Adults' and Children's Comprehension of Linguistic Disjunction

Masoud Jasbi[1] & Michael C. Frank[2]

[1] Harvard University

[2] Stanford University

Author Note

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

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

Abstract

Disjunction has played a major role in advancing theories of logic, language, and cognition, featuring as the centerpiece of debates on the origins and development of logical thought. Recent studies have argued that preschool children's comprehension of linguistic disjunction differs from adults in two ways. First, children are more likely to interpret *or* as *and* (conjunctive interpretations); Second, children are more likely to consider a disjunction as inclusive (lack of exclusivity implicatures). We first provide a comprehensive overview of previous developmental studies, showing that conjunctive interpretations are often due to task demands and participants' application of nonlinguistic strategies. We then provide the results of three studies on adults and preschool children, using binary and ternary forced-choice judgment tasks, as well as quantitative analysis of their spontaneous verbal feedback. We did not find evidence for conjunctive interpretations of disjunction in preschool children, supporting the hypothesis that conjunctive interpretations are due to task demands and application of nonlinguistic strategies. With respect to exclusivity implicatures, forced-choice judgment tasks suggested a slight tendency in children to accept a disjunction as inclusive more easily than adults. However, the quantitative analysis of children's spontaneous verbal feedback showed that children were sensitive to the exclusivity implicature of disjunction. More specifically, children explicitly mentioned the word *and* as a better alternative to *or* when both propositions were true, yet did not consider such an infelicitous disjunction "wrong". These results suggest that truth value judgment task alone can underestimate children's pragmatic competence, and should be accompanied by measures that are more sensitive to pragmatic inferences. Overall, our studies suggest that preschool children's understanding of logical connectives are much more adult-like than previously considered.

Keywords: conjunction, disjunction, implicatures, semantics, pragmatics, logic, language, language acquisition, language development

Word count: X

Adults' and Children's Comprehension of Linguistic 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)"

Linguists label the inference that the speaker does not know which sentence is true, as the IGNORANCE implication. Tarski (1941) also 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 a 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 IMPLICATURES (a term Grice coined for such implications) are inferences derived from our pragmatic reasoning on why the speaker used a disjunction like "A or B", instead of "A and B", or just "A". 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 argued for a general communicative principle: speakers strive to be as truthful, informative, relevant, and brief as they can. Therefore, a disjunction commonly results in the inference that the speaker could not have uttered only one of the disjuncts, probably because they were uncertain about its truth (ignorance implicature). 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 implicature). 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 semantic and pragmatic 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) differ from adults in their interpretations of disjunction in two ways. First, they are more likely to interpret *or* as *and* (Braine & Rumain, 1981; Neimark, 1970; Singh, Wexler, Astle-Rahim, Kamawar, & Fox, 2016; Tieu et al., 2016). This is often referred to as the conjunctive interpretation of disjunction. Second, preschool children are more likely to interpret a disjunction as inclusive.

In other words, unlike adults, children do not "compute exclusivity implicatures", and therefore consider a disjunction as felicitous when both disjuncts are true (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Chierchia et al., 2004; Crain, 2008). This is often referred to as children's "lack of exclusivity implicatures".

In the present study, we tested adults and preschool children's comprehension of linguistic disjunction in simple existential sentences and did not find evidence for substantial differences between adults' and preschool children's understanding of linguistic disjunction. We start with a broad review of the literature on children's acquisition of disjunction. Previous studies provide considerable support for the role of task demands in children and adults conjunctive interpretations of disjunction. Next we present three experiments that tested adults and children using two- and three-alternative forced-choice judgment tasks. We also collected and categorized children's spontaneous verbal responses in the same tasks. Both forced-choice judgments and quantitative analyses of spontaneous verbal feedback suggested that children did not interpret *or* as *and*. Therefore, we did not find evidence that children and adults differ in this respect. With respect to exclusivity implicatures, forced-choice judgment tasks suggested that children interpreted a disjunction as inclusive whether they were measured with two or three response options. On the other hand, adults considered a disjunction as inclusive if they had two response options, but showed sensitivity to the exclusivity implicature when they were given three options. However, the quantitative analysis of children's spontaneous verbal feedback suggested that children were also sensitive to the exclusivity implicature of disjunction. More specifically, children explicitly mentioned the word *and* as a better alternative to *or* when both propositions were true, yet they did not consider such a disjunction "wrong". In short, while the forced-choice measurements failed to capture children's sensitivity to exclusivity implicatures, such sensitivity was reflected in their spontaneous verbal feedback. Taken together, our studies support the hypothesis that task demands and issues of measurement could have contributed to the previously observed differences between preschool children and adults in their comprehension

of logical words (Katsos, 2014). In General Discussion, we discuss the implications of our studies 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 may have 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 theory[1].

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,

[1]The 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 also focus more on adult-like vs. non-adult-like behavior in children rather than "erroneous" 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.

preschool children can interpret a disjunction as exclusive or conjunctive. However, using a similar "give-item" task, Johansson and Sjolin (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 Sjolin (1975) argued that the linguistic *or* should be 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 Rumain (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 Rumain (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 Rumain (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 Rumain (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 Rumain (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' 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 & Rumain, 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; 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, 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 preschool 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 enrichment.

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

Table 1

Summary of tasks used in previous studies and their conclusions

Task	Conclusion
School test (Imperative)	Comprehension of *or* develops in high school.
Circle all that are bird or black!	Before that children often interpret it as conjunction.

Task	Conclusion
School test (Truth Value Judgment) *The bird is in the nest or the shoe is on the foot.*	Children (8-14 years) as well as some adults interpret *or* as a conjunction. This is likely due to task demands and application of nonlinguistic strategies.
Give-item *Give me all the green things or give me all the round things.*	Children (4-7 years) interpret disjunction as exclusive (choose-one). The inclusive (logical) concept of disjunction develops later.
Give-item + Truth Value Judgment *Give me all the green things or give me all the round things.* *Either there is X or there is Y in the box.*	Children (5-6 years) interpret *or* 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) *A troll ate a piece of pizza or an ice cream.*	Children (4-6 years) understand *or* 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 *or* 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 Rumain (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 (3AFC), and the analysis of children's open-ended verbal feedback 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) + Open-ended Feedback
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].

[2]There 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).

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 15 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 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 were on the card (e.g. CAT+DOG) and one was guessed (e.g. *cat*), the guess could be infelicitous or even false if interpreted exhaustively (e.g. **only** *cat*). In addition to acting as a control, such trials could show how often children derive exhaustive implicatures. Conjunction trials (e.g. *cat and dog*) were controls for disjunction trials. Conjunction trials were 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 only one animal was on the card (e.g. CAT) the disjunction guess (e.g. *cat or dog*) was true. When two animals were on the card (e.g. CAT+DOG), the disjunction

We expect similar behavior from labels that refer to non-maximal degrees of being "right" such as "a bit right" or "a little right".

guess (e.g. *cat or dog*) could be judged as true but infelicitous or even false. Such disjunction trials help us understand whether participants interpreted disjunction as inclusive or exclusive.

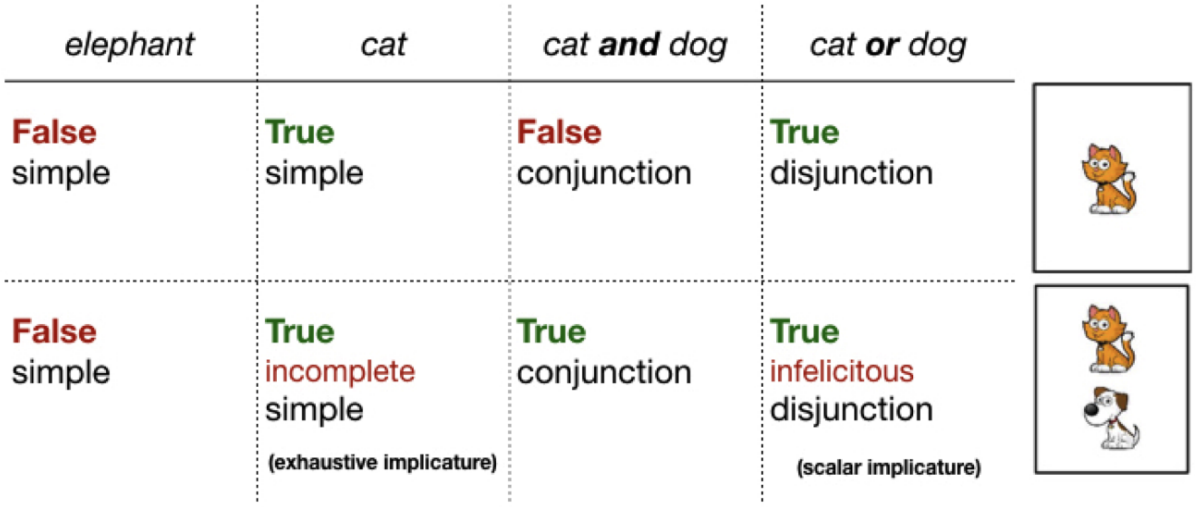

Figure 1. Rows show example cards and columns example utterances (guesses). Each cell represents a trial type.

