, and false otherwise. An exclusive disjunction is true when only one of the disjuncts is true and false otherwise. Let’s also assume a simple linking function in which false statements are judged as “wrong” and true statements as “right” (see Jasbi et al. (2019) for a discussion of linking assumptions in such a task). In Experiment 1, this purely truth-conditional account has the following predictions: First, conjunction guesses like “cat and dog” are wrong when only one of the animals is on the card and right when both are. Second, disjunction guesses are always right if they are interpreted as inclusive (because in all disjunction trials at least one of the animals was on the card). However, if disjunction is interpreted as exclusive, it should be judged as wrong when both animals are on the card. Finally, if truth conditions are all that matters, the addition of the intermediate option “kinda right” should not substantially affect the judgments.

The results of the 2AFC and 3AFC tasks in Experiment 1 support two conclusions. First, judgments for *and* matched logical conjunction and *or* inclusive disjunction. If adults

in our task interpreted *or* as strictly exclusive, we expected majority “wrong” responses when both disjuncts were true. Yet neither task showed substantial “wrong” responses for disjunction trials. Second, addition of an intermediate option affected response patterns in some trial-types. These trials fell into two major categories. First, those with responses split between “wrong” and “kinda right”. Second, those with responses split between “kinda right” and “right”. The first category of responses, namely the split between “wrong” and “kinda right”, only happened when there was one animal on the card (e.g. CAT) and the guess was a conjunction (e.g. *cat and dog*). In such trials, even though the guess was false, it was not completely incorrect; one of the animals was guessed right. Therefore, choosing the intermediate option could reflect the judgment that such guesses are better than those that fail to name any animal on the card.

The second category corresponds to trial types in which the guesses were literally true, but underinformative. When there were two animals on the card (e.g. CAT+DOG), guessing only one of them (e.g. *cat*) or guessing a disjunction of them (e.g. *cat or dog*) results in a true yet sub-optimal statement. In these cases, there was a better possible guess, namely a conjunction (e.g. *cat and dog*). Similarly, when only one animal was on the card (e.g. CAT), and Bob guessed a disjunction (e.g. *cat or dog*), the guess was true but not the best one. A simple utterance (e.g. *there is a cat*) would have been better. Therefore, disjunction guesses (with either one or both disjuncts being true) had intermediate acceptability. This is generally the case in the context of guessing and prediction: a guess or prediction with a disjunction may be true but it is not the most informative guess. There is always a better or more specific guess.

In a forced choice task with no intermediate option, participants may differ on how they respond to cases of intermediate acceptability. Some may decide to ignore the slight unacceptability and focus on the truth of the statement. Others may decide to focus on the fact that a better guess was not made and express this in their judgments. This decision is

independent of a participant’s judgment of the linguistic stimuli, and depends on several factors including what matters for the purposes of the task and what type of measurement is used. For example, in a two-alternative task, most adults may not consider non-truth-conditional violations grave enough to render a guess as “wrong”. Therefore, judgments in a 2AFC task match the truth of a guess. However, if a third intermediate option is provided, participants may opt to also express the unacceptability or infelicity of a guess in the task - depending on the label of the intermediate option. In a subsequent study, we found that participants opt for the intermediate option more often if it is labeled as “kinda right” rather than “neither” (Jasbi et al., 2019). Most importantly, children may differ from adults in how they approach intermediate judgments in forced choice tasks. This source of variation between children and adults has remained relatively unexplored, despite previous evidence for it (Katsos, 2014; Katsos & Bishop, 2011). The next two experiments provide evidence that children may differ from adults in how they deal with the intermediate nature of disjunction in forced-choice tasks.

Experiment 2: Children’s three-alternative forced choice judgments vs. open-ended verbal feedback

This experiment tested children’s comprehension of disjunction in the same guessing game and compared them to those of adults’. Since the 3AFC judgment task in Experiment 1 was better at capturing the nuances of adults’ pragmatic reasoning, we decided to first test children with the 3AFC task. We also provide an analysis of children’s open-ended and spontaneous verbal feedback to the guesses.

Methods

Participants. We recruited 42 English speaking children from the Bing Nursery School at Stanford University. Children were between 3;1 and 5;2 years old (Mean = 4;3).

Materials. We used the same set of cards and linguistic stimuli as the ones in study

1. There were 8 trial types and 2 trials per trial type for a total of 16 trials. We made two changes to make the experiment more suitable for children. First, instead of the fictional character Bob, a puppet named Jazzy played the guessing game with them. Jazzy wore a sleeping mask over his eyes during the game (Figure 18). Second, a pilot study showed that a scale with three alternatives is better understood and used by children if it is presented in the form of rewards to the puppet rather than verbal responses such as “wrong”, “a little bit right”, and “right”, or even hand gestures such as thumbs up, middle, and down. Therefore, we placed a set of red circles, small blue stars, and big blue stars in front of the children. These tokens were used to reward the puppet after each guess. During the introduction, the experimenter explained that if the puppet was right, the child should give him a big star; if the puppet was a little bit right, a little star, and if he was not right, a red circle. Katsos and Bishop (2011) was the first study to use three types of rewards as response options (small strawberry, big strawberry, huge strawberry) to test preschool children’s semantic and pragmatic competence. They reported that children used the intermediate response option when an utterance was pragmatically infelicitous.

Procedure. The experiment was carried out in a quiet room with a small table and two small chairs. Children sat on one side of the table and the experimenter and the puppet on the other side facing the children. The groups of circles, small stars, and big stars were placed in front of the child from left to right respectively. A deck of six cards was in front of the experimenter. Similar to study 1 with adults, study 2 had three phases: introduction, instruction, and test.

The goal of the introduction was for the experimenter to show the cards to the children and make sure they recognized the animals and knew their names. The experimenter showed the cards to the children and asked them to label each animal. All children recognized the animals and could label them correctly. In the instruction phase, children went through

three example trials. The experimenter explained that he was going to play with the puppet first, so that the child could learn the game. He removed the six introduction cards and placed a deck of three cards face-down on the table. From top to bottom (first to last), the cards had the following images: CAT, ELE, CAT+DOG (Table 4). The experimenter put the sleeping mask on the puppet’s eyes and explained that the puppet is going to guess what animal is on the cards. He then picked the first card and asked the puppet: “What do you think is on this card?” The puppet replied with “*There is a dog*”. The experimenter showed the CAT-card to the child and explained that when the puppet is “not right” he gets a circle³. He then asked the child to give the puppet a circle. Rewards were collected by the experimenter and placed under the table to not distract the child. The second trial followed the same pattern except that the puppet guessed “right” and the experimenter invited the child to give the puppet a big star. In the final trial of the instruction, the puppet guessed that *there is a cat* on the card when the card was CAT+DOG. The experimenter said that the puppet was “a little right” and asked the child to give him a little star.

In the test phase, the experimenter removed the three instruction cards and placed a deck of 16 randomized cards on the table. He explained that it was the child’s turn to play with the puppet. For each card, the puppet provided a guess and the child provided the puppet with a reward. The guesses were paired with each card in a way that allowed two trials per 8 trial types⁴.

Offline Annotations. While playing the game, children often provided spontaneous verbal reactions to the puppet’s guesses. During the analysis of the videos, these verbal responses were categorized into four types: 1. None, 2. Judgments, 3. Descriptions, and 4. Corrections. The first category referred to cases where children did not say anything and

³The pilot study had shown that some children struggle with understanding the word “wrong”, so “not right” was used instead.

⁴A more detailed description of the procedure as well as the randomization code for the test phase is available on the study’s online repository.

only rewarded the puppet. The second category (judgments) referred to positive/negative linguistic feedback that did not include information about the animals on the card. Examples include: “you are right!”, “yes”, “nope”, or “you winned!”. In the third category (descriptions), children labeled the animals on the card: “cat!”, “dog and elephant!”, “There is a cat and a dog!” etc. Finally, with correction, children provided extra “focus words” such as *just*, *only*, and emphasized *AND* that acted like corrections to what the puppet had said. Examples include: “Just a cat!”, “Both!”, “The two are!”, “Only cat!”, “cat AND dog” (with emphasis placed on *and*). In trials where the child provided both judgments as well as descriptions or corrections (e.g. “Yes! Cat!”), we placed the feedback into the more informative categories, namely description or correction.

Results

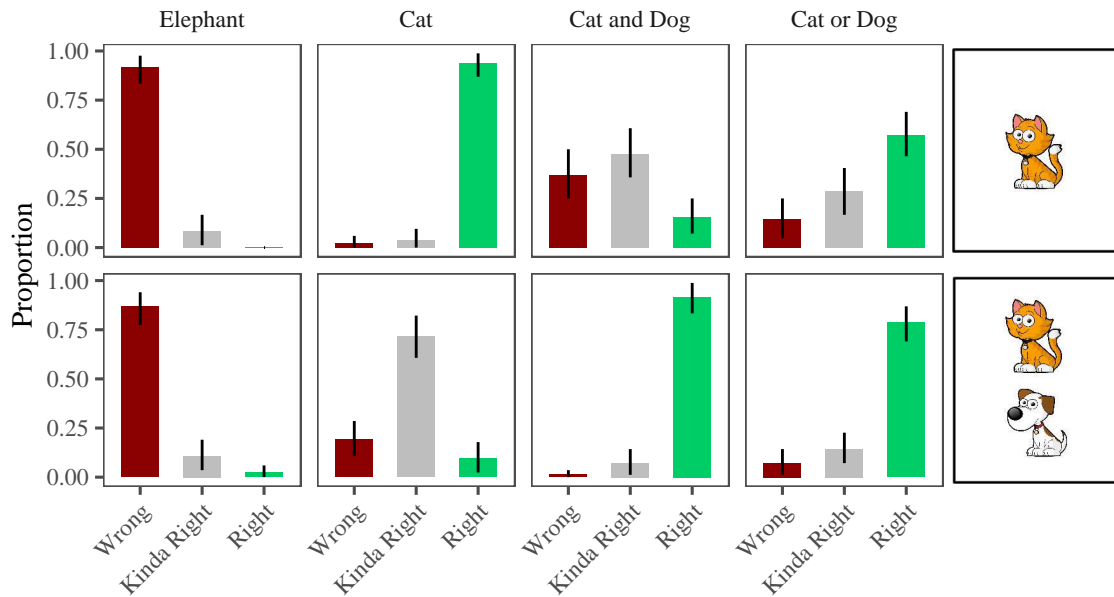


Figure 4. Children’s three-alternative forced-choice judgments in Experiment 2.

