Measuring Adults' and Children's Comprehension of 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 had a key role in advancing theories of logic, language, and cognition. Previous research suggests that adults and children might differ in their interpretation of linguistic disjunction in two ways. First, unlike adults, children might interpret a disjunction as conjunction. Second, children might interpret *or* as inclusive disjunction when adults interpret it as exclusive. We first review the long tradition of research on children's development of disjunciton. We show that previous research suggests conjunctive readings of disjunction are mainly due to task demands. Then we present three studies that assess adults and children's understanding of *and* and *or* using three different measures: binary truth value judgments, ternary truth value judgments, and free-form verbal feedback. We report that preschool children and adults do not differ in their binary judgments of disjunction. With ternary judgments, they show similar results except when both disjuncts are true. Adults tend to rate such disjunctions lower (exclusivity implicatures) while children consider them "right". In their free-form verbal feedback, however, children explicitly correct such infelicitous disjunctions and suggest that the connective *and* should have been used instead of *or*. These results suggest that forced-choice truth-value judgments may underestimate children's pragmatic competence. In order to capture children's semantic as well as pragmatic competence, we recommend complementing truth value judgment tasks with measures more sensitive to pragmatic infelicities.

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

Word count: X

Measuring Adults' and Children's Comprehension of Disjunction

Introduction

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

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

In addition to this IGNORANCE implication – that neither disjunct is known to be true – Tarski noted that a disjunction has at least two different interpretations: exclusive and inclusive. For example, a child may ask us to be taken to a hike in the morning and a theater in the afternoon, but we may respond: "No, we are going on a hike or we are going to the theater". He 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 provided an elegant explanation for the complex set of interpretations that linguistic disjunction receives. He argued that the literal meaning of *or* (i.e. its semantics) is captured by the truth conditions of logical inclusive disjunction. However, this literal meaning is enriched as speakers use a disjunction in context. Ignorance and exclusivity

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

Intricacies involved in the interpretation of disjunction raise a developmental issue: how does this complex linguistic knowledge develop in humans? When do children begin to interpret a disjunction? What are their early interpretations like? How do children learn the interpretations of disjunction? The present paper provides a broad overview of the literature that addresses these questions and builds on it in two ways. First, it further improves on previous experimental methods. Some previous experimental studies used complex experimental design, complex linguistic stimuli, or alternatively lacked appropriate controls such as a control connective or comprehension of adults in the same task. In section we review these issues and in section we present an experimental paradigm that avoids them. Second, most previous research tested children and adults using two-alternative forced-choice tasks (2AFC) (Crain & Thornton, 1998). Here, we report adults and children's judgments with both two and three alternatives (2AFC and 3AFC tasks). We also compare children's truth value judgments against their open-ended verbal feedback to the speaker. We find that

different tasks and different types of measurements are sensitive to different categories of meaning. In order to not understimate children's linguistic competence, it is important to develop tasks and measurements that can reliably capture the category of meaning under investigation.

Previous Research

Researchers have studied children's comprehension of logical connectives *and* and *or* 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 the choice of experimental task affected the conclusions of previous studies on disjunction.

Within the Piagetian program, researchers hypothesized that abstract logical (i.e. inclusive) disjunction forms 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 disjunciton 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). 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 in-class tests were not suitable for assessing children's understanding of logical connectives. For example, 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 "comparing visual and auditory information with little regard for the implied logical relationship in the verbal description." Participants responded with "true" if the individual disjuncts matched the pictures and false otherwise. Such a non-linguistic strategy would yield correct answers for conjunction but incorrect (conjunctive) answers for disjunction. This also explains why in Paris's study, conjunctive readings reduced with age and why using the word *either* along with *or* helped reduce conjunctive interpretations further.

Suppes and Feldman (1969) tested preschool children with a "give-item" task. They provided children with wooden blocks of different colors and shapes and used commands such as "give the things that are round or green." They found that depending on the exact phrasing of the command, preschool children can interpret a disjunction as exclusive or conjunctive. However, Johansson and Sjolin (1975)'s give-item task did not find considerable

[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 "errorneous" behavior. Depending on the context, a disjunction may be interpreted as exclusive, inclusive, or even conjunctive, and adults set the benchmark interpretation for children's performance in experimental tasks.

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, suggesting that they may have been due to task demands. In the truth value judgment task, a puppet described the contents of four boxes that each contained 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.

It is important to note that in Braine and Rumain (1981)'s judgment task, the puppet used 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. 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. Following Grice, Chierchia et al. (1998) argued that in order to capture children's semantic competence with *or*, experiments need to test the comprehension of disjunction in contexts that disjunction is not enriched with exclusivity implicatures. These contexts include embedding *or* under linguistic operators such as negation or conditionals.