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 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 15 in the appendix shows an example test trial.

Results

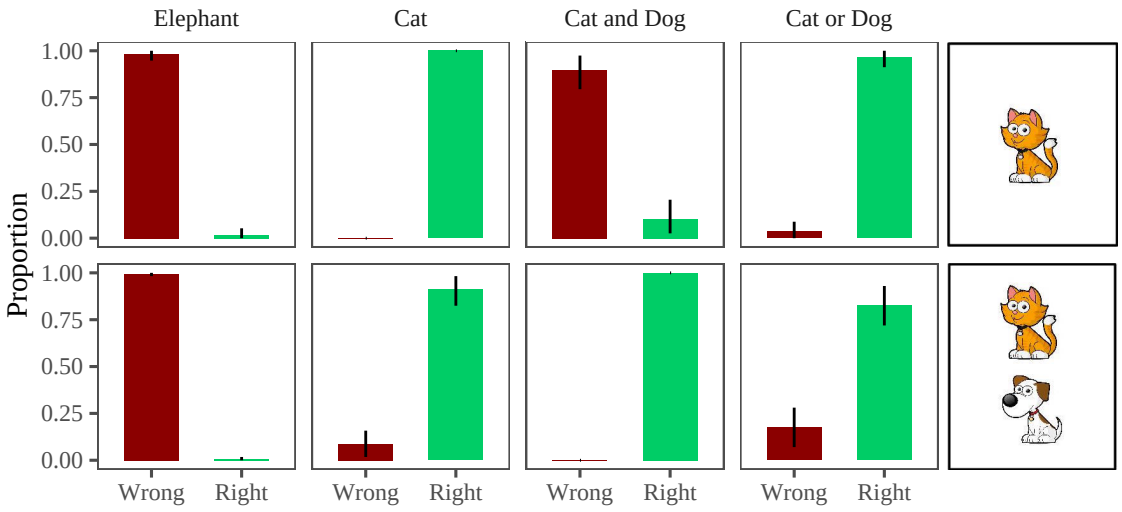

Figure 2. Adults' two-alternative forced choice judgments in Experiment 1. Columns represent example guesses and rows example cards.

Figure 2 shows the results for the adult 2AFC task. Starting with the leftmost column, participants judged false simple trials as "wrong". In such trials the guessed animal (e.g. *elephant*) was not on the card. In true simple and true-but-incomplete simple trials, the guessed animal (e.g. *cat*) was on the card and participants judged the guess "right". Moving to connective trials, when a conjunction (e.g. *cat and dog*) was false (e.i. only one animal was on the card) participants judged the guess "wrong". When the conjunction was true (i.e. both animals were on the card) they judged it "right". Both true disjunction trials and true-but-infelicitous disjunction trials were judged as "right". A disjunction guess (e.g. *cat or dog*) was true when one of the animals was on the card (e.g. CAT) and true-but-infelicitous when both were (e.g. CAT+DOG).

Figure 3 shows the results for the 3AFC judgment task. The addition of an intermediate response option did not affect false simple, true simple, and true conjunction trials. In false simple trials, the animal mentioned (e.g. *elephant*) was not on the card, and participants judged the guess "wrong". In true simple trials the animal mentioned (e.g. *cat*)

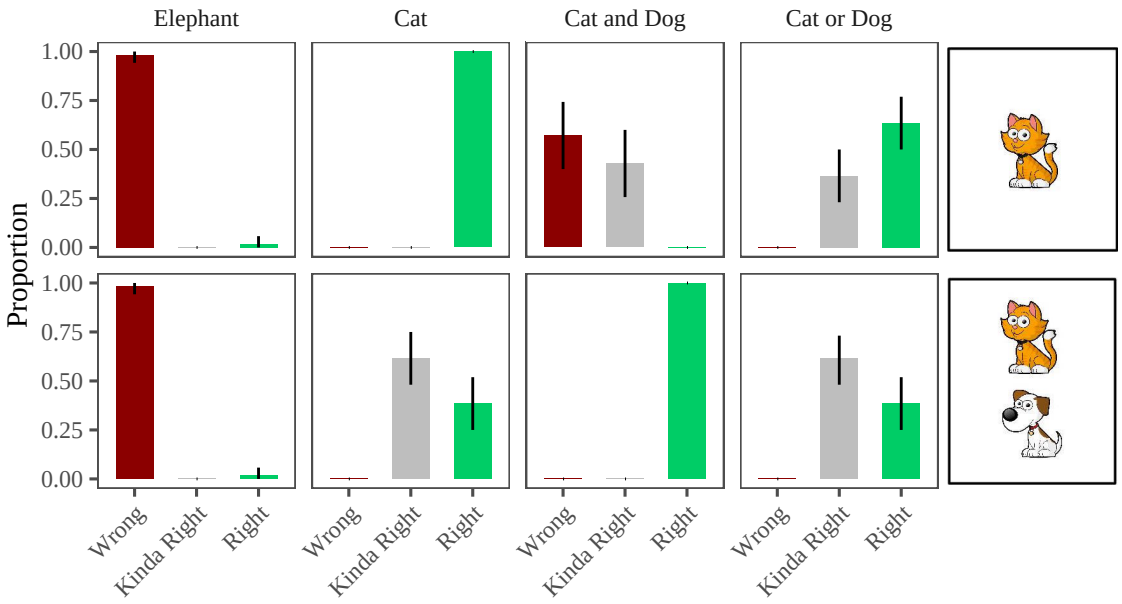

Figure 3. Adults' three-alternative forced choice judgments in Experiment 1.

was the only animal on the card and participant considered the guess "right". This was similar to true conjunction trials in which two animals were on the card (e.g. CAT+DOG) and the guess mentioned both (e.g. *cat and dog*). Participants judged true conjunction trials as "right" in both 2AFC and 3AFC tasks.

Four trial types showed different patterns of judgments in the 2AFC and the 3AFC tasks. In true-but-incomplete simple trials, 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" options. In false conjunction trials, only one animal was on the card (e.g. CAT), but two animals were guessed (e.g *cat and dog*). Most adults considered such false conjunctions "wrong" but some chose "kinda right". The intermediate option may have been used to express partial truth of the guess because one of the guessed animals was on the card. With true disjunction and true-but-infelicitous disjunction guesses, responses were split between "kinda right" and "right". It is likely that participants had different reasons for choosing "kinda right" in each disjunction trial type. In true disjunction trials, participants may have considered a simple guess (e.g. *there is a cat*) as more appropriate. In

true-but-infelicitous trials, 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. Since we are mainly interested in the differences between adults and children, we defer statistical analysis to Experiment 2 where we compare children and adults responses.

Discussion

Table 3

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

cat	dog	cat $\wedge$ dog	cat $\vee$ dog	cat $\oplus$ 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 map to "wrong" and true statements to "right."[3] In the 2AFC task, judgments for *and* matched logical conjunction and *or* inclusive disjunction. If adults in our task interpreted *or* as exclusive, we expected majority "wrong" responses when both disjuncts were true. This is not what we found.

[3]see Jasbi et al. (2019) for a discussion of linking assumptions in forced-choice truth-value judgment tasks.