Truth Value Judgments. Figure 4 shows the results for children’s 3AFC judgments. Starting with control trials (leftmost column in Figure 4), if the mentioned animal (e.g. *elephant*) was not on the card, children judged the guess as “wrong”. If the

mentioned animal (e.g. *cat*) was the only animal on the card, children judged the guess to be “right”. Here we ignore the results for exhaustivity trial types in which the animal mentioned (e.g. *cat*) was only one of the animals on the card (e.g. CAT+DOG). The reason is that such trials were used in the instruction phase to introduce the “little bit right” option, and the results are probably biased by the instructions. With conjunction guesses (e.g. *cat and dog*), when only one of the animals was on the card (e.g. CAT), most children judged the guess as “wrong” or “a little right”. However, if both animals were on the card (e.g. CAT+DOG), they judged the conjunction as “right”. With disjunction guesses (e.g. *cat or dog*), when only one of the animals mentioned was on the card (e.g. CAT), children considered the guess “right” or “kinda right”. If both animals were on the card (e.g. CAT+DOG), the disjunction was considered “right”.

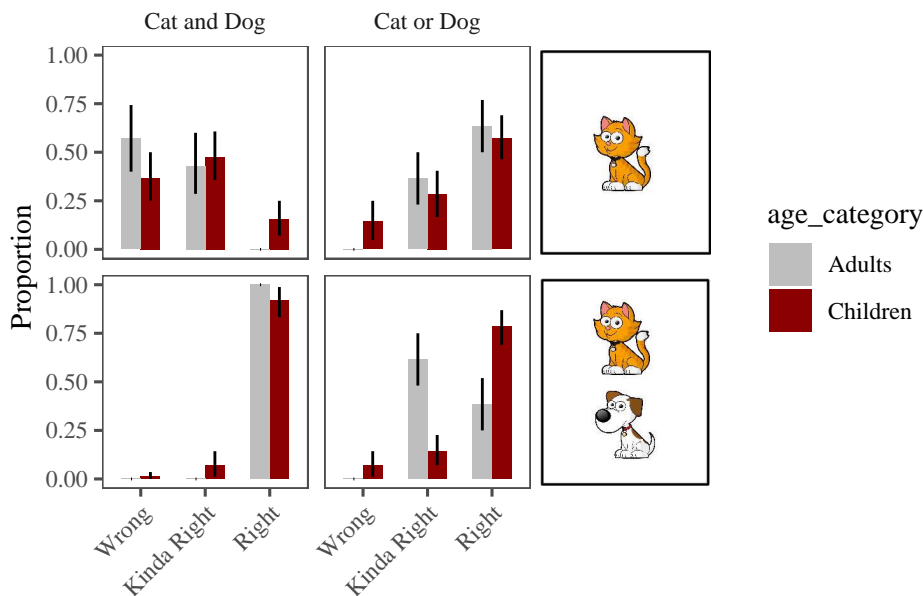


Figure 5. Comparison of Adults’ 3AFC judgments from Experiment 1 and Children’s 3AFC judgments from Experiment 2.

MIKE→

Figure 5 compares the results for children and adults’ 3AFC judgments in the conjunction and disjunction trials. Overall, the results look very similar. To quantify possible

differences between adults and children more precisely and model both our 3AFC task as well as the subject-level clustering of data, we decided to fit ordinal mixed-effects logistic models. We adopted the Bayesian framework and used the R package “brms” (Bürkner, 2017).

First, we fit separate ordinal mixed-effects logistic models for adults and children. The models included the fixed effect of trial-type and maximal random-effects structures (Barr, Levy, Scheepers, & Tily, 2013), i.e. random intercepts and slopes for participants and items (cards).⁵ Second, we fit an ordinal mixed-effects model to the combined dataset of adults and children with the added interaction effect of “age category” (adult vs. child), with “adults” set as the intercept.⁶ For all models, the response variable had three ordered levels: “wrong”, “kinda right”, and “right”. The trial types “two animals, *and*”, “two animals, *or*”, and “one animal, *and*” constituted the (dummy-coded) fixed effects of the model, with “one animal, *or*” set as the intercept. The priors over trial types were set to $\mathcal{N}(0, 10)$. For other parameters, default weakly informative priors – Student-t (3, 0, 10) and Cholesky LKJ Correlation (1) – were used as endorsed in “brms” documentation. All four chains converged after 4000 samples (with a burn-in period of 2000 samples).

The left panel of Figure 6 shows the mean and 90% highest posterior density intervals (HPDIs) for three contrasts of interest estimated from separate ordinal models for adults and children. Because predictors were dummy-coded, it is possible to examine contrasts of interest by computing the difference between coefficients for pairs of conditions we wish to contrast. The x-axis shows the contrasts between the following trial-types: conjunction vs. disjunction trials when only one animal was on the card (one animal, *and-or*), trials with one vs. two animals on the card when the guess was a disjunction (*or*, two-one animal), and conjunction vs. disjunction trials when two animals were on the card (two animals, *and-or*). Adults’ and children’s estimated coefficients are similar in sign to one another in the first and third contrasts, though adults’ are more extreme. The first contrast (one animal, *and-or*)

⁵response \sim trial type + (1 + trial type|sid) + (1 + trial type|card)

⁶response \sim trial type * age category + (1 + trial type|sid) + (1 + trial type|card)

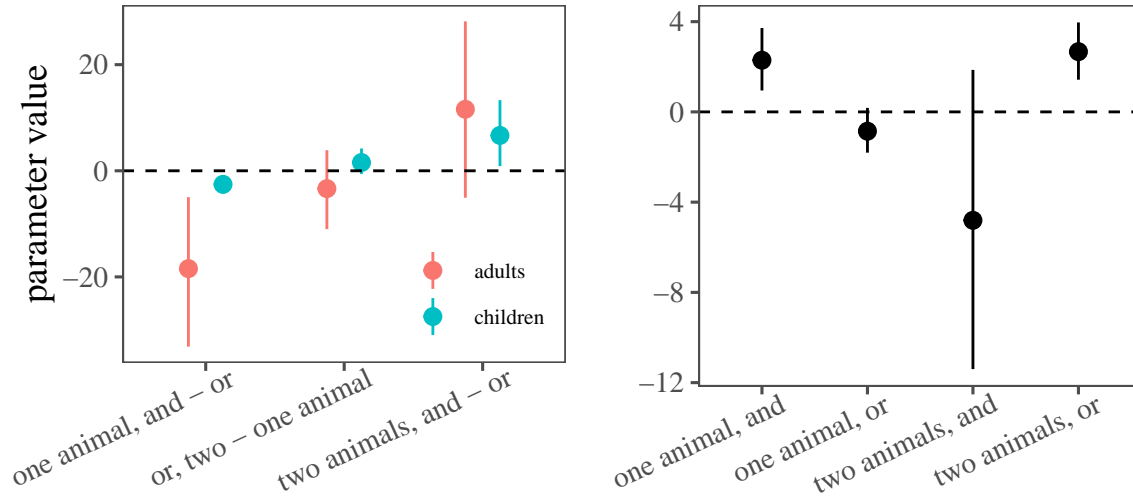


Figure 6. Left: The mean and 90% highest posterior density interval for the coefficients estimated in separate ordinal logistic regressions for adults and children, Right: Mean and 90% highest posterior density interval for the interaction coefficients in the combined dataset.

suggests that when only one animal was on the card, both adults and children had lower judgments for guesses with conjunction compared to guesses with disjunction. The third contrast suggests that when two animals were on the card, conjunction received higher judgments. For the second contrast, adults show negative values while children show positive values. This suggests that when the guess was a disjunction, children had slightly higher ratings when both animals were on the card compared to only one; adults on the other hand had slightly lower ratings and were more likely to choose “kinda right” over “right”. This is consistent with the hypothesis that children “compute exclusivity implicatures” at a lower rate than adults and are more likely to consider a disjunction acceptable when both disjuncts are true (Barner et al., 2011; Noveck, 2001; Papafragou & Musolino, 2003). However, we will see that children’s spontaneous feedback does not support this hypothesis.

To understand the extent to which children’s judgments differed from adults’ we will look at the mean and the 90% HPDIs of the interaction coefficients from the combined model shown in the right panel of Figure 6. We use this model to focus on the likely reliable differences between children and adults. Based on the 90% HPDIs, we can infer that

judgments of adults and children differed in two ways. First, when the guess was a disjunction (e.g. *cat or dog*) and two animals were on the card (e.g. CAT+DOG), children gave higher ratings than adults. This is again related to the issue of exclusivity implicatures discussed before. Second, when the guess was a conjunction (e.g. *cat and dog*) but only one animal was on the card (e.g. CAT), children gave slightly higher ratings than adults did. This could be either due to children’s judgments being more lenient or their judgments focusing more on matching labels to pictures rather than the connective. We will see more evidence for the label matching hypothesis in Experiment 3. <–MIKE

Spontaneous open-ended feedback. We also categorized and annotated children’s spontaneous and free-form verbal reactions to the puppet’s guesses. Table 6 summarizes the definitions and examples for each category and Figure 7 shows the results. We should point out that each trial type had a similar number of “None” cases. Some children remained silent throughout the experiment and only provided rewards to the puppet. In Experiment 3, we explicitly asked children to provide feedback and therefore, had no “None” response category. In the discussion and analysis here we will not comment further on the “None” category but focus on the other three categories.

Starting with control trials (the leftmost column), when the guessed animal was not on the card (e.g. *elephant*), children either provided judgments like “No!” or described what was on the card: “cat” or “cat and dog”. However, when the guessed animal (e.g. *cat*) was the only animal on the card (e.g. CAT), most children provided a positive judgment like “Yes”. When the animal guessed (e.g. *cat*) was only one of the animals on the card (e.g. CAT+DOG), children described what was on the card, for example “cat and dog”.