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, Crain, Guasti, Gualmini, & Meroni, 2001; Chierchia et al., 2004), nuclear scope of the negative quantifier *none* (Gualmini & Crain, 2002), restriction and nuclear scope of *not every* (Notley, Thornton, & Crain, 2012), and prepositional phrases headed by *before* (Notley, Zhou, Jensen, & Crain, 2012), 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. A conclusion that stands in sharp contrast to the earlier conclusions from the give-item tasks. Since under the Gricean account, exclusive interpretation of disjunction is the result of pragmatic (scalar) implicatures, the earlier emergence of inclusive interpretations is considered as further evidence that children face delay in pragmatic development that results in generating scalar inferences (Barner, Brooks,

& Bale, 2011; Noveck, 2001; Papafragou & Musolino, 2003).

However, all the studies in the Gricean program use the Truth Value Judgment Task (Crain & Thornton, 1998). 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, it is possible that truth value judgment tasks are simply not suitable for capturing 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 applying the nonlinguistic strategies, as discussed by earlier studies (Braine & Rumain, 1981; Paris, 1973).

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

However, a recent replication of Tieu et al. (2016) by Skordos, Feiman, Bale, and Barner (2018) 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. (2018) 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
In-class tests	Children (6-10 years) interpret *or* as *and* (conjunctive interpretation). Older children consider disjunction as exclusive. Only adults interpret it as inclusive.
Give-item	Children (4-7 years) interpret disjunction as exclusive (choose-one). The inclusive (logical) concept of disjunction develops later.
Give-item + Truth Judgment (not controlling for speaker ignorance)	Children (5-6 years) interpret *or* as exclusive in commands but ignore its contribution in in truth value judgments and interpret it as a conjunction. Interpretation of disjunction in commands develops earlier than the knowledge of its truth conditions.

Task	Conclusion
Truth Judgment (controlling for speaker ignorance)	Children (4-6 years) understand the truth conditions of *or* similar to inclusive disjunction. Inclusive interpretation develops earlier than exclusive interpretations. Two studies report majority conjunctive interpretations too.
Truth Judgment (controlling for speaker ignorance, controlling for number of alternatives)	Children (4-6 years) understand the truth conditions of *or* similar to inclusive disjunction.

To summarize, our review of previous literature suggests that experimental tasks 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 interpreations 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 may 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 disapppeared when Skordos et al. (2018)'s replication controlled for the number of alternatives in the task. Finally, early studies which included in-class tests found that even college students can interpret *or* as a conjunction if participants find the task difficult or confusing.

Present Study

This study does not find evidence for conjunctive interpretations. It finds evidence for the possible role of measurement in our conclusions on children's exclusivity implicatures.

Previous studies sometimes used complex linguistic stimuli or relatively complex designs that may have increased the application of non-linguistic strategies. Some studies violated "speaker ignorance"; i.e. had the speaker utter the disjunction when the truth of the propositions were known to the speaker. Some studies did not use the conjunction word (e.g. *and*) in control trials, or did not use adults as control participants. Finally, some studies tested the disjunction word in linguistic environments that collapse interpretive differences between a conjunction and a disjunction. The experimental paradigm reported here builds on and improves previous studies by controlling for all these factors.

In the studies reported here, we used simple existential sentences (e.g. *there is a cat or a dog*) and tested the interpretation of participants in a simple and easy to understand guessing game. The guessing game provided a context in which the speaker was ignorant with respect to which alternatives were true. The game is essentially a variant of the truth value judgment task and used conjunction trials as well as adult participants as controls. Adults provided different interpretations for conjunction and disjunction trials in the task. Furthermore, we tested children's interpretations in two different ways, using forced choice tasks with 2 and 3 options (2AFC and 3AFC tasks), as well as free-form verbal responses.

Study 1: Adult's 2AFC and 3AFC Judgments

The goal of this study was to examine adults' comprehension of *and* and *or* as a benchmark for children's comprehension. Participants saw a card, read a description, and had to evaluate the description with respect to what they saw on the card. In test trials, the descriptions contained the conjunction word *and* and the disjunction word *or*. 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". The two conditions were otherwise identical[2]. 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 with cartoon images of a cat, a dog, and an elephant (Figure 1). There were two types of cards: cards with only one animal and cards with two animals. There were three types of guesses: simple (e.g. *There is a cat*), conjunctive (e.g. *There is a cat and a dog*), and disjunctive (e.g. *There is a cat or a dog*). In each guess, the animal labels used in the guess and the animal images on the card could have no overlap (e.g. Image: dog, Guess: *There is a cat or an elephant*), partial overlap (e.g. Image: Cat, Guess: *There is a cat or an elephant*), or total overlap (e.g. Image: cat and elephant, Guess: *There is a cat or an elephant*). Crossing the number of animals on the card, the types of guesses, and the overlap between the guess and the card yields 12 different possible trial types. We chose 8 trial types (Figure 2), to balance the number of one-animal vs. two-animal cards, simple vs. connective guesses, and expected true vs. false trials.

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

[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). We expect similar behavior from labels like "a bit right" and "a little right" which refer to non-maximal degrees of being "right".