However, if truth conditions were all that mattered, the addition of the intermediate option (kinda right) in the 3AFC task should not have substantially affected the judgments. In fact it did not in false simple trials, true simple trials, and true conjunction trials. These cases showed unequivocal "wrong" and "right" judgments. But in four other trial types, the intermediate option (kinda right) reflected more nuanced judgments. Responses in these trial-types fell into two patterns. First, responses in the false conjunction trials were split between "wrong" and "kinda right" responses. 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. Second, in true-but-incomplete simple trials, true disjunction trials, and true-but-infelicitous disjunction trials, the judgments were split between "kinda right" and "right". These trial types included guesses that 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, a conjunction (e.g. *cat and dog*) was the optimal guess. Similarly, when only one animal was on the card (e.g. CAT), a disjunction guess (e.g. *cat or dog*) was true but not optimal. A simple guess (e.g. *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 best guess or prediction.

In a forced choice task, 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 incompleteness 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 acceptability of disjunction.

Experiment 2: Children's 3AFC judgments and open-ended 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 Experiment 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 15). 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 circle[4]. 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 types[5].

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 (none) referred to cases where children did not say anything and 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, for example: "you are right!", "yes", "nope", or "you winned!". In the third

[4]The pilot study had shown that some children struggle with understanding the word "wrong", so "not right" was used instead.

[5]A more detailed description of the procedure as well as the randomization code for the test phase is available on the study's online repository.

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 added functional elements such as focus words *just* and *only*, or emphasized the connective *AND*. 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

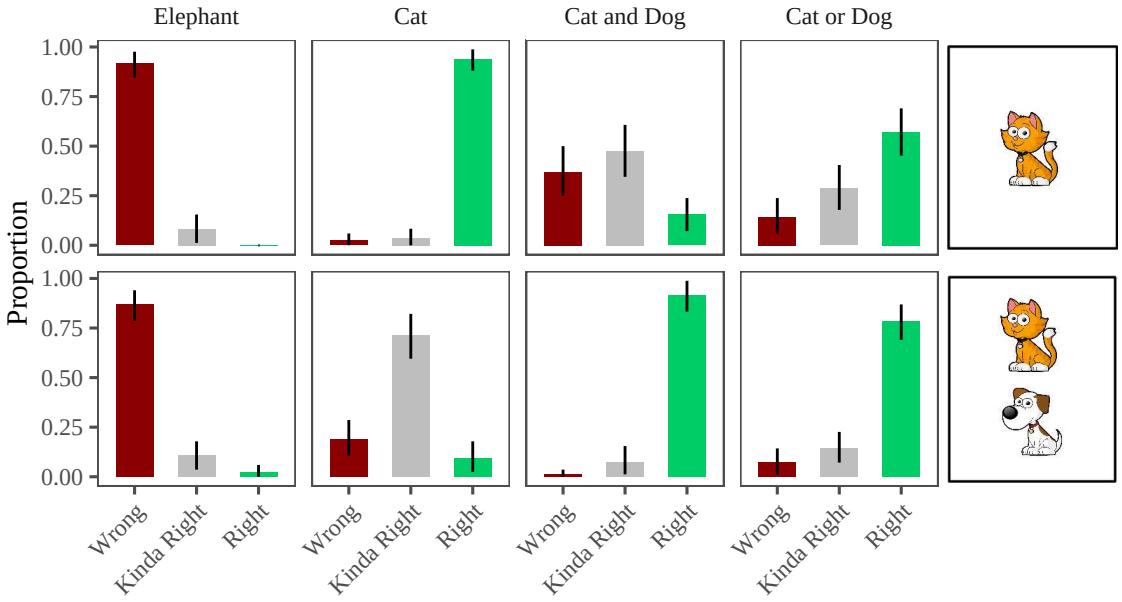

Figure 4. Children’s three-alternative forced-choice judgments in Experiment 2.

Three-Alternative Forced-Choice Judgments. Figure 4 shows the results for children’s 3AFC judgments. Starting with the leftmost column in Figure 4, children judged false simple trials as “wrong”. In these trials the mentioned animal (e.g. *elephant*) was not on the card. Moving to the second column, children judged true simple trials as “right”. In these trials the mentioned animal (e.g. *cat*) was the only animal on the card. Here we ignore the results for true-but-incomplete trials 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. Moving to the third column, children judged false conjunction trials as "wrong" or "a little right". In these trials, only one animal was on the card (e.g. CAT), but two were mentioned (e.g. *cat and dog*). In true conjunction trials, both mentioned animals were on the card and children judged the guess as "right". Finally, in true disjunction trials only one animal was on the card and children considered the guess (e.g. *cat or dog*) as either "right" or "kinda right". In true-but-infelicitous disjunction trials both animals were on the card and children judged the disjunction "right".

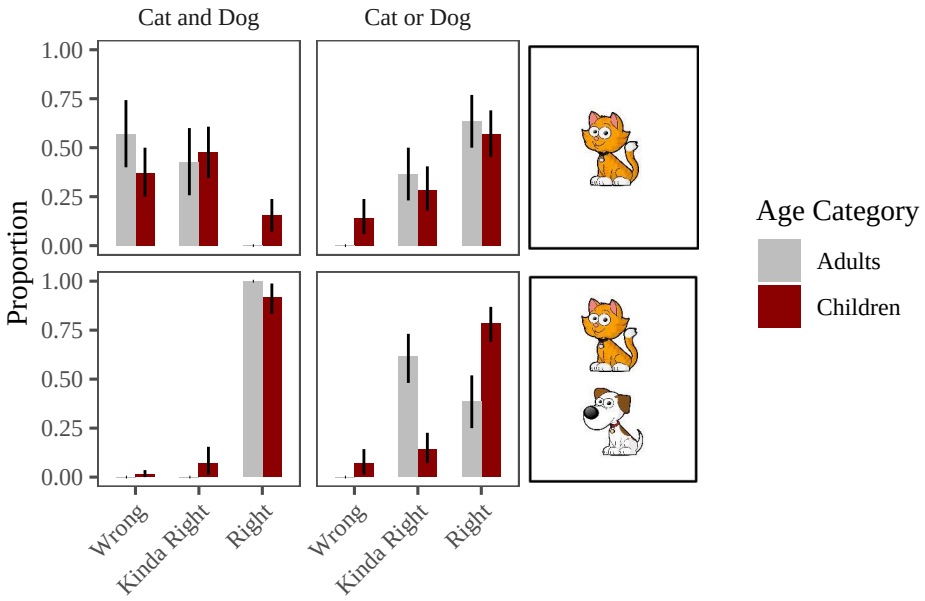

Figure 5. Comparison of Adults' 3AFC judgments from Experiment 1 and Children's 3AFC judgments from Experiment 2.

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. Since ordinal and multinomial logistic models with complex random effects structures are not easily fit in standard frequentist packages, 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).[6] 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.[7] For all models, the response variable had three ordered levels: "wrong", "kinda right", and "right". The trial types "T,Con" (true conjunction), "T.in,Dis" (true-but-infelicitous disjunction), and "F,Con" (false conjunction) constituted the (dummy-coded) fixed effects of the model, with "T,Dis" (true disjunction) 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).

Figure 6 shows the means and the 95% highest posterior density intervals (HPDIs) for our model coefficients. The left panel of Figure 6 shows the results from separate ordinal models for adults and children. It helps us understand how adults and children interpreted conjunction and disjunction separately. 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 three contrasts of interest. First, both adults and children rated false conjunction trials lower than true disjunction trials (F,Con - T,Dis) [children's 95% HPDI: -4.99, -0.16]. Although, adults seem to have had more negative judgments with false conjunction trials than children. Second, both adults and children judged true conjunction trials better than true disjunction trials (T,Con - T,Dis). Nevertheless, the 95% credible intervals for both groups contain zero. Finally, adults judged

[6] response ~ trial type + (1 + trial type|sid) + (1 + trial type|card)

[7] response ~ trial type * age category + (1 + trial type|sid) + (1 + trial type|card)