In trials with conjunction (e.g. *cat and dog*), children showed a high rate of corrections and descriptions when there was only one animal on the card (e.g. CAT). In their corrections, children used the focus particles *just* and *only* as in “just a cat” or “only a cat”. When both animals were on the card (e.g. CAT+DOG) and a conjunction was used (e.g. *cat and dog*),

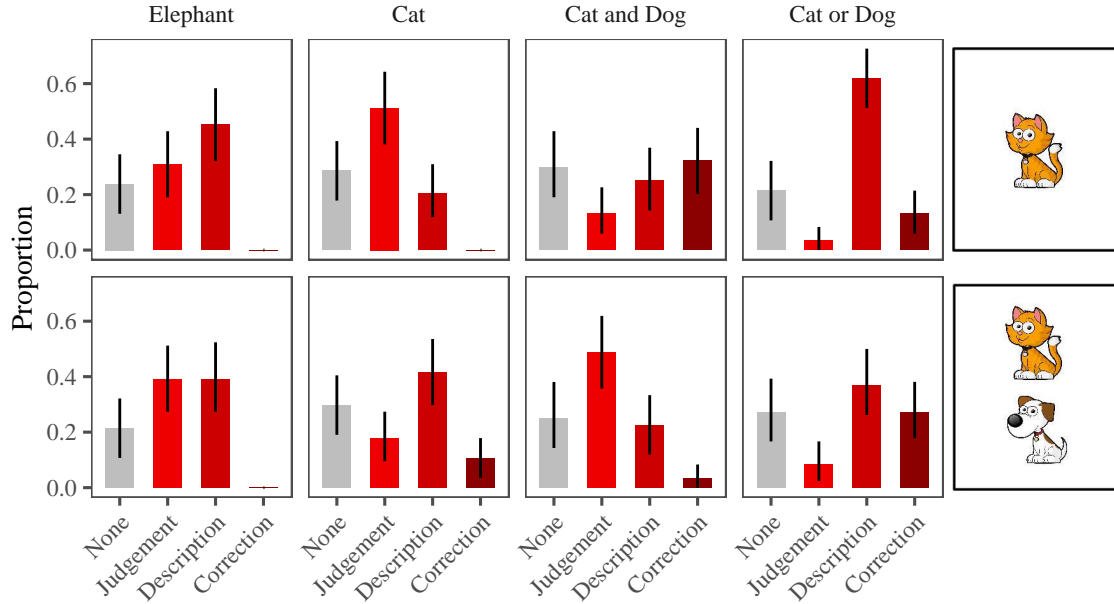


Figure 7. Children’s open-ended Feedback in Experiment 2.

children predominantly provided positive judgments like “Yes!” and “You are right”. With disjunction (e.g. *cat or dog*), when only one of the animals was on the card (e.g. CAT), most children simply described what was on the card, saying for example “cat”. However, when both animals were on the card (e.g. CAT+DOG), children corrected the puppet by saying “Both!” or emphasizing *and* as in “cat AND dog!”.

MIKE—>To quantify and compare the distribution of children’s feedback in different trial-types, we used a Bayesian mixed-effects multinomial regression model with the fixed effect of trial-type as well as random intercepts and slopes for participants and items (cards). The dependent measure was children’s feedback categories of judgment, description, and correction, with correction set as the reference category. The trial types “one animal, *or*”, “one animal, *and*”, and “two animals, *and*” constituted the (dummy-coded) fixed effects of the model with “two animals, *or*” set as the intercept. Priors and convergence information were identical to those reported for our previous models.

Figure 8 shows the mean and 90% HPDIs of the multinomial model co-efficients, with

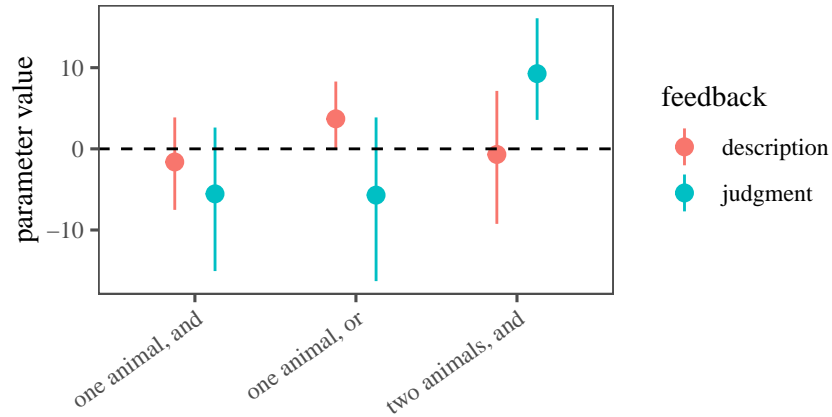


Figure 8. The mean and 90% highest posterior density interval of the coefficients of interest in experiment 2’s mixed-effects multinomial logistic model on children’s feedback.

the x-axis separating trial types. Starting from the left, the HPDIs for judgments and descriptions over corrections for the “one animal, and” trial-type included zero. This suggests that the feedback distribution was similar in conjunction trials with one animal and disjunction trials with two animals. Both trial types received a relatively high number of corrections. Considering the “one animal, *or*” trials, the HPDIs for judgments and descriptions over corrections included zero but in the case of descriptions, only very marginally so. As we will see in the next experiment’s replication of children’s feedback, children provide more descriptions than corrections when only one of the disjuncts is true. Finally with “two animal, *and*” trials, the HPDI for descriptions over corrections covers zero while that of judgments over corrections stays above zero. This suggests that when a conjunction was used and both animals were on the card, children provided more affirmative judgments than corrections. Overall, our statistical modeling suggests that patterns of children’s feedback were similar for conjunction trials when only one proposition was true and disjunction trials when both propositions were true. In such trial-types children provided higher rates of corrections than other trial-types.<—MIKE

Discussion

In study 2, we used a 3AFC judgment task to test children's comprehension of logical connectives *and* and *or*. We compared these results to those found in the 3AFC judgment task of study 1 with adults. The general comparison showed that adults and children had similar patterns of judgments, except when both disjuncts were true. In such cases, adults judged the disjunctive guess as not completely right while most children judged it as completely right. There was even a slight preference among children to reward the puppet more in such cases, compared to cases of disjunction when only one disjunct was true.

To consider another measure of children's comprehension, we also looked at children's spontaneous open-ended verbal feedback to the puppet's guesses. Our analyses suggested that children recognized false and infelicitous utterances with the connectives and provided appropriate corrective feedback. As expected from an adult-like understanding of connectives, children corrected the puppet most often when there was only one animal on the card and the guess was conjunctive, or when there were two animals on the card and the guess was disjunctive. Perhaps the most important finding was that children increased their corrective feedback in disjunctive guesses where both disjuncts were true, compared to those with only one true disjunct. These findings differ from the results of the 3AFC judgment task, which suggested that children did not find any infelicity with disjunctive guesses when both disjuncts were true.

The analysis of children's open-ended feedback raises two important issues. First, it runs counter to what the 3AFC judgment task suggests with respect to exclusivity implicatures. The forced-choice task suggests that children find such underinformative utterances as unproblematic while analysis of their spontaneous feedback shows that they provided more corrections to such utterances. Second, one of the explanations for why children fail to derive implicatures is that they cannot access the stronger alternative to the

disjunction word *or*, namely *and* (Barner et al., 2011). However, in the context of the guessing game, some children explicitly mentioned the word *and*, as the word the puppet should have said instead of *or*. Interestingly, these children continued to reward the puppet and considered the guess “right”, even though they corrected him. This raises the possibility that forced-choice truth value judgments underestimate children’s pragmatic knowledge. In study 3, we used both a 2AFC truth judgment task and an analysis of children’s open-ended feedback. If the findings of study 2 were on the right track, we expected to replicate the same pattern in study 3, and find that children’s open-ended feedback better reflects their sensitivity to pragmatic violations than the results of the 2AFC judgments.

Experiment 3: Children’s 2AFC judgments and open-ended feedback

This study used the same paradigm as Experiment 2 but measured children’s judgments using a binary forced choice task. Similar to Experiment 2, children’s open-ended feedback was also analyzed. The main hypothesis was that preschool children provide corrective feedback to the puppet if both disjuncts are true, but they do not consider this infelicity to be grave enough to render the guess “wrong” in a 2AFC judgment task. The main hypothesis along with relevant analyses and predictions were preregistered in an “As Predicted” format⁷.

Methods

Participants. We recruited 50 English speaking children from the Bing Nursery School at Stanford University. Children were between 3;6 and 5;9 years old (Mean = 4;7).

Materials. Experiment 3 was similar to Experiment 2 but differed in how children provided their judgments. Based on the findings in Experiment 2, we first focused on verbal

⁷The As Predicted PDF document is accessible at <https://aspredicted.org/x9ez2.pdf>.

feedback, instead of forced-choice responses. We used two different ways of measuring children’s judgments. First, we encouraged children to provide verbal feedback to the puppet. They were asked to say “yes” when the puppet was right and “no” when he was not right. Importantly, they were also asked to help him say it better. In each trial, after children were done with this initial open-ended feedback, we asked the classic truth value judgment forced choice question: “Was Jazzy (the puppet) right?”. This question elicited a 2AFC response for each trial independent of children’s earlier open-ended response. These two measures allowed us to compare open-ended and binary forced-choice judgments in the same paradigm and for the same trials.

Procedure. The setup and procedure were similar to Experiment 2, except there were no rewards. As in previous studies, participants sat through three phases: introduction, instruction, and test. The introduction phase made sure children knew the names of the animals on the cards. In the instruction phase, they received four training trials, as shown in Table 5 in the Appendix section.

As in Experiment 2, the experimenter put a sleeping mask over the puppet’s eyes and explained that Jazzy (the puppet) was going to guess what animal was on the cards. He then picked the first card and asked the puppet: “What do you think is on this card?” The puppet replied with “There is a dog”. The experimenter showed the cat-card to the child and said: when Jazzy is “not right”, tell him “no”. He then asked the child to say “no” to the puppet. The second trial followed the same pattern except that the puppet guessed “right” and the experimenter invited the child to say “yes” to the puppet. There were two more instruction trials before the test phase began. The test phase contained 16 randomized trials, half of which contained guesses with the words *and* and *or*⁸.