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-guess scenario was chosen at random. The order of trial types was also randomized. Figure 3 shows an example test trial.

Table 2

Summary of study 1 methods with adult participants

Study	N	Age	Mode	Response Options
Study 1 - 2AFC	57	Adults	Online (Mturk)	Wrong, Right
Study 1 - 3AFC	52	Adults	Online (Mturk)	Wrong, Kinda Right, Right

Results

In this section, we first present the results of the 2AFC and 3AFC tasks with adults. Then we discuss how these results can be interpreted with respect to the semantics and pragmatics of disjunction in the context of the guessing game.

Figure 4 shows the results for the adult 2AFC task. The two left columns show the simple guesses and serve as controls. The results show that if the animal mentioned in the guess was not on the card (e.g., elephant), participants judged the guess to be "wrong"; if the animal was on the card (e.g., cat), participants judged the guess to be "right". The next two columns of Figure 4 show the results for the test conditions, namely conjunction and disjunction. An *and*-guess (e.g. cat and dog) was considered "wrong" if only one of the animals was on the card, and "right" if both were. An *or*-guess (e.g. cat or dog) was "right" whether one or both animals were on the card. The patterns of "right" and "wrong" responses in the binary task match the expectations for truth and falsehood of logical

conjunction and (inclusive) disjunction.

Figure 5 shows the results for the 3AFC judgment task. For four trial types, the results were identical to the 2AFC task. In the first and second trial types, if the animal mentioned was not on the card (e.g. elephant), participants judged the guess as "wrong", regardless of whether one animal was on the card or two. In the third trial type, if the animal mentioned (e.g. cat) was the only animal on the card, participants judged the guess as "right". Finally, if there were two animals on the card and the puppet mentioned them using *and* (e.g. cat and dog), all participants considered the guess "right".

The four remaining trial types showed different patterns of judgments than their counterparts in the 2AFC task. If the animal mentioned (e.g. cat) was only one of the animals on the card, participant judgments were divided between "right" and "kinda right". When only one of the animals was on the card (e.g. cat) and the guess was a conjunction (e.g cat and dog), most adults considered the guess "wrong" but some chose "kinda right", possibly suggesting that the intermediate option was used to express partial truth of a guess. With *or*-guesses (e.g. cat or dog), if the card had only one of the animals (e.g. cat), most participants considered the guess "right" while some considered it "kinda right". It is possible that those who chose "kinda right" considered the completely right guess to be "cat". For *or*-guesses with both animals on the card, adults were split between "kinda right" and "right" responses. The choice of "kinda right" over "right" in such trials can be interpreted as a sign that adults were sensitive to the infelicity of a disjunction when conjunction was more appropriate. In the next section, we discuss the nature of pragmatic reasoning in the context of this guessing game.

Discussion

Consider the following truth conditions for *and* and *or*: A conjunction with *and* is true when both conjuncts are true and false otherwise. An inclusive disjunction with *or* is true when at least one disjunct is true, and false otherwise. An exclusive disjunction is true when

only one of the disjuncts is true and false otherwise. Let's also assume a simple linking function in which false statements are judged as "wrong" and true statements as "right" (see Jasbi et al. (2019) for a discussion of linking assumptions in such a task). In the context of study 1, this purely truth-conditional account has the following predictions: First, conjunction guesses like "cat and dog" are wrong when only one of the animals is on the card and right when both are. Second, disjunction guesses are always right if they are interpreted as inclusive, because in all such trials at least one of the animals is present on the card. However, if disjunctive guesses are exclusive, they are right when one of the animals is on the card and wrong when both are. Finally, the addition of a third intermediate option between wrong and right should not substantially affect the judgments.

Figure 4 shows that in the binary task, judgments for *and* best match the predictions of the truth-conditional account for logical conjunction. For *or*, the judgments match the predictions of inclusive disjunction, rather than exclusive disjunction. However, the ternary judgments deviated from a purely truth-conditional account in four trial types: (i) trials with simple guesses when two animals were shown on the card; (ii) disjunction trials with one animal; (iii) disjunction trials with both animals on the card; and (iv) conjunction trials with one animal on the card.

These trial types fall into two major categories with respect to their response patterns. First, those in which participants chose "kinda right" and "right" (i-iii); Second, those in which participants chose "wrong" and "kinda right" (iv). The first category corresponds to trial types in which the guesses were literally true, but pragmatically infelicitous. In trial types (i) to (iii), there were always better alternative guesses. When there were two animals on the card (e.g. cat and dog), a guess mentioning only one of them (e.g. there is a cat) was technically true but a better guess would have been one mentioning both animals with *and* (e.g. cat and dog). This was also the case for disjunctive guesses (e.g. cat or dog) when both animals were on the card. When only one animal was on the card (e.g. cat), a simple guess (e.g. there is a cat) was more appropriate than a disjunctive one (e.g. there is a cat or a dog),

even though a disjunctive guess is literally true.