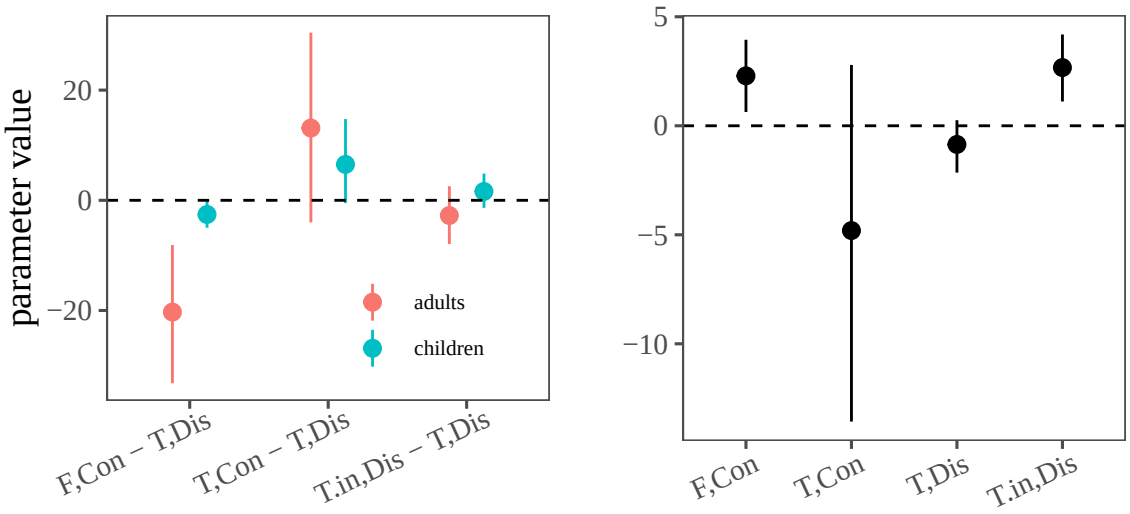

Figure 6. Left: The mean and 95% highest posterior density intervals for the coefficients estimated for each trial type in separate ordinal logistic regressions for adults and children. "F,Con - T,Dis" shows comparison of false conjunction and true disjunction trials; "T,Con - T,Dis" true conjunction vs. true disjunction trials; "T.in,Dis - T,Dis" true-but-infelicious disjunction vs. true disjunction trials. Right: Mean and 95% highest posterior density interval for the interaction coefficients (age category, adults as intercept) in the adult-child combined dataset. The x-axis labels stand for trial types shown in Figure 5.

true disjunction trials slightly better than true-but-infelicitous ones while children judged true-but-infelicitous disjunction trials slightly better. However, the 95% credible intervals for both groups contained zero. The means and credible intervals computed separately for adults and children match truth conditions of conjunction and disjunction: false conjunction trials were judged negatively and differently from true conjunction and disjunction trials.

To provide a precise estimate of the differences between adults' and children's judgments, we look at the means and the 95% HPDIs of the interaction coefficients in the combined adult-child dataset (Figure 6, Right). For false conjunction and true-but-infelicitous disjunction trials, the 95% credible intervals do not contain zero. This suggests that children's and adults' judgments differed in these two trial types. In both trial

types, children's judgments were higher than adults' judgments. This is consistent with two hypotheses: first that children are more lenient than adults, and second that children focus more on animal labels matching animal pictures on the card (Paris, 1973). Higher ratings in true-but-infelicitous trials are consistent with a third hypothesis as well: that children "compute exclusivity implicatures" at a lower rate than adults (Barner et al., 2011; Noveck, 2001; Papafragou & Musolino, 2003). However, we will see in the next section that children's spontaneous feedback does not support this hypothesis. Finally, we would like to add that we ran a model with children's age as a predictor and did not find any evidence for an effect of children's age on their forced-choice judgments.

Open-ended Verbal Feedback. We also categorized and annotated children's spontaneous and free-form verbal feedback 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 the leftmost column of Figure 7), in false simple trials the guessed animal was not on the card (e.g. *elephant*) and children either provided judgments like "No!" or descriptions like "cat" or "cat and dog". Moving to the second column, in true simple trials the guessed animal (e.g. *cat*) was the only animal on the card and most children provided positive judgments like "Yes". In true-but-incomplete trials, the animal guessed (e.g. *cat*) was only one of the two animals on the card (e.g. CAT+DOG) and children provided a description of the the card, for example "cat and dog".

In false conjunction trials, only one of the animal was on the card when two were guessed (e.g. *cat and dog*) In such trials, children provided a high number of corrections and

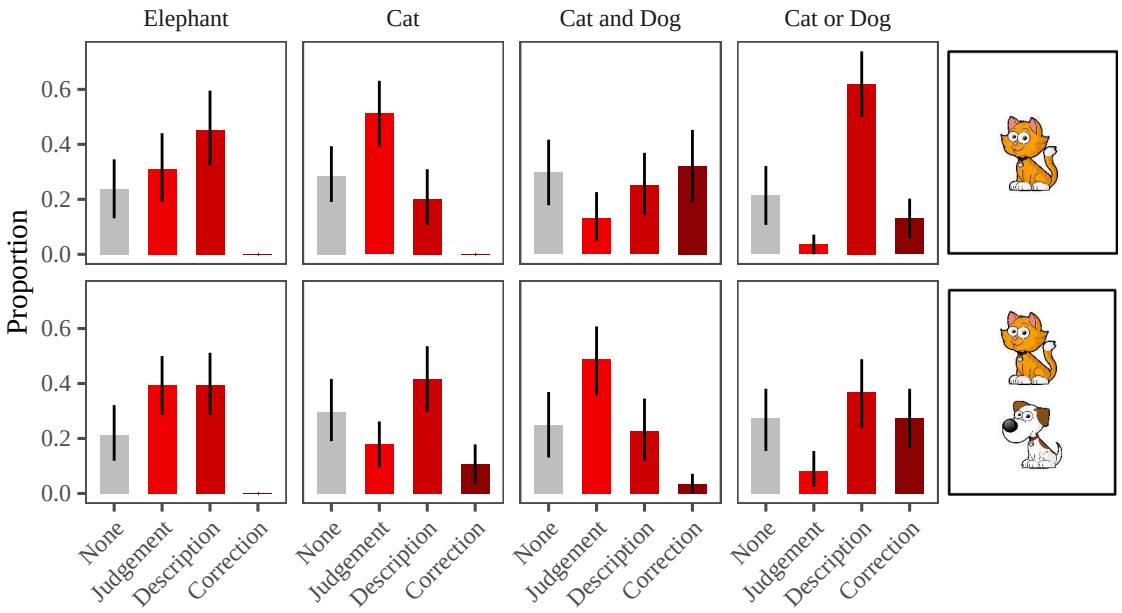

Figure 7. Children's open-ended vebal feedback in Experiment 2 classified as "judgments" (yes/no), "descriptions" (e.g. cat, dog, cat and dog), and "corrections" (e.g. just a cat, only a cat, "cat AND dog", Both!).

descriptions. In their corrections, children used the focus particles *just* and *only* as in "just a cat" or "only a cat". In true conjunction trials, both animals were on the card and children predominantly provided positive judgments like "Yes!". With true disjunction trials, only one of the guessed animals was on the card and most children simply provided a description of what was on the card (e.g. "cat"). However, in true-but-infelicitous disjunction trials, both animals were on the card yet children corrected the puppet by saying "Both!" or emphasizing *and* as in "cat AND dog!".

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).[8] The dependent measure was children's feedback categories of judgment, description, and correction, with correction set as the reference category. The trial types "T,Dis", "F,Con",

[8] feedback ~ trial type + (1 + trial type|sid) + (1 + trial type|card)

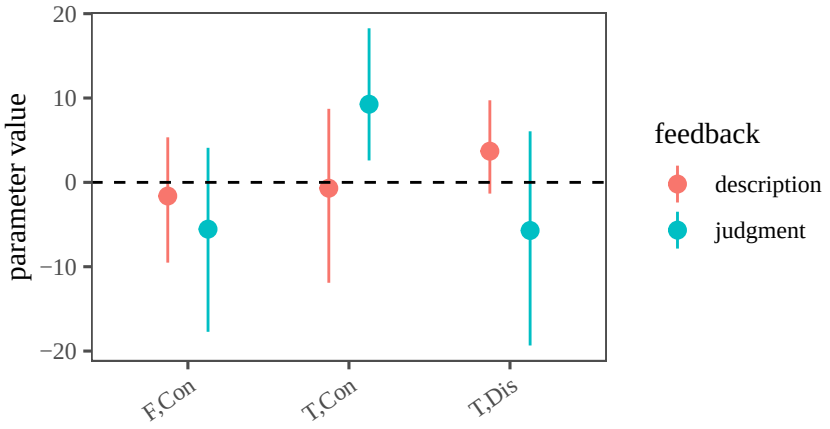

Figure 8. The mean and 95% highest posterior density interval of the coefficients of interest in experiment 2's mixed-effects multinomial logistic model on children's feedback. The category "correction" was set as the reference.

and "T,Con" constituted the (dummy-coded) fixed effects of the model with "T.in,Dis" set as the intercept. Priors and convergence information were identical to those reported for our previous models.