⁸The randomization code as well as the details of the methods are available on this paper’s online repository.

Results

We first look at the results of the 2AFC judgement task for each trial type and compare them to those of the adults' in Experiment 1. Then we analyze children's open-ended responses and compare them to the forced choice responses obtained in the same trial types. For the 2AFC judgments we excluded 26 trials (out of total 800) where children either did not provide a Yes/No response or provided both (i.e. "Yes and No"). The exclusions were almost equally distributed among different types of guesses and cards. In the analysis of children's open-ended feedback, we excluded 8 trials (out of total 800) where children either did not provide any feedback or their feedback could not be categorized into the existing categories.

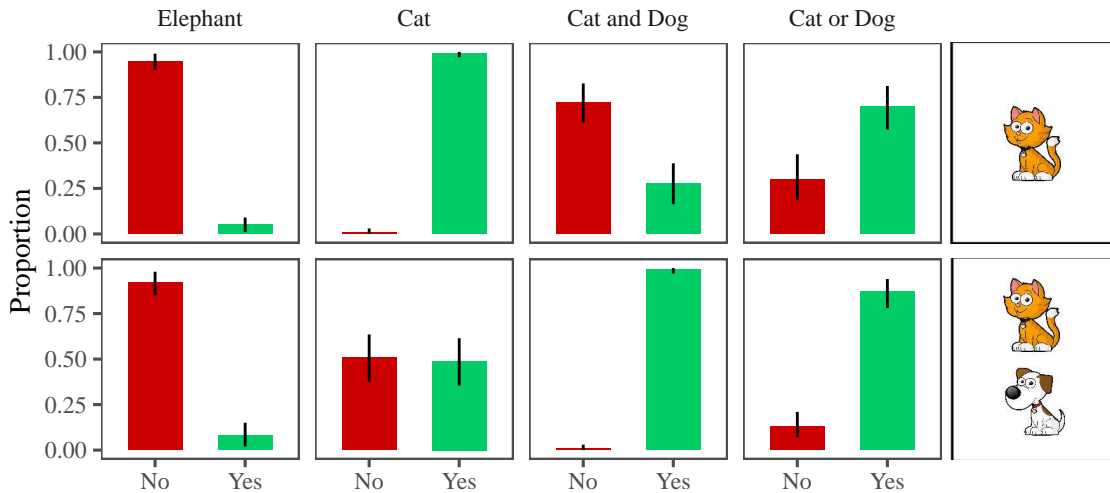


Figure 9. Children's two-alternative forced-choice judgments in Experiment 3.

Two-Alternative Forced Choice Judgments. Figure 9 shows children's 2AFC judgments. Starting with control trials, when the guessed animal (e.g. *elephant*) was not on the card, children considered the guess "wrong". When the guessed animal (e.g. *cat*) was the only animal on the card (e.g. CAT), children considered the guess "right". However, if it was only one of the animals on the card (e.g. CAT+DOG), children were equally split between "wrong" and "right" judgments. This is in contrast to adults who unanimously considered such guesses as "right" in their 2AFC judgments (Figure 2). There are two possible

explanations for this difference. First, some children may interpret a simple guess like “there is a cat” exhaustively as “there is **only** a cat”. Second, some children may consider leaving out an animal as a grave violation even though they do not interpret the guess as there is **only one** animal on the card. The first explanation is unexpected for a theory of acquisition that assumes children are overall more logical or literal as interpreters than adults (Noveck, 2001).

In conjunction trials, when the card had two animals (e.g. CAT+DOG) and both were guessed (e.g. *cat and dog*) children unanimously considered the guess as “right”. When the card (e.g. CAT) had only one of the guessed animals (e.g. *cat and dog*), children were more likely to consider the guess “wrong”. This is similar to adults’ binary judgments, but different in extent: adults were more consistent and unanimous in rejecting such guesses. In disjunction trials, when the card (e.g. CAT) had only one of the guessed animals (e.g. *cat or dog*), most children considered the guess “right”. This is again similar to adults but differs from them in extent: adults more consistently and unanimously judged such guesses as “right”. When the card had two animals (e.g. CAT+DOG) and both were guessed using a disjunction (e.g. *cat or dog*), children considered the guess right.

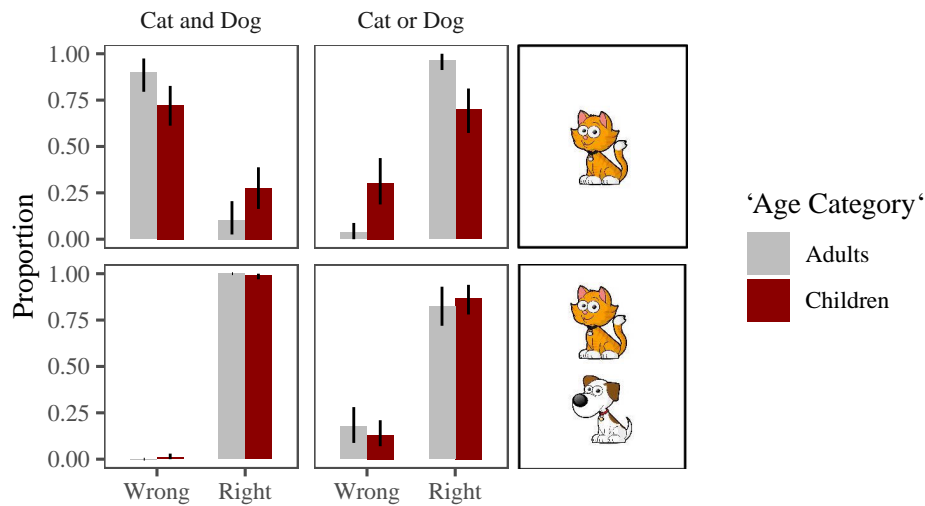


Figure 10. The comparison of 2AFC judgment tasks for conjunction and disjunction trials in adults (Experiment 1) and children (Experiment 3).

Figure 10 provides a side-by-side comparison of adults' and children's 2AFC judgments for conjunction and disjunction trials. The judgments are very close and differ only very slightly in trial types where there is only one animal on the card. To quantify trial-type differences in adults and children, we fit separate Bayesian mixed-effects binomial logistic regressions for each group. To capture differences between adults and children, we fit a similar model to the combined dataset of adults and children and added age category (adult vs. child) as an interaction term. As before, the model included the fixed dummy coded effect of trial-type (Levels: "one animal, *or*" (reference) - "two animals, *and*" - "one animal, *and*" - "two animals, *or*"). As before, the models had random intercepts and slopes for participants. Details of priors and convergence were similar to previous models as well.

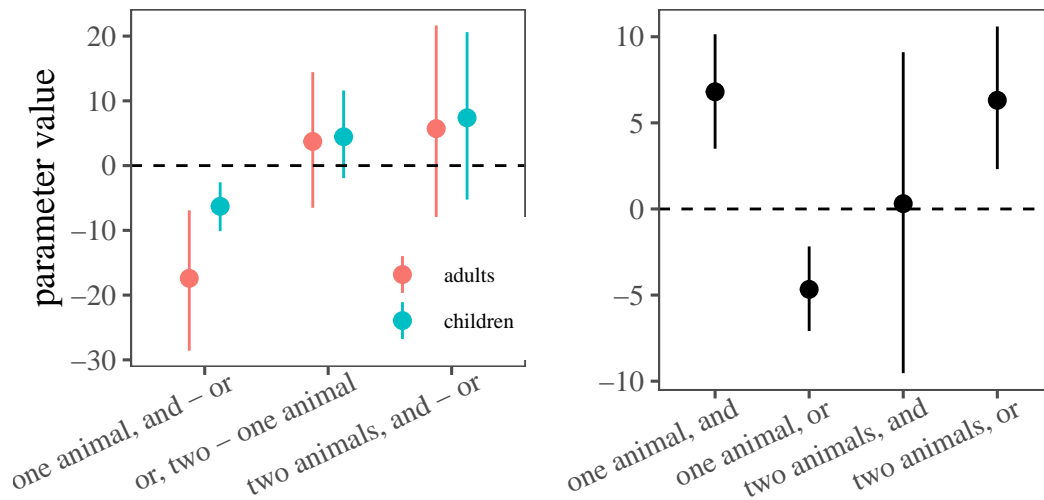


Figure 11. Left: Mean and 90% highest posterior density interval for parameter values estimated in separate ordinal logistic regressions for adults and children, Right: Mean and 90% highest posterior density interval for the interactive effect of age category (adult vs. child) on 3AFC judgments.

The left panel of Figure 11 shows the mean and 90% HPDI for three contrasts of interest estimated from separate binomial models for adults and children. The x-axis shows the contrasts between the following trial-types: conjunction vs. disjunction trials when only one animal was on the card (one animal, *and-or*), trials with one vs. two animals on the card

when the guess was a disjunction (*or*, two-one animal), and conjunction vs. disjunction trials when two animals were on the card (two animals, *and-or*). Similar to the results reported in Experiment 2, adults' and children's estimated coefficients are similar in sign to one another, though adults' are more extreme in the first contrast. The first contrast (one animal, *and-or*) suggests that when only one animal was on the card, both adults and children had lower judgments for guesses with conjunction compared to guesses with disjunction. The second and third contrasts suggest that on average, children and adults did not differentiate between disjunction trial-types with one vs. two animals or trial-types with two animals when the guess was a disjunction vs. a conjunction. The majority considered all such trials as "right" in the binary truth value judgment task.