The second category of responses, namely "wrong" and "kinda right", only happened in one trial type: when there was one animal on the card (e.g. cat) and the guess was a conjunction (e.g. cat and dog). While the majority of participants considered such guesses as "wrong", some considered them not as bad as failing to name any of the animals on the card. In other words, the pattern of judgments captured the fact that such conjunctive guesses correctly name one of the animals on the card but not both. Overall, the comparison of forced choice judgments with two and three alternatives suggests that two alternatives better captured the truth-conditional meaning of the connectives, but underestimated adult pragmatic reasoning in the guessing game.

Study 2: Children's three-alternative forcied choice judgments vs. open-ended verbal feedback

The goal of this study was to examine children's interpretations of *and* and *or* in the guessing game and compare them to those of the adults'. Since the 3AFC judgment task in study 1 was better at capturing the nuances of adults' pragmatic reasoning, we decided to first test children using the 3AFC task. We also analyzed children's open-ended verbal feedback about the guesses in the same task.

Methods

Table 3

Summary of Study 2 Methods

Study	N	Age	Mode	Response Option
Study 2	42	3;1-5;2 (M = 4;3)	Study Room	Circle (wrong), Little Star (little right), Big Star (right)

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

Materials. We used the same set of cards and linguistic stimuli as the ones in study 1. There were 8 trial types and 2 trials per trial type for a total of 16 trials. We made two changes to make the experiment more suitable for children. First, instead of the fictional character Bob, a puppet named Jazzy played the guessing game with them. Jazzy wore a sleeping mask over his eyes during the game (Figure 6). 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 is right, the child should give him a big star, if he is a little bit right, a little star, and if he is not right, a red circle.

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, elephant, cat and dog (Table 4). He 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[3]. 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 had a cat and a dog on it. The experimenter said that the puppet was "a little right" and asked the child to give him a little star.

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

In the test phase, the experimenter removed the three instruction cards and placed a deck of 16 randomized cards on the table[4]. 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 paried with each card in a way that allowed two trials per 8 trial types (see).

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

[4]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.

Offline Annotations. While playing the game, children often provided spontaneous verbal reactions to the puppet's guesses. During the analysis of the videos, these verbal responses were categorized into four types: 1. None, 2. Judgments, 3. Descriptions, and 4. Corrections. The first category referred to cases where children did not say anything and only rewarded the puppet. Judgments referred to linguistic feedback such as "you are right!", "yes", "nope", or "you winned". Such feedback only expressed judgments and complemented the rewards. Descriptions were cases that the child simply mentioned what was on the card: "cat!", "dog and elephant!", "There is a cat and a dog!" etc. Finally, corrections referred to feedback that provided "focus words" (e.g. *just*, *only*, *AND*) that acted like corrections to what the puppet had said. Examples include: "Just a cat!", "Both!", "The two are!", "Only cat!", "cat AND dog" (with emphasis placed on *and*). In trials where the child provided both judgments as well as descriptions or corrections (e.g. "Yes! Cat!"), we placed the feedback into the more informative categories, namely description or correction.

Results

Figure 7 shows the results for children's 3AFC judgments. Starting from the left column, if the mentioned animal was not on the card (e.g. elephant), children judged the guess as "wrong". If the animal mentioned (e.g. cat) was the only animal on the card, children judged the guess to be "right". Here we ignore the results for trial types in which the animal mentioned was one of the animals on the card. 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.

In conjunctive guesses (e.g. *cat and dog*), when only one of the animals mentioned was on the card, children judged the guess as "wrong" or "a little bit right". However, if both animals were on the card, they judged it "right". In disjunctive guesses (e.g. *cat or dog*), when only one of the animals mentioned was on the card, children considered the guess "right" or "kinda right". If both animals were on the card, it was considered "right".

Figure 8 compares the results for children and adults' 3AFC judgments in the conjunction and disjunction trials. The major difference between adults and children's responses was disjunctive trials with two animals on the card. Most children considered such trials as "right" while most adults considered them as "kinda right". In the next section, we use Bayesian regression modeling to compare adults' and children's three-alternative responses more systematically.

Analysis and Statistical Modeling. We used the R package RStan for Bayesian statistical modeling to fit separate ordinal mixed-effects logistic models for adults' and children's judgments. The response variable had three ordered levels: *wrong*, *kinda right*, and *right*. The trial types *One-Animal-OR*, *Two-Animals-OR*, *One-Animal-AND* constituted the (dummy-coded) fixed effects of the model with *Two-Animals-AND* set as the intercept. The model also included by-subject random intercepts. The priors over trial types and the random intercepts were set to $\mathcal{N}(0, 10)$. We also included parameters C_1 and C_2, the two cutpoints delimiting the logistic for 1) *wrong* and *kinda right* and 2) *kinda right* and *right* responses, drawn with the prior $\mathcal{N}(0, 1)$.[5] All four chains converged after 3000 samples (with a burn-in period of 1500 samples).