Figure 8 shows the means and 95% credible intervals of the multinomial model coefficients, with the x-axis separating trial types. Starting from the left, the credible intervals for judgments and descriptions over corrections for the "F,Con" trial-type included zero. This suggests that the feedback distribution was similar in false conjunction trials and true-but-infelicitous disjunction trials. Both trial types received a relatively high number of corrections. With "T,Con" trials, the credible interval for descriptions over corrections covers zero while that of judgments over corrections stays above zero. This suggests that with true conjunctions children provided more affirmative judgments like "yes" than corrections. Finally with "T,Dis" trials, the credible intervals 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 in true disjunction trials. Overall, our statistical modeling suggests that children were sensitive to false and infelicitous trials, providing more

corrections to the puppet in such contexts. We should also add that we included children's age as a predictor in a model and did not find any evidence for the effect of children's age on their verbal feedback.

Discussion

In Experiment 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 Experiment 1 with adults. The general comparison showed that adults and children had similar patterns of judgments with respect to the truth conditions of the connectives. Both groups had negative judgments for false conjunction statements and positive judgments for true conjunction and disjunction trials. Furthermore, we did not find any effect of children's age on their forced-choice judgments. This suggests that 3-to-5-year-old children understood the semantics of linguistic conjunction and disjunction in an adult-like manner. However, the results also showed that children's judgments differed from adults in two small ways. First, children were more likely to consider the guess "right" (and reward the puppet) when the guess was a false conjunction. Second, children were more likely to consider the guess "right" (and reward the puppet) when the guess was an infelicitous disjunction. This second difference is consistent with the hypothesis that children "lack exclusivity implicatures". However, our analysis of children's spontaneous verbal feedback provided evidence against such an interpretation.

To consider another measure of children's comprehension, we looked at children's spontaneous open-ended verbal feedback to the puppet's guesses. The results showed that children recognized false and infelicitous statements with the connectives, and provided appropriate corrective feedback. As expected from an adult-like understanding of connectives, children corrected the puppet most often when a conjunction was false (i.e. only one proposition was true), or when a disjunction was infelicitous (i.e. both propositions were

true). We did not find evidence for any age effect on children's verbal feedback either. Together with the results of forced choice judgments, these results suggest that children in this age range understood the truth conditions of conjunction and disjunction, and showed sensitivity to the pragmatic infelicity of a disjunction when both disjuncts were true. In fact, in their corrections, children often explicitly mentioned *and* as the correct connective that should have been used in such contexts. In the next Experiment, we replicate these results in a two-alternative forced-choice task.

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 if the disjunction is true but infelicitous. However, 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" format[9].

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 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.

[9]The As Predicted PDF document is accessible at https://aspredicted.org/x9ez2.pdf.

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*[10].

[10]The 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.

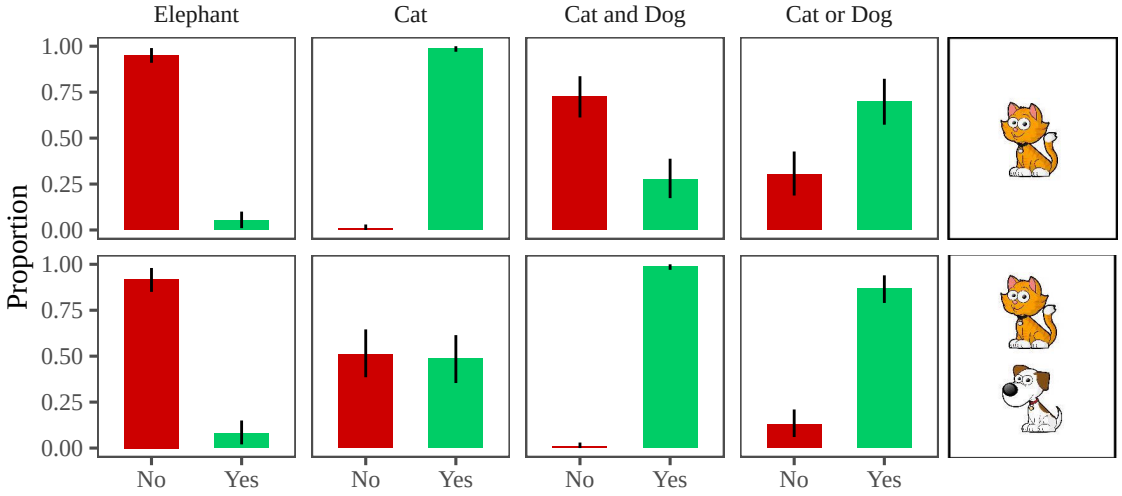

Figure 9. Children's two-alternative foced-choice judgments in Experiment 3.

Two-Alternative Forced Choice Judgments. Figure 9 shows children's 2AFC judgments. Starting with the leftmost column and false simple trial types, the guessed animal (e.g. *elephant*) was not on the card and children considered the guess "wrong". Moving to the next column and true simple trials, the guessed animal (e.g. *cat*) was the only animal on the card and children considered the guess "right". In true-but-incomplete trials only one of the animals on the card was guessed and children's judgments were equally split

between "wrong" and "right". 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 interpreters than adults (Noveck, 2001).

In false conjunction trials, only one of the two guessed animals was on the card and most children considered 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 true conjunction trials, children unanimously judged the guess "right", similar to adults. In true disjunction trials, the card had only one of the guessed animals and 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". Finally, with true-but-infelicitous disjunction trials, children considered the guess right.

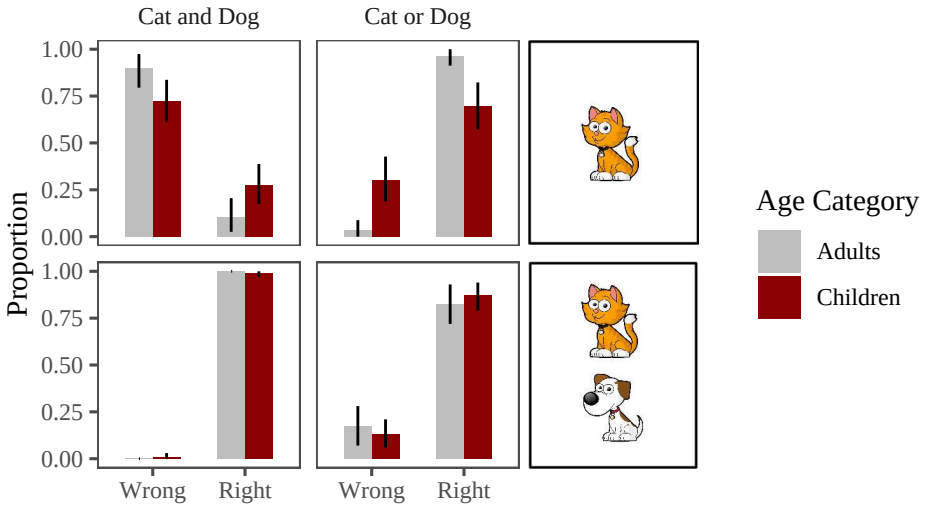

Figure 10. The comparison of 2AFC judgment tasks for conjunction and disjunction trials in adults (Experiment 1) and children (Experiment 3).

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 models included the fixed dummy coded effect of trial-type (Levels: "T,Dis" (reference) - "T,Con" - "F,Con" - "T.in,Dis"). The models also included random intercepts and slopes for participants and items. Details of priors and convergence were similar to previous models as well.