To understand the extent to which children's judgments differed from adults' we looked at the mean and the 90% HPDIs of the interaction coefficients from the combined binomial model shown in the right panel of Figure 11. We used this model to focus on the likely reliable differences between children and adults, even though they may be small. Based on the 90% HPDIs, we can infer that judgments of adults and children differed in three ways. First, when the guess was a conjunction and only one animal was on the card, children were slightly more likely than adults to choose "right". Second, when the guess was a disjunction and only one animal was on the card, children were slightly more likely to choose "wrong". Notice that these two differences between children and adults are compatible with the label-matching account (Paris, 1973) but not the non-adult-like pragmatic enrichment account (Singh2016, Tieu et al., 2016). Under the pragmatic enrichment account, children should be more likely to say "wrong" to a disjunction with one disjunct true, but not more likely to say "right" to a conjunction when only one conjunct is true. However, the label-matching account predicts both these outcomes, because it posits that in both cases children disregard (or don't pay attention to) the connective and judge based on the match between labels and objects, which is partial in both cases.

Finally the third difference; our combined model suggested that children were slightly more likely than adults to judge the guess “right” when it was a disjunction and two animals were on the card. This is consistent with the hypothesis that children compute exclusivity implicatures at a lower rate than adults. However, as we will see in the next section, this hypothesis is undermined by the data from children’s feedback, which point to children’s sensitivity to the infelicity of disjunction when both disjuncts are true.

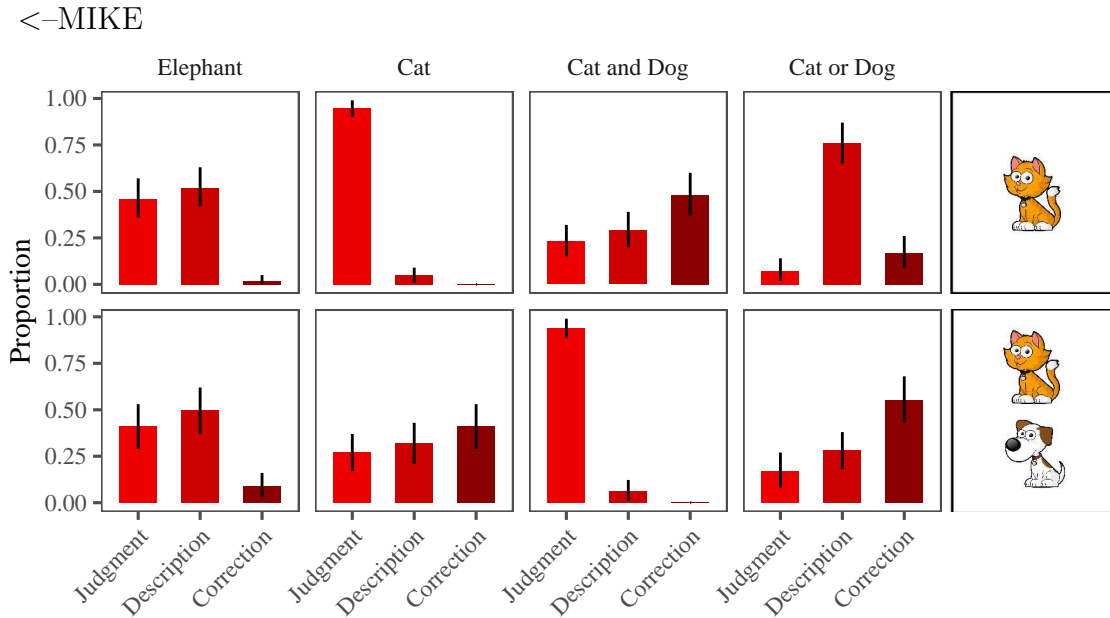


Figure 12. Children’s Open-ended Feedback in Experiment 3.

Open-ended Feedback. Figure 12 shows the distribution of children’s feedback to the puppet in Experiment 3 (see Table 6 for the definitions and examples of feedback categories). Similar to Experiment 2, Children’s feedback showed four main patterns. First when the puppet guessed an animal not on the card (e.g. *elephant*), there was a split pattern between negative judgments like “No!” and simply mentioning the animal on the card (e.g. “Cat!”). Second, almost all children responded with positive judgments like “Yes!” when the puppet’s guess accurately matched what was on the card. This was the case in trials where there was only one animal on the card and the puppet mentioned it, as well as trials where there were two animals on the card and the puppet mentioned both with a conjunction.

Third, children provided the largest number of corrective feedback in trials where the guess was either false or infelicitous. These included three trial types: (a) two animals on the card (e.g. CAT+DOG) but the puppet only guessed one (e.g. *cat*); (b) the puppet guessed two animals with conjunction (e.g. *cat and dog*) but only one of them was on the card (e.g. CAT); and (c) two animals on the card (e.g. CAT+DOG), and the puppet guessed both but used a disjunction (e.g. *cat or dog*). Finally, there was a pattern of feedback unique to disjunction trials with only one animal on the card. In such cases, almost all children simply named the animal on the card (e.g. “cat!”).

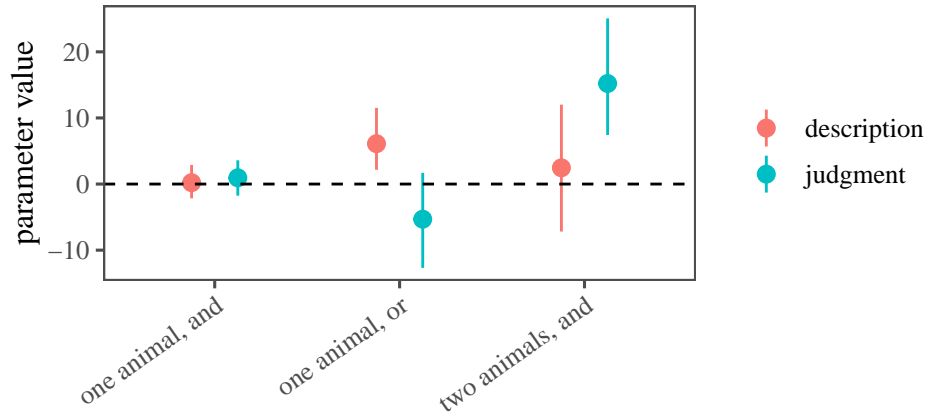


Figure 13. The mean and 90% highest posterior density interval of the coefficients of interest in experiment 3's mixed-effects multinomial logistic model on children's feedback.

MIKE—>To quantify and compare the distribution of children's feedback in connective trial-types, we used a Bayesian mixed-effects multinomial regression model with the fixed effect of trial-type as well as random intercepts and slopes for participants and items (cards). Similar to our analysis in Experiment 2, the dependent measure was children's feedback categories of judgment, description, and correction, with correction set as the reference category. The trial-types “one animal, *or*”, “one animal, *and*”, “two animals, *and*” constituted the (dummy-coded) fixed effects of the model with “two animals, *or*” set as the intercept. Priors and convergence information were identical to those reported for our previous models.

Figure 13 shows the mean and 90% HPDIs of the multinomial model co-efficients. These results replicate the findings on children’s feedback reported in Experiment 2. Starting from the left, the HPDIs for judgments and descriptions over corrections for the “one animal, *and*” trial-type included zero. This suggests that the feedback distribution was similar in conjunction trials with one animal and disjunction trials with two animals. Considering the “one animal, *or*” trials, the HPDI for judgments over corrections included zero but not the HPDI for descriptions. This suggests that children provided more descriptions than corrections when only one of the disjuncts was true. Finally with “two animal, *and*” trials, the HPDI for descriptions over corrections includes zero while that of judgments over corrections stays above zero. This suggests that when a conjunction was used and both animals were on the card, children provided more affirmative judgments than corrections. Overall, the results confirm the findings reported in Experiment 2: in connective trials, children were more likely to provide corrections in trial-types that were either false or infelicitous. <—MIKE

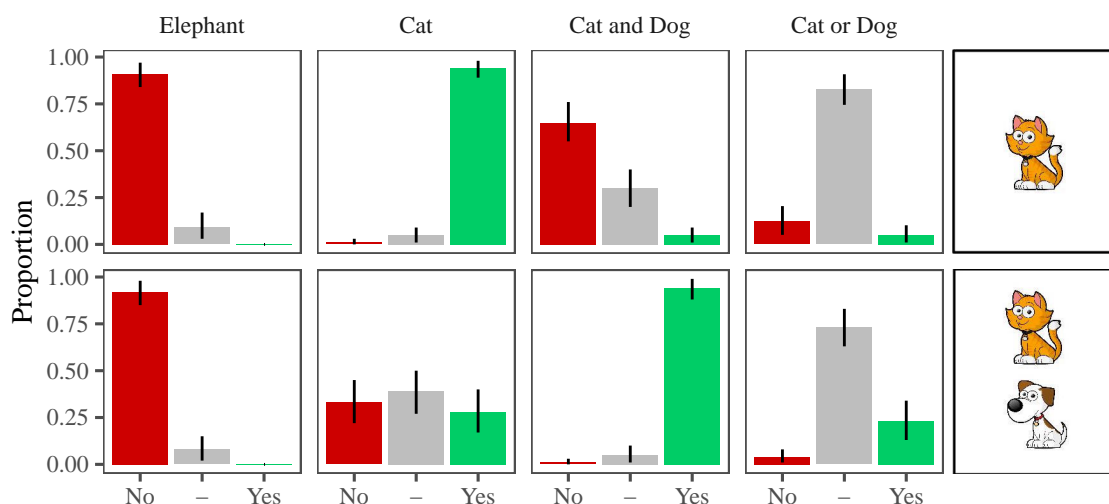


Figure 14. Children’s open-ended feedback in Experiment 3. The x-axis shows whether children spontaneously provided a yes (green), no (red), or other response (grey).

In Figure 14 we break down children’s open-ended feedback based on whether they spontaneously provided a simple judgment of *Yes!* and *No!*, or not. Responses that were not

yes/no judgments are grouped in a middle category shown with a dash. The goal here is to compare children's open-ended and spontaneous binary judgments with their forced-choice binary judgments shown in Figure 9. Notice that both were provided in the same task and essentially one right after the other (forced-choice after spontaneous). Children's open-ended judgments and their forced choice judgments in study 3 show similar patterns for all types of guesses except for disjunctive ones. In trials that the puppet guessed a disjunction, the vast majority of children refused to provide a yes/no judgment when they were not forced to. Instead, they described the animal on the card or provided corrections to the puppet's infelicitous disjunctive guess.