We made inferences based on the highest-posterior density (HPD) intervals for the coefficients estimated from each model. Because predictors are dummy-coded, it's possible to examine contrasts of interest by computing the difference between coefficients for pairs of conditions we wish to contrast. In naming the coefficients like b(OR,2), OR/AND represents the connective used and the number represents the number of animals on the card. Figure 9 shows the contrasts of interest: *b(OR, 2)-b(OR,1)* represents the difference between the estimated coefficients for the disjunction trials with two animal on the card and those with only one; *b(OR, 2)* represents the difference between the estimated coefficients for the conjunction trials with two animals and the disjunction trials with two animals; and so on.

Overall, adults' and children's estimated coefficients are similar in sign to one another,

[5] We used a tight prior in this case to decrease posterior correlations between cutpoints and intercept.

though adults' are more extreme. In the conjunction trials (*b(AND, 2)-b(AND,1)*), children and adults showed a strong preference for the cards with two animals rather than one. At the same time, given two animals on the card, children and adults showed a preference for *and* rather than *or* (*b(OR, 2)*). However, with only one animal on the card, children and adults preferred a disjunctive guess (*b(OR, 1)-b(AND,1)*). These results are compatible with the truth conditions of conjunction and disjunction.

The main difference between adults and children shows up in the contrast between the disjunctive trial types: two animals vs. only one (*b(OR, 2)-b(OR,1)*). On average, children rated disjunction trials with two animals higher than those with only one. Adults on the other hand showed the opposite pattern: they rated disjunction trials with two animals lower. This pattern is compatible with current accounts of pragmatic development that suggest children's interpretations tend to be more literal than adults (Barner et al., 2011; Noveck, 2001; Papafragou & Musolino, 2003).

The slight preference children show for cards with two animals when the guess is disjunctive (e.g. "cat or dog") is also compatible with the account proposed by Singh et al. (2016) and Tieu et al. (2016). However, the effect seems much smaller here than was reported in their studies. The comparison with conjunction trials makes it clear that overall, children are not interpreting *or* as a conjunction. The effect in this study can be more accurately described as a preference in truth value judgments for both disjuncts being true rather than a conjunctive interpretation of disjunction. The results from children's spontaneous linguistic feedback provide more evidence that children are not interpreting *or* as a conjunction. We will discuss these results next.

Table 5

Definitions and Examples for the Feedback Categories.

Category	Definition	Examples
None	no verbal feedback	

Category	Definition	Examples
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!"

Children's open-ended feedback. As explained in section , we also categorized and annotated children's spontaneous and free-form verbal reactions to the puppet's guesses. Table 5 summarizes the definitions and examples for each category and Figure 10 shows the results. We should point out that each trial type had a similar number of "None" cases. Some children remained more or less silent throughout the experiment and only provided rewards to the puppet. In the next study we ask children to provide feedback explicitly and therefore we have no "None" responses. In the discussion and analysis here we will not comment further on the "None" category but focus on the other three categories.

In the leftmost column, when the guessed animal was not on the card (e.g. "there is an elephant"), children either provided judgments like "No!" or described what was on the card (e.g. "cat" or "cat and dog"). However, when the guessed animal was the only animal on the card (e.g. "there is a cat"), most children provided a positive judgment like "Yes". When the animal guessed was only one of the animals on the card, children described what was on the card (e.g. cat and dog).

In the critical trial types with conjunction and disjunction, children showed a high rate of corrections and descriptions when there was only one animal on the card (e.g. cat) and the guess was a conjunction (e.g. "there is a cat and a dog"). In their corrections, children used the focus particles *just* and *only* as in "just a cat" or "only a cat". However, when both animals were on the card and a conjunction was used (e.g. "there is a cat and a dog"), children predominantly provided positive judgments like "Yes!" and "You are right".

Considering disjunctive guesses like "cat or dog", when only one of the animals was on the card, most children simply described what was on the card (e.g. "cat"). However, when both animals were on the card, children corrected the puppet by saying "Both!" or emphasizing *and* as in "cat AND dog!".

We performed chi-squared goodness-of-fit tests to compare the feedback distributions in the critical conditions with *and* and *or*. Here we focus on those trials (the four bar charts on the right of Figure 10). Children's linguistic feedback showed three patterns. First, the one-animal conjunctive and two-animal disjunctive (top left and bottom right) trials contained a higher proportion of corrections than the other trial types. These were trials where the guesses were either false or infelicitous. In the conjunction trials, a comparison of the feedback distribution in one-animal and two-animal conditions was statistically significant ($\chi^2(3, 83) = 201.65$, $p < .0001$), suggesting that children gave different feedback to true and false guesses. A similar numerical trend was present in the disjunction trials, but it was not significant ($\chi^2(9, 4) = 12$, $p = 0.21$).