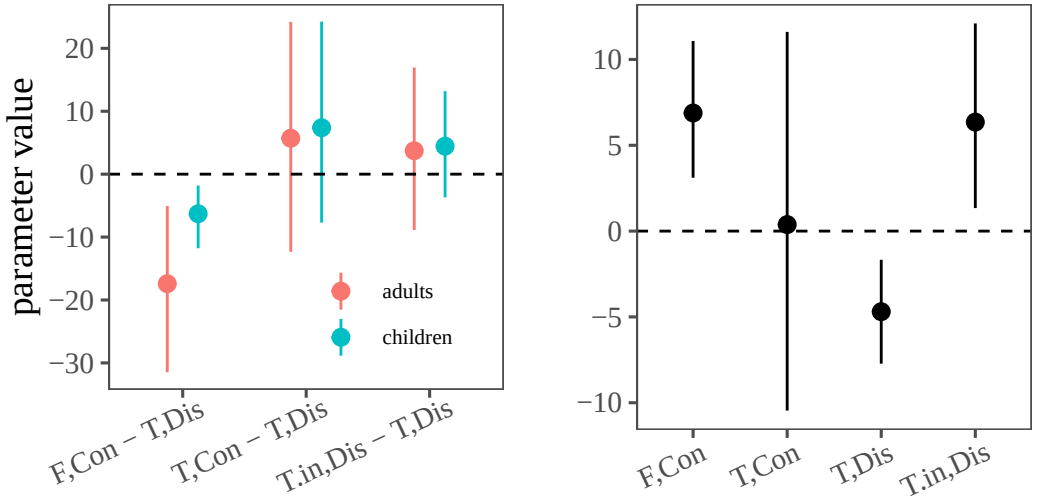

Figure 11. Left: The means and 95% highest posterior density intervals for parameter values estimated in separate ordinal logistic regressions for adults and children. "F,Con - T,Dis" shows the comparison of false conjunction and true disjunction trials; "T,Con - T,Dis" true conjunction vs. true disjunction trials; "T.in,Dis - T,Dis" true but infelicitous disjunction vs. true disjunction trials. Right: The means and 95% highest posterior density intervals for the interactive effect of age category (child vs. adult, adult intercept) on 2AFC judgments. The x-axis labels represent false conjunction, true conjunction, true disjunction, and true but infelicitous disjunction trials.

The left panel of Figure 11 shows the means and 95% HPDIs for three contrasts of

interest shown on the x-axis, estimated from separate binomial models for adults and children. First, for both adults and children, the 95% credible intervals for "F,Con - T,Dis" do not contain zero. This suggests that for both groups judgments of false conjunction trials were lower and different than true disjunction trials. Second, 95% credible intervals for "T,Con - T,Dis" estimated for adults and children contains zero. Therefore, adults and children had similar judgments for true conjunction and true disjunction trials. Third, the 95% credible intervals for "T.in,Dis - T,Dis" contains zero as well, suggesting that children and adults judged true-but-infelicitous disjunction trials similar to the true disjunction trials. Overall, the separate binomial models show that judgment patterns match the truth conditions of conjunction and disjunction. False conjunction trials were judged negatively and differently from true conjunction and disjunction trials.

To estimate the extent to which adults' and children's judgments differed from each other, we looked at the means and the 95% credible intervals of the interaction coefficients computed in the combined binomial model (Figure 11, Right). Based on the 95% credible intervals, we can infer that judgments of adults and children differed in three ways. First, children judged false conjunction trials slightly more positively than adults (F,Con). Second, they judged true disjunction trials slightly more negatively than adults (T,Dis). 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). The pragmatic enrichment account predicts that children would rate a true disjunction more negatively than adults, but does not predict more positive judgments for false conjunction trials. However, the label-matching account predicts both these outcomes, because it posits that in both cases, the match between animal labels and animal pictures affects children's judgments. Finally the third difference, children judged true but infelicitous disjunction trials more positively than adults did. 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. We would like to add that we ran a model with children's age as a predictor and did not find any evidence for an effect of children's age on their forced-choice binary judgments.

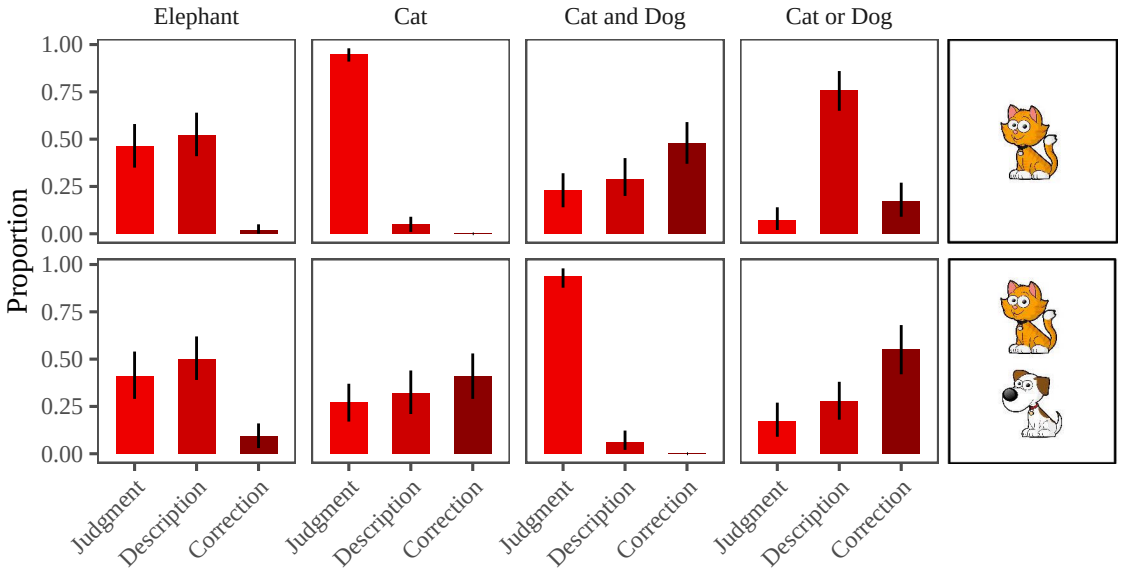

Figure 12. Children's open-ended vebal feedback in Experiment 3 classified as "judgments" (yes/no), "descriptions" (e.g. cat, dog, cat and dog), and "corrections" (e.g. just a cat, only a cat, "cat AND dog", Both!)..

Open-ended Verbal 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 in false simple trial types when the puppet guessed an animal not on the card (e.g. *elephant*), there was a split pattern between negative judgments like "No!" and descriptions like "Cat!". Second, almost all children responded with positive judgments like "Yes!" in true simple and true conjunction trial types. These are the trials where the puppet's guess correctly matched what was on the card. Third, children provided corrections in trials where the guess was either false or infelicitous. These included three trial types. First, true but incomplete simple trials in which two animals were on the card (e.g. CAT+DOG) but the puppet only guessed one (e.g. *cat*). Second, false conjunction trials in

which the puppet guessed two animals (e.g. *cat and dog*) but only one of them was on the card (e.g. CAT). Third, true but infelicitous disjunction trials in which two animals were 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 true disjunction trials. In these trials, the puppet used a disjunction (e.g. *cat or dog*) but only one of the animals was on the card. In such cases, almost all children simply named the animal on the card (e.g. "cat!").

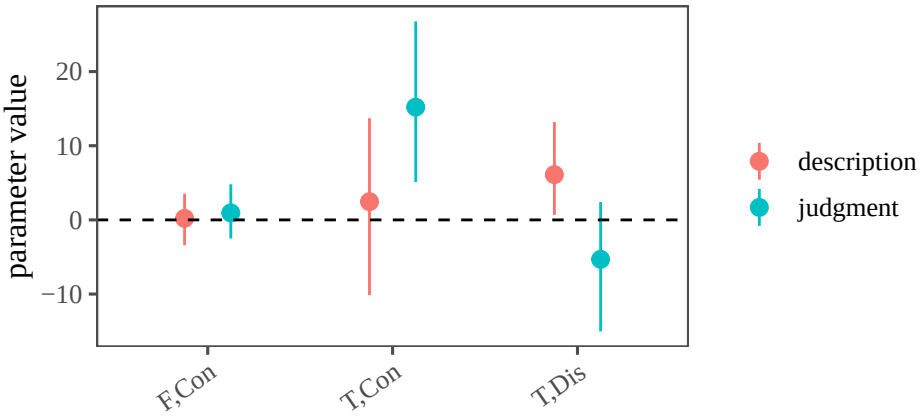

Figure 13. The means and 95% highest posterior density intervals of the coefficients of interest in experiment 3's mixed-effects multinomial logistic model on children's feedback.