One way to interpret these results is that disjunctive guesses (with at least one disjunct true) are considered neither right nor wrong. When children were forced to provide wrong/right responses in the experimental context, some conformed to the adult patterns of judgment and some did not. However, it is possible that such deviations from adult judgments do not reflect differences in the comprehension of disjunction, but rather differences in how children map their comprehension of disjunction onto the notions of "right" and "wrong" when forced to do so.

Figure 15 shows the proportion of feedback categories other than yes/no judgments on the x-axis. Our goal here is to display the trial types with corrective feedback (blue and red). These trial types include: (1) conjunction when only one conjunct is true (e.g. guess: *cat and a dog*, card: CAT), (2) disjunction when both disjuncts are true (e.g. guess: *cat or dog*, card: CAT+DOG), and (3) simple guesses when two animals were on the card (e.g. guess: *cat*, CAT+DOG). As explained before, these trial types involved guesses that were either false or infelicitous. Furthermore, the type of corrective feedback children provided matched the type of mistakes made in the guesses. With conjunctive guesses (e.g. *cat and a dog*) when there was only one animal on the card (e.g. CAT), children provided exclusive corrections such as "just a cat" or "only a cat!", suggesting that the other animal in the guess (e.g. *dog*) should

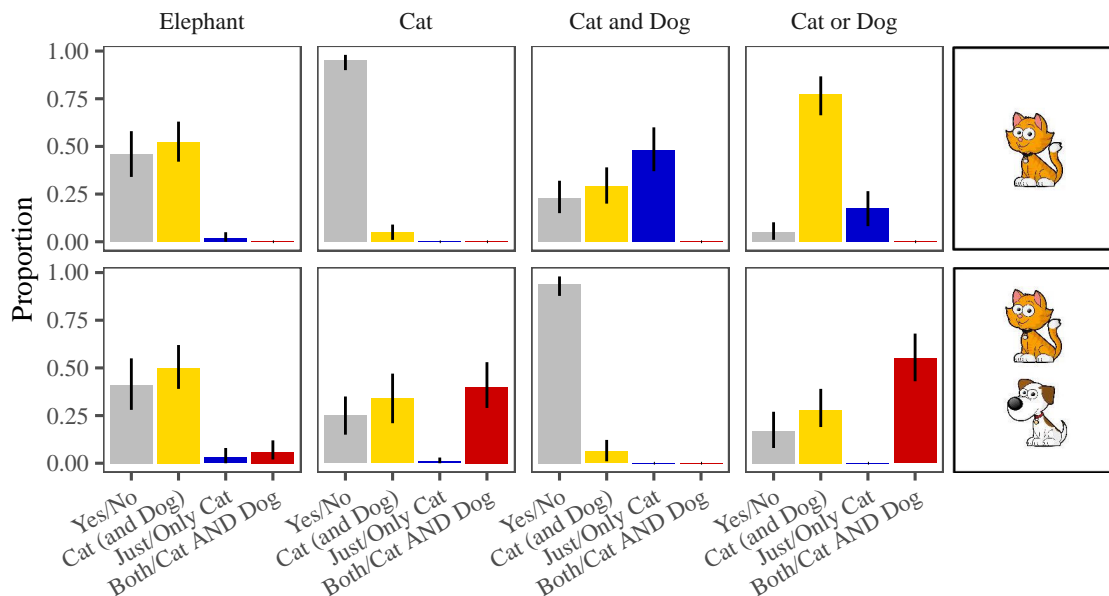


Figure 15. Children's open-ended feedback in different trial types of Experiment 3.

have been excluded. When two animals were on the card (e.g. CAT+DOG) and the puppet used a disjunctive guess (e.g. *cat or dog*), or a simple guess (e.g. *cat*), children provided inclusive feedback such as “cat AND dog” or “both”, suggesting that another animal should have been included. This is particularly notable in the case of disjunction since both animals were mentioned, but children still emphasized that the connective *and* should have been used, or that *both* animals mentioned were actually on the card. Such corrective comments hint at a deeper understanding of differences between the meaning of disjunction and conjunction.

Discussion

Experiment 3 measured children's comprehension of logical connectives in two ways: First, with analyzing their open-ended feedback and second, with a two-alternative forced choice task. First, we asked children to say *yes* to the puppet if he was right and *no* if he was wrong. However, children could provide any form of feedback they wanted. Second, we followed children's open-ended feedback with a 2AFC question: “Was the puppet right?” This way, we could measure children's comprehension in two different ways in the same trial.

Ideally, both measures should show similar results. However, the findings were similar for conjunctive guesses, but not disjunctive ones. Children avoided binary right/wrong feedback with disjunction and preferred to provide more nuanced feedback.

The 2AFC responses followed the predicted pattern: conjunctive guesses were judged wrong if only one conjunct was true, and right if both were true. Disjunctive guesses were judged right whether one or both disjuncts were true. There was no significant difference in the 2AFC task between the responses of children and those of adults in Experiment 1. Children's open-ended feedback in Experiment 3 replicated the findings of Experiment 2. Children provided more corrective feedback in false and infelicitous trials than in true and felicitous ones. The corrective feedback was tailored to the puppet's mistake. If the puppet used a conjunction when there was only one animal on the card, children pointed out that the other animal should have been excluded from the guess. They used the exclusive adverbials *just* and *only* in their feedback. If the puppet used a disjunction when both animals were on the card, children stressed *and* or *both*, implying that both animals should have been included.

While the 2AFC results suggested that children took no issue with disjunctive guesses when both disjuncts were true, the analysis of their corrective feedback showed that they provide appropriate corrections in such cases and emphasized that the connective *and* would have been a better guess. Taking both measures together, we conclude that even though children are aware of the problem with such guesses, they do not consider them *wrong*.

General Discussion

We reported three studies on adults' and preschool children's comprehension of disjunction. The first study used two- and three-alternative forced choice judgment tasks with adults. In the 2AFC task, adult interpretations closely matched the semantic accounts

of *and* and *or* as conjunction and inclusive disjunction. The 2AFC judgments did not register robust signs of pragmatic infelicities. However, the 3AFC judgments showed signs of pragmatic infelicities, especially in disjunctive guesses with true disjuncts. When two animals where on the card (e.g. cat and dog) and the guess used *or* (e.g. *There is a cat or a dog!*), participants were more likely to choose “kinda right” rather than “right”.

The second study used a 3AFC judgment task with four-year-old children. It also included an exploratory analysis of children’s open-ended verbal feedback to the puppet in the experimental setting. Children’s interpretations were similar to those of adults in the 3AFC task and only differed for pragmatically infelicitous disjunctions. When both disjuncts were true, adults tended to judge disjunctive guesses as “kinda right”. This was evidence for the pragmatic infelicity of such guesses. While, children judged such disjunctive statement as “right”, the analysis of their open-ended feedback showed that they took issue with such statements as well, and provided appropriate corrective feedback.

In the third study, we focused on eliciting open-ended verbal feedback from children and followed it with a 2AFC question. Children’s 2AFC responses reflected the semantics of the connectives *and* and *or* as conjunction and inclusive disjunction. There was no significant difference between children and adults in the 2AFC task. Analysis of children’s open-ended feedback replicated the findings in study 2. Children provided more corrective feedback in false and pragmatically infelicitous trials with the connectives than in felicitous trials. The comparison of the 2AFC task and children’s open-ended responses showed that children are sensitive to the infelicity of disjunctions with true disjuncts, even though they consider them to be “right” guesses.

Previous studies had suggested that adults and preschool children differ in their interpretation of disjunction in two ways. First, unlike adults, children might interpret a disjunction as conjunction (Singh et al., 2016; Tieu et al., 2016). Second, children might interpret *or* as inclusive disjunction when adults interpret it as exclusive (Crain, 2012). The

studies reported here provide evidence for the hypothesis that these differences may be an artifact of the experimental task and the type of measurement (Skordos et al. (n.d.), Katsos (2014)).

Considering the first difference, in the 2AFC and 3AFC judgment tasks we found only small (but significant) preferences for both disjuncts being true rather than only one. Combining the 2AFC and the verbal feedback results, we expect that a child with strong conjunctive interpretation of disjunction should have rejected a disjunctive guess when only one disjunct was true, provided a “Just/Only” feedback, and accepted the guess when both disjuncts were true without providing a correction. We found no child in our sample that showed this pattern of responses. Two children who consistently rejected a disjunction when only one disjunct was true, provided corrective feedback when one or both disjuncts were true. Therefore while it is possible that some children interpret *or* as *and*, our results did not show a common or consistent effect.

We would like to add that conjunctive interpretations of disjunction, even when robustly observed, can have at least two potential explanations. First, non-linguistic interpretive strategies and preferences, due to task demands or unknown connective meaning (Clark, 1973; Paris, 1973), and second, non-adult-like pragmatic enrichment, hypothesized to occur in free-choice contexts in adult language (Singh et al., 2016; Tieu et al., 2016). As explained in section , previous research provides substantial evidence for task-related increase in conjunctive readings of disjunction (Braine & Romain, 1981; Neimark & Slotnick, 1970; Paris, 1973; Skordos et al., n.d.). In order to show instances of pragmatically enriched conjunctive readings in preschool children, it is important to first rule out task-related conjunctive interpretations.

Considering the second difference, namely the lower rate of exclusivity inferences in preschool children, our studies provided evidence that the choice of measurement may play an important role. In the 3AFC judgment task when two animals were on the card

(e.g. card: cat and dog, guess: “There is a cat or a dog”), adults were more likely to choose “kinda right” than children were. Children mostly chose “right”. However, in their free-form feedback, children corrected such utterances and suggested that the connective *and* should have been used instead of *or*.

There have been at least four major proposals to account for children’s perceived low rate of “implicature computation”: processing difficulty (Pouscoulous, Noveck, Politzer, & Bastide, 2007; Reinhart, 2004), non-adult-like lexical entry (Barner et al., 2011; Horowitz, Schneider, & Frank, 2017), pragmatic tolerance (Katsos & Bishop, 2011), and the role of experimental measurement (Katsos, 2014). Below we argue that the first three cannot explain the reported results of children’s forced judgments and free-form feedback, and that these results highlight the role of experimental measurement as a source of perceived differences in children and adults pragmatic inferences.