Second, the one-animal disjunctive trials (top right) showed the highest proportion of "descriptions". These are trials in which the guess is correct but not specific enough: it leaves two possibilities open. These trials were significantly different from the one-animal trials for conjunction ($\chi^2(3, 83) = 62.16$, $p < .0001$). Finally, the two-animal conjunctive trials (bottom left) showed the highest proportion of "judgments" such as *You are right!*. This was not surprising given that these trials represented the optimal guessing scenario. These trials had a significantly different feedback distribution from the matching disjunction trials ($\chi^2(3, 84) = 184.98$, $p < .0001$).

Discussion

In study 2, we used a 3AFC judgment task to test children's comprehension of logical connectives *and* and *or*. We compared these results to those found in the 3AFC judgment task of study 1 with adults. The general comparison showed that adults and children had

similar patterns of judgments, except when both disjuncts were true. In such cases, adults judged the disjunctive guess as not completely right while most children judged it as completely right. There was even a slight preference among children to rewarded the puppet more in such cases, compared to cases of disjunction when only one disjunct was true.

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

The analysis of children's open-ended feedback raises two important issues. First, it runs counter to what the 3AFC judgment task suggests with respect to exclusivity implicatures. The forced-choice task suggests that children find such underinformative utterances as unproblematic while analysis of their spontaneous feedback shows that they provided more corrections to such utterances. Second, one of the explanations for why children fail to derive implicatures is that they cannot access the stronger alternative to the disjunction word *or*, namely *and* (Barner et al., 2011). However, in the context of the guessing game, some children explicitly mentioned the word *and*, as the word the puppet should have said instead of *or*. Interestingly, these children continued to reward the puppet and considered the guess "right", even though they corrected him. This raises the possibility that forced-choice truth value judgments underestimate children's pragmatic knowledge. In study 3, we used both a 2AFC truth judgment task and an analysis of children's open-ended

feedback. If the findings of study 2 were on the right track, we expected to replicate the same pattern in study 3, and find that children's open-ended feedback better reflects their sensitivity to pragmatic violations than the results of the 2AFC judgments.

Study 3: Children's 2AFC judgments and open-ended feedback

This study used the same paradigm as study 2 but focused on children's open-ended feedback and aimed at replicating the findings in study 2. The main hypothesis was that four-year-olds provide corrective feedback to the puppet if both disjuncts are true, but they do not consider this infelicity to be grave enough to render the guess itself "wrong" in a 2AFC judgment task. The main hypothesis along with relevant analyses and predictions were preregistered in an "As Predicted" format[6].

Methods

Table 6

Summary of Study 3 Methods

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

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. Study 3 was similar to Study 2 but differed in how children provided their judgments. Based on the findings in Study 2, we focused on verbal feedback, instead of rewards. We used two different ways of measuring children's judgments. First, we encouraged children to provide verbal feedback to the puppet. They were asked to say "yes" when the puppet was right and "no" when he was not right. They were also asked to help

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

him say it better. After children were done with this initial open-ended feedback, for each trial we asked a forced choice yes/no judgment 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 Study 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 7.

As in Study 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*[7].

Table 7

Instruction Trials for Study 3.

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

[7]The randomization code as well as the details of the methods are available on this paper's online repository.

Card	Guess	Response
DOG	there is a dog!	Yes!

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 Study 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.

Two-Alternative Forced Choice Judgments. Figure 11 shows children's 2AFC judgments. In the leftmost column, when the animal guessed was not on the card (e.g. elephant), children considered the guess "wrong". When the animal guessed was the only animal on the card (e.g. cat), children considered the guess "right". However, if the animal guessed (e.g. cat) was only one of the animals on the card, children were equally split between "wrong" and "right" judgments. On the other hand, almost all adults considered such guesses "right" in their 2AFC judgments (Figure 4). In such trial types, children seem to interpret the guess "there is a cat" as "there is **only** a cat", while adults do not. This difference between children and adults is unexpected for a theory of meaning acquisition that assumes children are overall more logical or literal as interpreters than adults (Noveck, 2001).

In the trials with *and* and *or*, children's judgments were similar to those of adults. Figure 12 compares adults' and children's 2AFC judgments. In trials with conjunction, when only one of the animals was on the card, most children considered the guess "wrong". This is similar to adults' judgments, but different in extent: adults were more consistent and

unanimous in rejecting such guesses. A mixed effects logistic regression with the fixed effect of age category (adult vs. child) and random effect of subject found no significant difference between adults' and children's responses in such trials (see Table 8, Conjunction - One Animal).

Table 8

Mixed effects logistic models for conjunction and disjunction trials when only one disjunct was true, in 2AFC judgments of adults and children, using `glmer` *in R's lme4 package. Formula:* $Response \sim AgeCategory + (1|Subject)$.