To quantify and compare the distribution of children's feedback in trial-types with connectives, 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 "F,Con", "T,Con", "T,Dis" constituted the (dummy-coded) fixed effects of the model with "T.in,Dis" set as the intercept. Priors and convergence information were identical to those reported for our previous models.

Figure 13 shows the means and 95% credible intervals of the multinomial model co-coefficient. These results replicate the findings on children's feedback reported in Experiment 2. Starting from the left, the credible intervals for judgments over corrections as

well as descriptions over corrections for false conjunction trials (F,Con) include zero. This suggests that the feedback distribution was similar in false conjunction and true-but-infelicitous disjunction trials. In true disjunction trials (T,Dis), the credible interval for judgments over corrections includes zero but not that of descriptions. Therefore, children provided more descriptions than corrections in true disjunction trials (T,Dis). Finally with true conjunction trials (T,Con), the credible interval for descriptions over corrections includes zero, but not judgments over corrections. This suggests that with true conjunctions, children provided more affirmative judgments like "yes" than corrections. Overall, the results confirm the findings reported in Experiment 2: children were more likely to provide corrections in trial-types that were either false or infelicitous.

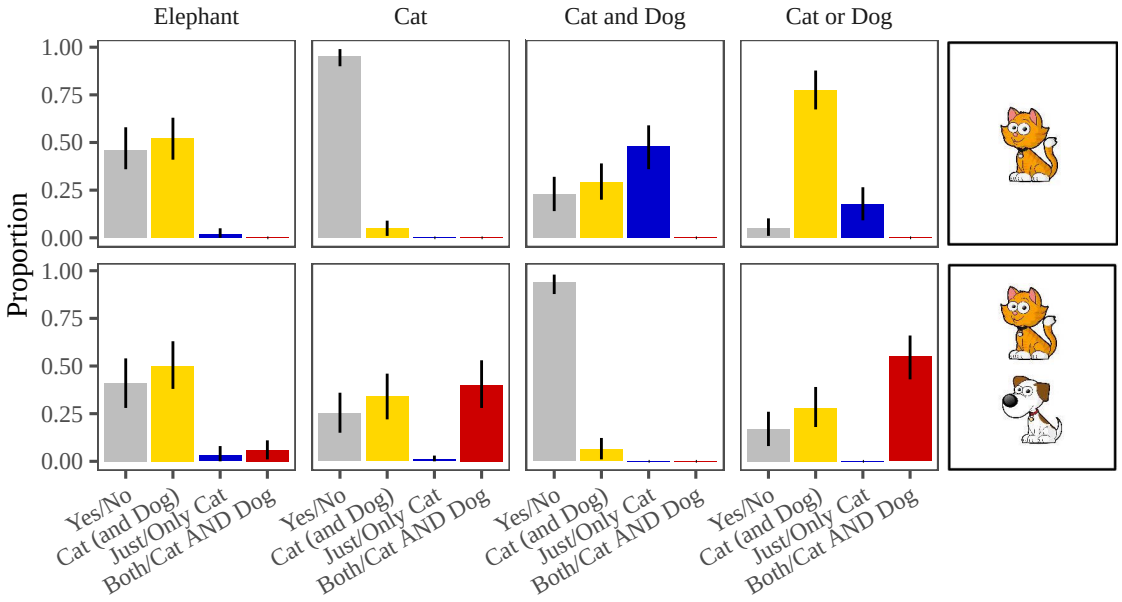

Figure 14. Children's open-ended feedback in different trial types of Experiment 3.

To better appreciate the pattern of spontaneous corrections provided by children, Figure 14 breaks down corrections into two sub-categories: those using exclusive focus words such as *only* and *just* (blue) and those using inclusive focus elements such as *both* and emphasizing *AND*. Our goal here is to focus on the trial types with corrective feedback (blue and red). The type of corrective feedback children provided in these trial types matched the type of mistakes made in the guesses. With conjunction 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 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 deep 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. The 2AFC responses followed the predicted pattern: a false conjunction was judged "wrong" and a true conjunction "right". Disjunction guesses were judged right whether they were true or true-but-infelicitous. 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. Taking both measures into account, we conclude the following: children understand the truth conditions of linguistic disjunction as shown by their 2AFC judgments; they also understand that when

both statements are true, a conjunction is a better (more felicitous) description than a disjunction, and their verbal feedback reflects that. However, they do not consider an infelicitous disjunction as "wrong".

General Discussion

This study did not find substantial differences between adults' and children's interpretation of linguistic disjunction in simple existential sentences. In the two-alternative forced-choice task, children and adults provided similar judgments that matched the inclusive interpretation of disjunction. In the three-alternative forced-choice task, adults chose the intermediate option more often when both propositions were true, suggestive they are sensitive to the exclusivity implicature of disjunction. While children did not do so in their forced-choice judgments, the quantitative analysis of children's verbal feedback showed that children were also sensitive to exclusivity implicatures of disjunction. These results provide evidence for the hypothesis that previously observed differences between children and adult's interpretation of disjunction may have been an artifact of the experimental task and the type of measurement used (Katsos, 2014; Paris, 1973; Skordos et al., n.d.).

The results reported here have two main implications for developmental semantics and pragmatics. First, children's conjunctive interpretations of disjunction in some of the previous studies have been attributed to a particular theory of pragmatic implicatures (Fox, 2007) and a developmental account in which children differ from adults with respect to the set of alternatives they generate while computing such implicatures (Singh et al., 2016; Tieu et al., 2016). However, as explained in our literature review, there is substantial evidence that conjunctive interpretations, even when robustly observed, are likely due to task demands and application of non-linguistic strategies (Braine & Rumain, 1981; Neimark & Slotnick, 1970; Paris, 1973; Skordos et al., n.d.). Therefore, in order to show instances of pragmatically enriched conjunctive readings in preschool children, it is crucial to first rule

out task-related conjunctive interpretations. This can be done by including trials in which the disjunction word (e.g. *or*) is replaced by a nonsense word. If it is truly the disjunction word that children enrich pragmatically via non-adult-like alternatives, then trials with the disjunction word should elicit higher conjunctive interpretations than control trials with the nonsense word.

Second, there are three major proposals to account for children's observed lower rate of scalar implicatures in experimental tasks (Noveck, 2001; Papafragou & Musolino, 2003). The first proposal focuses on processing difficulty, suggesting that implicature computations are cognitively taxing and children lack the appropriate processing resources (Pouscoulous, Noveck, Politzer, & Bastide, 2007; Reinhart, 2004). The second proposal is that children have not learned the scale (e.g. <*or*, *and*>), which allows for derivation of adult-like scalar implicatures (Barner et al., 2011; Horowitz, Schneider, & Frank, 2017). According to this proposal, children either lack the meaning for *or*, lack the meaning for *and*, or have not assigned *and* as the stronger alternative to *or*. Finally, the third proposal is that children are more tolerant of pragmatic infelicities than adults (Katsos & Bishop, 2011). When a speaker uses a linguistic form (e.g. a disjunction) that is true but not felicitous, children tolerate it and consider it "right" but adults do not.

The experimental results presented here do not fit the predictions of any of these accounts. We found that children are more likely than adults to judge a disjunction "right" when both propositions are true. This phenomenon is often referred to as "lack of scalar implicatures" in children. Yet, we also found that children are more liklely than adults to judge a simple guess (e.g. cat) as "wrong" when there are two animals (e.g. CAT+DOG). In other words, children were more likely to interpret a simple guess (e.g. "there is a cat") exhaustively (e.g. "there is *only* a cat"). Let's call this pattern "surplus of exhaustivity implicatures" in children. Neither the processing account nor the tolerance account predict "lack of scalar (exclusivity) implicatures" as well as "surplus of exhaustivity implicatures" in

preschool children. Whether children struggle with processing pragmatic inferences, or they are more tolerant of pragmatic violations, we should observe "lack of implicatures" across the board.