1. Processing difficulty. First, processing accounts locate the problem in children’s processing capacities such as working memory. They suggest that pragmatic computations are cognitively taxing and children lack the appropriate processing resources to carry them out appropriately. A prediction of processing accounts (at least in their current format) is that children will show reduced implicature computations for all types of implicatures – scalar or not. This prediction was not borne out in our experimental results here. In Experiment 3, children were much more likely than adults to call a simple guess (e.g. *There is a cat!*) “wrong” if there were two animals on the card (e.g. cat and dog). In other words, children’s interpretations were much more exhaustive than adults. Processing accounts do not predict that children may derive implicatures at a higher rate than adults but this is what we found, at least for the traditional interpretation of the judgment task.

2. Non-adult-like Lexicon. Several proposals blame the structure of the child’s lexicon for the alleged failure in deriving implicatures. The assumption is that the child’s lexical entry for scalar items must include three elements for successful derivation: 1. the

922 semantics of the weak term (e.g. *some, or*) 2. the semantics of the strong term (e.g. *all, and*);
923 and possibly 3. a scale that recognizes the stronger term as an alternative to the weaker one
924 (e.g. $\langle \textit{some}, \textit{all} \rangle$, $\langle \textit{or}, \textit{and} \rangle$). Each of these elements have been pinpointed as the source
925 of the problem in previous studies (Barner et al., 2011; Horowitz et al., 2017; Katsos &
926 Bishop, 2011). However none of them seem to apply to the results reported here.

927 If children in this study lack the semantics of the connective *or*, we would expect them
928 to either perform at chance or default to a conjunctive interpretation. Neither prediction was
929 borne out in studies 2 and 3. Furthermore, children’s free-form linguistic feedback in both
930 studies suggested that children understood disjunction well enough to provide relevant
931 feedback. So this explanation seems unlikely. The problem cannot be that children do not
932 know the meaning of *and* either. Children’s performance in both study 2 and 3 for
933 conjunction trials show that they understand its meaning very well. Finally, comparing
934 children’s truth value judgments and their free-form verbal feedback, we found that many
935 children judged a disjunction with true disjuncts as “right”, yet went on to correct the
936 puppet and explicitly mention *and* as the connective he should have used. If children could
937 not access the stronger alternative, they could not have mentioned it in their feedback either.
938 And if accessing the stronger alternative would have resulted in expressing sub-optimal
939 judgments, they should not have judged the guess as “right”.

940 **3. Pragmatic Tolerance.** Katsos & Bishop (2011) suggested that children tend to
941 tolerate pragmatic infelicities more than adults. They showed that when children were
942 provided with a 2AFC judgment task, they considered a description with the scalar term
943 *some* as “right” when *all* was more informative (e.g. *The turtle played with some of the*
944 *balls.*, Scene: the turtle played with all the balls.) However, when they are presented with
945 three options (small, big, and huge strawberries) in a 3AFC task, they choose the middle
946 option in the same type of trials. They argued that children tolerate pragmatic infelicities
947 and do not regard them as “wrong”. As in a processing account, the tolerance account

predicts that scalar and ad-hoc implicatures will be similarly affected. However, our results did not match those of Katsos & Bishop (2011). When children were presented with a 3AFC task, they chose the highest reward (and not the middle option) for uses of *or* when *and* was more informative. Second, and more importantly, we found different patterns for exhaustive and scalar inferences as mentioned before. This is not predicted by the tolerance account unless we assume that children are more tolerant towards violations of scalar inferences than they are towards exhaustive ones. While this is not currently assumed in the literature, it is a possible adjustment. However, we would like address this issue by focusing on another related factor: the role of measurement in estimates of children’s pragmatic capacity (Katsos, 2014).

4. The Role of Measurement. Two observations in the current studies provide support for the hypothesis that methodological issues, and more specifically issues of measurement contribute to the differences found between adults and children in pragmatic capacity. First, Experiment 1 showed that even for adults, the estimates of adult infelicity rates may differ based on the number of alternatives in the forced choice task. A 2AFC task underestimated adults’ sensitivity to pragmatic infelicity. In fact, in a follow up study, we systematically varied the number of response options and replicated the results presented here (see Jasbi et al., 2019). Second, children’s open-ended linguistic feedback in the experimental context better reflected their sensitivity to pragmatic nuances than the forced-choice judgment tasks. Third, children showed a higher rate of “wrong” judgments for cases of exhaustive inferences (simple guesses with two animals on the card) than adults did. While a difference in sensitivity to ad-hoc vs. scalar implicatures has been reported and argued for before (Horowitz et al., 2017; Stiller, Goodman, & Frank, 2015), a higher sensitivity than adults is not predicted by any of the current accounts.

Figure 16 shows a summary of the factors that are proposed to affect pragmatic computations. As Pouscoulous & Noveck (2009) and Katsos (2014) have suggested, the

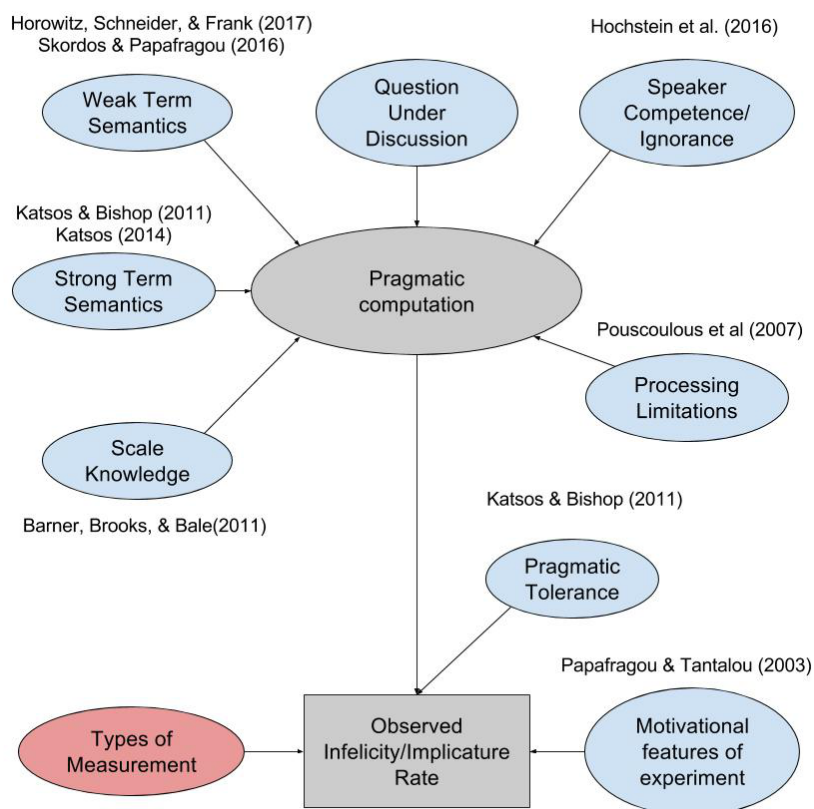


Figure 16. Factors that could affect pragmatic computations and the estimates of these computations in the experimental settings

central issue is “the rate” at which children and adults manifest pragmatic reasoning in the experimental setting. No one doubts children’s capacity to perform such computations. At issue is the extent to which children and adults compute specific implicatures. As Katsos (2014) pointed out, it seems reasonable to assume that all these factors play some part here. What matters is the degree to which each contributes to the outcome.

The results of the studies reported here suggest that it is important to distinguish between factors that affect pragmatic computations and those that affect the observed “rate” in an experimental setting. As we showed in Experiment 1, given the number of alternatives in the forced choice task (2AFC vs. 3AFC), we may get different estimates of adults’ rate of infelicity judgments, but we cannot assume that there is a difference in adults’ pragmatic

capacities in these two tasks. A similar situation exists when we compare children's forced choice measures of infelicity and their open-ended feedback.

In order to better understand the differences between adults and children's semantic and pragmatic capacities, it is necessary to have a good understanding of how our measurements affect estimates of adults and children's performance in the experimental tasks. Children may be no more capable of making exhaustive inferences than adults and no less capable of making scalar inferences either. They may simply have a different construal of the wrong-right scale and of what the forced-choice task is about. The concepts "right" and "wrong" are as much subject to developmental change and differences between adults and children as are scalar items that constitute the focus of our studies. Relying on a single type of measurement increases the risk of measurement-specific conclusions. Using multiple measurements in the same task can provide converging evidence for felicity/infelicity or presence/absence of specific inferences. Ultimately, in order to capture semantic and pragmatic competences of adults and (especially) children, we need to develop methods that can reliably tap into specific dimensions of meaning.

Conclusion

We provided three studies that tested adults and children's comprehension of disjunction in existential sentences using three different measures: binary forced-choice truth value judgments (2AFC task), ternary forced-choice truth value judgments (3AFC task), and free-form verbal feedback. The results suggested that for each population, different measures were sensitive to different aspects of meaning. The binary measure captured children and adults intuitions about truth values well: it showed that they considered a disjunction as inclusive in existential sentences of the guessing game. Ternary judgments provided evidence for adults' pragmatic inferences: adults often considered a disjunction when both disjuncts were true as "kinda right" and not completely right. For children, the ternary judgments did

1009 not register such an effect, but their free-form verbal feedback did. When both disjuncts
1010 were true, children verbally corrected the puppet and suggested that he should have said *and*
1011 instead of *or*. The combination of children's truth valued judgments and their verbal
1012 feedback suggested that on average, children in our sample understood that when both
1013 propositions were true, the conjunction and disjunction of them were also true, yet
1014 conjunction made a more appropriate and felicitous utterance.

1015 Since Tarski's original observations on disjunction, research in semantics and
1016 pragmatics has shown that the variety of interpretations Tarski observed are in fact distinct
1017 types of meaning observed in many aspects of language and connected to distinct processes
1018 that generate them. Therefore, while the inclusive interpretation is hypothesized to be part
1019 of *or*'s semantics, exclusivity and ignorance interpretations are analyzed as distinct
1020 pragmatic inferences generated separately. This theoretical insight has in turn led
1021 developmental researchers to seek distinct developmental mechanisms for each type of
1022 meaning. The results of the studies reported here suggest that as more and more varieties of
1023 meaning become subject of experimental studies, we also need to develop measures especially
1024 suited to capture the specific aspect of meaning under investigation.