Trial Data	Coefficient	Standard Error	Z-Value	P-value
Conjunction - One Animal	-2.05	2.86	-0.72	0.47
Disjunction - One Animal	1.34	1.79	0.75	0.45

In conjunctive guesses where both animals were on the card, both children and adults were unanimous in considering the guess "right". In disjunctive trials when only one of the animals was on the card, 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". Yet again, a mixed effects logistic regression with the fixed effect of age (adult vs. child) and random effect of subject found no significant difference between adults' and children's responses in such trials (see Table 8, Disjunction - One Animal). Adults and children showed almost identical patterns of judgments in trials where there was two animals on the card and the guess used the connective *or*. Children and adults did not differ in their rate of rejecting disjunctive guesses when both disjuncts were true.

Finally, there is a small but significant preference in children's judgments of disjunctive statements for both disjuncts to be true. Comparing the disjunctive trials with one animal and two animals on the card, a mixed-effects logistic model with the fixed effect of disjunction type and the random effect of subjects found that children had a slight preference for both animals to be on the card (b= 1.85, se= 0.56, z= 3.32, $p < 0.001$).

There was a similar small trend in children's three-alternative judgments in study 2. While this was quite small compared to the other effects observed in these studies, it nevertheless indicated a difference between children's and adults' truth judgments. We return to this in more detail in section **??** of the General Discussion.

Open-ended Feedback. Figure 13 shows the distribution of children's feedback to the puppet in Study 3 (see Table 5 for the definitions and examples of feedback categories). There were no "None" responses in this study since the experimenter explicitly asked children to provide feedback to the puppet. The distribution of the responses in the other three categories (Judgment, Description, and Correction) revealed a successful replication of Study 2.

Children's feedback showed four main patterns. First when the puppet guessed an animal not on the card (e.g. "There is an elephant!"), there is a split pattern between negative judgments like "No!" and simply mentioning the animal on the card (e.g. "Cat!"). Children provided no corrections on such trials, at least the way we have defined them. Second, almost all children responded with positive judgments like "Yes!" when the puppet's guess accurately matched what was on the card. This was the case in trials where there was only one animal on the card (e.g. cat) and the puppet mentioned it (e.g. "There is a cat!"), as well as trials where there were two animals on the card and the puppet mentioned both with a conjunction (e.g. "There is a cat and a dog!"). Third, children provided the largest number of corrective feedback in trials where the guess was either false or infelicitous. These included three trial types: (a) the ones where there were two animals on the card (e.g. cat and dog) but the puppet only guessed one (e.g. "There is a cat!"); (b) the ones where the puppet guessed two animals with conjunction (e.g. "There is a cat and a dog!") but only one of them was on the card (e.g. cat); and (c) the ones where there were two animals on the card (e.g. cat and dog), and the puppet guessed both but used a disjunction (e.g. "There is a cat or a dog!"). Finally, there was a pattern of feedback unique to disjunctive trials (e.g. "There is a cat or a dog!") with only one animal on the card (e.g. cat). In such cases, almost

all children simply named the animal on the card (e.g. “Cat!”).

Figure 14 breaks down children’s open-ended feedback based on whether children said *Yes!*, *No!*, or said something else. Responses that were not yes/no judgments are grouped in a middle category shown with a dash. The goal here is to compare children’s open-ended judgments with their forced choice judgments shown in Figure 11. Children’s open-ended judgments and their forced choice judgments in study 3 show similar patterns for all types of guesses except for disjunctive ones. In trials that the puppet guessed a disjunction, the vast majority of children refused to provide a yes/no judgment when they were not forced to. Instead, they described the animal on the card or provided corrections to the puppet’s infelicitous disjunctive guess.

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

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

animal (e.g. dog) should have been excluded. When two animals were on the card (e.g. cat and dog) and the puppet used a disjunctive guess (e.g."There is a cat or a dog!"), or simple guess (e.g."There is a cat¡'), children provided inclusive feedback, 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 on the card.

Discussion

Study 3 measured children's comprehension of logical connectives in two ways: First, with analyzing their open-ended feedback and second, with a two-alternative forced choice task. First, we asked children to say *yes* to the puppet if he was right and *no* if he was wrong. However, children could provide any form of feedback they wanted. Second, we followed children's open-ended feedback with a 2AFC question: "Was the puppet right?" This way, we could measure children's comprehension in two different ways in the same trial. Ideally, both measures should show similar results. However, the findings were similar for conjunctive guesses, but not disjunctive ones. Children avoided binary right/wrong feedback with disjunction and preferred to provide more nuanced feedback.

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

children stressed *and* or *both*, implying that both animals should have been included.

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

General Discussion

We reported three studies on adults' and preschool children's comprehension of the logical connectives *and* and *or*. The first study used two- and three-alternative forced choice judgment tasks with adults. In the 2AFC task, adult interpretations closely matched the semantic accounts of *and* and *or* as conjunction and inclusive disjunction. The 2AFC judgments did not register robust signs of pragmatic infelicities. However, the 3AFC judgments showed signs of pragmatic infelicities, especially in disjunctive guesses with true disjuncts. When two animals where on the card (e.g. cat and dog) and the guess used *or* (e.g. *There is a cat or a dog!*), participants were more likely to choose "kinda right" rather than "right".