Non-adult-like lexicon does not seem to explain the results presented here either. Our experiments showed that preschool children differentiated *or* from *and*, interpreting each similar to adults (modulo exclusivity). Therefore, it is unlikely that children did not know the meaning of the weak member of the scale (i.e. *or*) or the strong member of the scale (e.i. *and*). Moreover with true-but-infelicitous disjunction trials, many children who judged the disjunction as "right" also informed the puppet in their verbal feedback that *and* should have been used instead. Mentioning *and* as the more felicitous alternative to *or* undermines the argument that children are not aware of *and* as the "scale-mate" to *or*. Taken together, the results of children's forced-choice judgments and their verbal feedback suggest children understood that the puppet should have used *and* instead of *or*, yet they did not consider this infelicity grave enough to render the guess "not right".

These results amplify the methodological concerns raised by Katsos (2014), and more specifically the issue of measuring semantic vs. pragmatic knowledge. The truth value judgment task uses the notion of an utterance being "right/wrong". However, it is not clear how different theoretical concepts such as entailment, presupposition, implicature, or infelicity link to the "right/wrong" scale and affect participant judgments in the truth value judgment task. More importantly, it is not clear whether this linking is the same for adults and children. The experiments presented here suggest that adults and children may differ on what semantic or pragmatic violations they consider "wrong" or "right", and that some (but perhaps not all) observations suggesting a lack of implicatures in children's comprehension may be due to the methods used for measuring pragmatic competence in children.

Since Tarski's original observations on disjunction, research in semantics and pragmatics has shown that the variety of interpretations Tarski observed are in fact distinct

types of meaning observed in many aspects of language and connected to distinct processes that generate them. Therefore, while the inclusive interpretation is hypothesized to be part of the semantics of a linguistic disjunction, exclusivity and ignorance interpretations are analyzed as distinct pragmatic inferences generated separately. This theoretical insight has in turn lead developmental researchers to seek distinct developmental mechanisms for each type of meaning. The results of the studies reported here suggest that as more and more varieties of meaning become subject to experimental studies, we also need to develop measures especially suited to capture the specific aspect of meaning under investigation.

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 *only/just*, emphasized *and* or used *both*	"only cat", "just elephant", "both!", "cat AND dog!"

References

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Braine, M. D., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*(1), 46–70.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

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

Somerville, MA: Cascadilla Press.

Chierchia, G., Crain, S., Guasti, M. T., & Thornton, R. (1998). "Some" and "or": A study on the emergence of logical form. In *Proceedings of the Boston University conference on language development* (Vol. 22, pp. 97–108). Somerville, MA: Cascadilla Press.

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 283–300). Basingstoke: Palgrave Macmillan.

Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. *Cognition*, *2*(2), 161–182.

Crain, S. (2008). The interpretation of disjunction in universal grammar. *Language and Speech*, *51*(1-2), 151–169.

Crain, S., Gualmini, A., & Meroni, L. (2000). The acquisition of logical words. *LOGOS and Language*, *1*, 49–59.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). Basingstoke: Palgrave Macmillan.

Goro, T., & Akiba, S. (2004). The acquisition of disjunction and positive polarity in Japanese. In *Proceedings of the 23rd West Coast conference on formal linguistics* (pp. 251–264). Somerville, MA: Cascadilla Press.

Grice, H. P. (1975). Logic and conversation. *1975*, 41–58.

Gualmini, A., & Crain, S. (2002). Why no child or adult must learn de Morgan's laws. In *Proceedings of the Boston University conference on language development*. Somerville,

926 MA: Cascadilla Press.

927 Gualmini, A., Crain, S., & Meroni, L. (2000). Acqisition of disjunction in conditional
928 sentences. In *Proceedings of the boston university conference on language development*.

929 Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2017). The trouble with quantifiers:
930 Exploring children's deficits in scalar implicature. *Child Development*.

931 Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to*
932 *adolescence: An essay on the construction of formal operational structures* (Vol. 84).
933 London: Routledge.

934 Jasbi, M., Waldon, B., & Degen, J. (2019). Linking hypothesis and number of response
935 options modulate inferred scalar implicature rate. *Frontiers in Psychology*, *10*.

936 Johansson, B. S., & Sjolin, B. (1975). Preschool children's understanding of the coordinators
937 "and" and "or". *Journal of Experimental Child Psychology*, *19*(2), 233–240.

938 Katsos, N. (2014). Scalar implicature. In D. Matthews (Ed.), *Pragmatic development in first*
939 *language acquisition* (Vol. 10, p. 183—198). Amsterdam: John Benjamins.

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

942 Neimark, E. D. (1970). Development of comprehension of logical connectives: Understanding
943 of "or". *Psychonomic Science*, *21*(4), 217–219.

944 Neimark, E. D., & Slotnick, N. S. (1970). Development of the understanding of logical
945 connectives. *Journal of Educational Psychology*, *61*(6p1), 451.

946 Nitta, N., & Nagano, S. (1966). Basic logical operations and their verbal expressions: Child's
947 conception of logical sum and product. *Research Bulletin of the National Institute for*

Educational Research, Tokyo, *7*, 1–27.

Notley, A., Thornton, R., & Crain, S. (2012a). English-speaking children's interpretation of disjunction in the scope of "not every". *Biolinguistics*, *6*(1), 32–69.

Notley, A., Zhou, P., Jensen, B., & Crain, S. (2012b). Children's interpretation of disjunction in the scope of "before": A comparison of English and Mandarin. *Journal of Child Language*, *39*(03), 482–522.

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

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

Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, *16*(2), 278–291.

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

Reinhart, T. (2004). The processing cost of reference set computation: Acquisition of stress shift and focus. *Language Acquisition*, *12*(2), 109–155.

Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., & Fox, D. (2016). Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics*, *24*(4), 305–352.

Skordos, D., Feiman, R., Bale, A., & Barner, D. (n.d.). *Do children interpret "or" conjunctively?* Retrieved from https://osf.io/2srxk/

970 Su, Y. (2014). The acquisition of logical connectives in child Mandarin. *Language*
971 *Acquisition*, *21*(2), 119–155.

972 Su, Y., & Crain, S. (2013). Disjunction and universal quantification in child mandarin.
973 *Language and Linguistics*, *14*(3), 599–631.

974 Suppes, P., & Feldman, S. (1969). *Young children's comprehension of logical connectives.*
975 *ERIC*. Department of Health, Education, Welfare. Office of Education.

976 Tarski, A. (1941). *Introduction to logic and to the methodology of the deductive sciences*.
977 Oxford University Press.

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

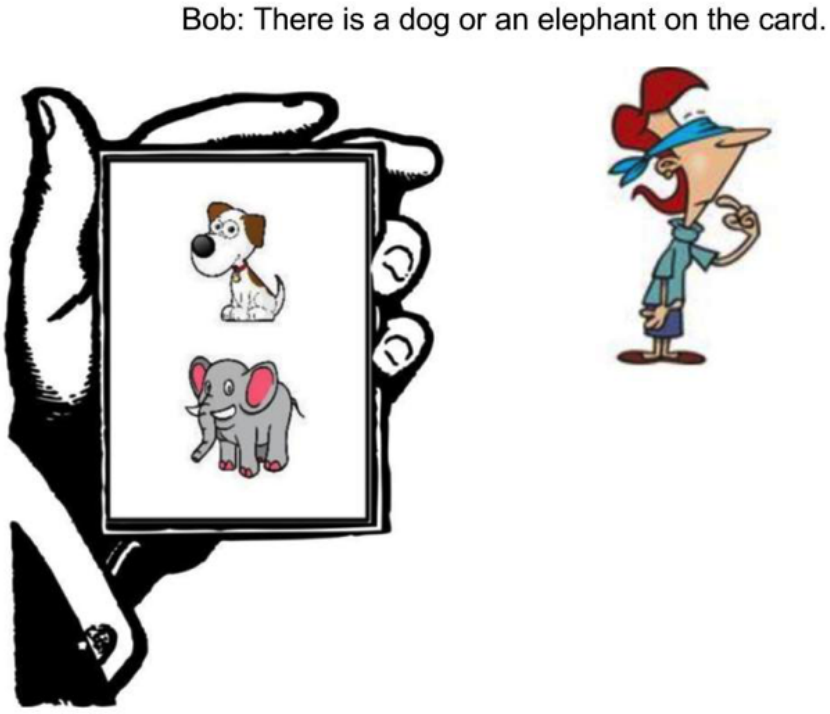

Figure 15. Cards used in the connective guessing game, sample scenario for adults in Experiment 1, and the puppet in children's experiments with the sleeping mask on and off.