References

1026

Supplementary Materials

Table 4

Instruction Trials.

Card	Guess	Reward
CAT	There is a dog!	Circle
ELEPHANT	There is an elephant!	Big Star
CAT-DOG	There is a dog!	Little Star

Table 5

Instruction Trials for Experiment 3.

Card	Guess	Response
CAT	there is a dog!	No!
ELEPHANT	there is an elephant!	Yes!
DOG-ELEPHANT	there is a cat!	No!
DOG	there is a dog!	Yes!

Table 6

Definitions and Examples for the Feedback Categories.

Category	Definition	Examples
None	no verbal feedback	
Judgment	provided verbal judgment mirroring the reward	“No!”, “Yes!” , “You are right!”
Description	mentioned the animal(s) on the card	“elephant”, “cat and dog”
Correction	used focus particles like <i>only/just</i> , emphasized <i>and</i> or used <i>both</i>	“only cat”, “just elephant”, “both!”, “cat AND dog!”

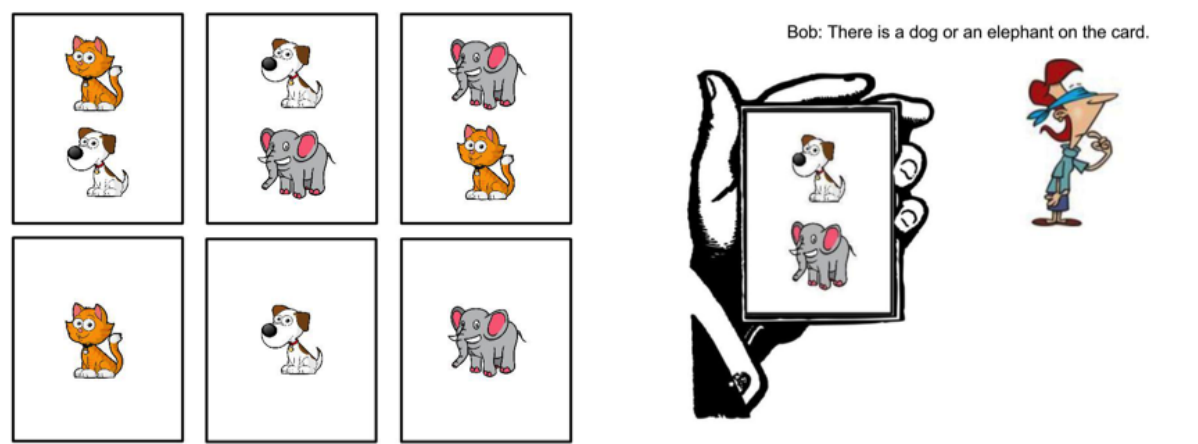
- 1027 Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar
1028 alternatives in children's pragmatic inference. *Cognition*, 118(1), 84–93.
- 1029 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
1030 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,
1031 68(3), 255–278.
- 1032 Braine, M. D., & Romain, B. (1981). Development of comprehension of “or”: Evidence for a
1033 sequence of competencies. *Journal of Experimental Child Psychology*, 31(1), 46–70.
- 1034 Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan.
1035 *Journal of Statistical Software*, 80(1), 1–28.
- 1036 Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition
1037 of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings*
1038 *of the 25th Boston University conference on language development* (pp. 157–168).
1039 Somerville, MA: Cascadilla Press.
- 1040 Chierchia, G., Crain, S., Guasti, M. T., & Thornton, R. (1998). “Some” and “or”: A study
1041 on the emergence of logical form. In *Proceedings of the Boston University conference*
1042 *on language development* (Vol. 22, pp. 97–108). Somerville, MA: Cascadilla Press.
- 1043 Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004).
1044 Semantic and pragmatic competence in children's and adults' comprehension of or. In
1045 I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 283–300). Basingstoke:
1046 Palgrave Macmillan.
- 1047 Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings.
1048 *Cognition*, 2(2), 161–182.
- 1049 Crain, S. (2008). The interpretation of disjunction in universal grammar. *Language and*

- 1050 *Speech*, 51(1-2), 151–169.
- 1051 Crain, S. (2012). *The emergence of meaning*. Cambridge: Cambridge University Press.
- 1052 Crain, S., Gualmini, A., & Meroni, L. (2000). The acquisition of logical words. *LOGOS and*
1053 *Language*, 1, 49–59.
- 1054 Goro, T., & Akiba, S. (2004). The acquisition of disjunction and positive polarity in
1055 Japanese. In *Proceedings of the 23rd West Coast conference on formal linguistics* (pp.
1056 251–264). Somerville, MA: Cascadilla Press.
- 1057 Grice, H. P. (1975). Logic and conversation. 1975, 41–58.
- 1058 Gualmini, A., & Crain, S. (2002). Why no child or adult must learn de Morgan’s laws. In
1059 *Proceedings of the Boston University conference on language development*. Somerville,
1060 MA: Cascadilla Press.
- 1061 Gualmini, A., Crain, S., & Meroni, L. (2000). Acquisition of disjunction in conditional
1062 sentences. In *Proceedings of the boston university conference on language development*.
- 1063 Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2017). The trouble with quantifiers:
1064 Exploring children’s deficits in scalar implicature. *Child Development*.
- 1065 Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to*
1066 *adolescence: An essay on the construction of formal operational structures* (Vol. 84).
1067 London: Routledge.
- 1068 Jasbi, M., Waldon, B., & Degen, J. (2019). Linking hypothesis and number of response
1069 options modulate inferred scalar implicature rate. *Frontiers in Psychology*, 10.
- 1070 Johansson, B. S., & Sjölin, B. (1975). Preschool children’s understanding of the coordinators
1071 “and” and “or”. *Journal of Experimental Child Psychology*, 19(2), 233–240.

- 1072 Katsos, N. (2014). Scalar implicature. In D. Matthews (Ed.), *Pragmatic development in first*
1073 *language acquisition* (Vol. 10, p. 183—198). Amsterdam: John Benjamins.
- 1074 Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of
1075 informativeness and implicature. *Cognition*, 120(1), 67–81.
- 1076 Neimark, E. D. (1970). Development of comprehension of logical connectives: Understanding
1077 of “or”. *Psychonomic Science*, 21(4), 217–219.
- 1078 Neimark, E. D., & Slotnick, N. S. (1970). Development of the understanding of logical
1079 connectives. *Journal of Educational Psychology*, 61(6p1), 451.
- 1080 Nitta, N., & Nagano, S. (1966). Basic logical operations and their verbal expressions: Child’s
1081 conception of logical sum and product. *Research Bulletin of the National Institute for*
1082 *Educational Research, Tokyo*, 7, 1–27.
- 1083 Notley, A., Thornton, R., & Crain, S. (2012a). English-speaking children’s interpretation of
1084 disjunction in the scope of “not every”. *Biolinguistics*, 6(1), 32–69.
- 1085 Notley, A., Zhou, P., Jensen, B., & Crain, S. (2012b). Children’s interpretation of
1086 disjunction in the scope of “before”: A comparison of English and Mandarin. *Journal*
1087 *of Child Language*, 39(03), 482–522.
- 1088 Noveck, I. A. (2001). When children are more logical than adults: Experimental
1089 investigations of scalar implicature. *Cognition*, 78(2), 165–188.
- 1090 Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the
1091 semantics–pragmatics interface. *Cognition*, 86(3), 253–282.
- 1092 Paris, S. G. (1973). Comprehension of language connectives and propositional logical
1093 relationships. *Journal of Experimental Child Psychology*, 16(2), 278–291.

- 1094 Pouscoulous, N., & Noveck, I. A. (2009). Going beyond semantics: The development of
1095 pragmatic enrichment. In S. Foster-Cohen (Ed.), *Language acquisition* (pp. 196–215).
1096 Berlin: Springer.
- 1097 Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental
1098 investigation of processing costs in implicature production. *Language Acquisition*,
1099 *14*(4), 347–375.
- 1100 Reinhart, T. (2004). The processing cost of reference set computation: Acquisition of stress
1101 shift and focus. *Language Acquisition*, *12*(2), 109–155.
- 1102 Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., & Fox, D. (2016). Children interpret
1103 disjunction as conjunction: Consequences for theories of implicature and child
1104 development. *Natural Language Semantics*, *24*(4), 305–352.
- 1105 Skordos, D., Feiman, R., Bale, A., & Barner, D. (n.d.). *Do children interpret “or”*
1106 *conjunctively?* Retrieved from <https://osf.io/2srxk/>
- 1107 Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool
1108 children. *Language Learning and Development*, *11*(2), 176–190.
- 1109 Su, Y. (2014). The acquisition of logical connectives in child Mandarin. *Language*
1110 *Acquisition*, *21*(2), 119–155.
- 1111 Su, Y., & Crain, S. (2013). Disjunction and universal quantification in child mandarin.
1112 *Language and Linguistics*, *14*(3), 599–631.
- 1113 Suppes, P., & Feldman, S. (1969). *Young children’s comprehension of logical connectives*.
1114 *ERIC*. Department of Health, Education, Welfare. Office of Education.
- 1115 Tarski, A. (1941). *Introduction to logic and to the methodology of the deductive sciences*.
1116 Oxford University Press.

- 1117 Tieu, L., Yatsushiro, K., Cremers, A., Romoli, J., Sauerland, U., & Chemla, E. (2016). On
1118 the role of alternatives in the acquisition of simple and complex disjunctions in french
1119 and japanese. *Journal of Semantics*.





elephant	cat	cat and dog	cat or dog	
simple control false	simple control true	<i>connective</i> control false	<i>connective</i> critical true	
simple control false	simple control true exhaustivity implicature	<i>connective</i> control true	<i>connective</i> critical true scalar implicature	

Figure 17. Cards used in the connective guessing game.



Figure 18. The puppet, Jazzy, with and without the sleeping mask.