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

In the third study, we focused on eliciting open-ended verbal feedback from children and followed it with a 2AFC question. Children's 2AFC responses reflected the semantics of

the connectives *and* and *or* as conjunction and inclusive disjunction. There was no significant difference between children and adults in the 2AFC task. Analysis of children's open-ended feedback replicated the findings in study 2. Children provided more corrective feedback in false and pragmatically infelicitous trials with the connectives than in felicitous trials. The comparison of the 2AFC task and children's open-ended responses showed that children are sensitive to the infelicity of disjunctions with true disjuncts, even though they consider them to be "right" guesses.

Previous studies had suggested that adults and preschool children differ in their interpretation of disjunction in two ways. First, unlike adults, children might interpret a disjunction as conjunction (Singh et al., 2016; Tieu et al., 2016). Second, children might interpret *or* as inclusive disjunction when adults interpret it as exclusive (Crain, 2012). The studies reported here provide evidence for the hypothesis that these differences may be an artifact of the experimental task and the type of measurement (Skordos et al. (2018), Katsos (2014)).

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

We would like to add that conjunctive interpretations of disjunction, even when robustly observed, can have at least two potential explanations. First, non-linguistic interpretive strategies and preferences, due to task demands or unknown connective meaning

(Clark, 1973; Paris, 1973), and second, pragmatic enrichment, common in free-choice contexts (Singh et al., 2016; Tieu et al., 2016). As explained in section , previous research provides substantial evidence for task-related increase in conjunctive readings of disjunction (Braine & Rumain, 1981; Neimark & Slotnick, 1970; Paris, 1973; Skordos et al., 2018). In order to show instances of pragmatically enriched conjunctive readings in preschool children, it is important to first rule out task-related conjunctive interpretations.

Considering the second difference, namely the lower rate of exclusivity inferences in preschool children, our studies provided evidence that the choice of measurement may play an important role. In the 3AFC judgment task when two animals were on the card (e.g. card: cat and dog, guess: "There is a cat or a dog"), adults were more likely to choose "kinda right" than children were. Children mostly chose "right". However, in their free-form feedback, children corrected such utterances and suggested that the connective *and* should have been used instead of *or*.

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

1. Processing difficulty. First, processing accounts locate the problem in children's processing capacities such as working memory. They suggest that pragmatic computations are cognitively taxing and children lack the appropriate processing resources to carry them out appropriately. A prediction of processing accounts (at least in their current format) is that children will show reduced implicature computations for all types of implicatures – scalar or not. This prediction was not borne out in our experimental results here. In Study

3, children were much more likely than adults to call a simple guess (e.g. *There is a cat!*) "wrong" if there were two animals on the card (e.g. cat and dog). In other words, children's interpretations were much more exhaustive than adults. Processing accounts do not predict that children may derive implicatures at a higher rate than adults but this is what we found, at least for the traditional interpretation of the judgment task.

2. Non-adult-like Lexicon. Several proposals blame the structure of the child's lexicon for the alleged failure in deriving implicatures. The assumption is that the child's lexical entry for scalar items must include three elements for successful derivation: 1. the semantics of the weak term (e.g. *some*, *or*) 2. the semantics of the strong term (e.g. *all*,*and*); and possibly 3. a scale that recognizes the stronger term as an alternative to the weaker one (e.g. <*some*, *all*>, <*or*, *and*>). Each of these elements have been pinpointed as the source of the problem in previous studies (Barner et al., 2011; Horowitz et al., 2017; Katsos & Bishop, 2011). However none of them seem to apply to the results reported here.

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

3. Pragmatic Tolerance. Katsos & Bishop (2011) suggested that children tend to

tolerate pragmatic infelicities more than adults. They showed that when children were provided with a 2AFC judgment task, they considered a description with the scalar term *some* as "right" when *all* was more informative (e.g. *The turtle played with some of the balls.*, Scene: the turtle played with all the balls.) However, when they are presented with three options (small, big, and huge strawberries) in a 3AFC task, they choose the middle option in the same type of trials. They argued that children tolerate pragmatic infelicities and do not regard them as "wrong". As in a processing account, the tolerance account predicts that scalar and ad-hoc implicatures will be similarly affected. However, our results did not match those of Katsos & Bishop (2011). When children were presented with a 3AFC task, they chose the highest reward (and not the middle option) for uses of *or* when *and* was more informative. Second, and more importantly, we found different patterns for exhaustive and scalar inferences as mentioned before. This is not predicted by the tolerance account unless we assume that children are more tolerant towards violations of scalar inferences than they are towards exhaustive ones. While this is not currently assumed in the literature, it is a possible adjustment. However, we would like address this issue by focusing on another related factor: the role of measurement in estimates of children's pragmatic capacity (Katsos, 2014).

4. The Role of Measurment. Two observations in the current studies provide support for the hypothesis that methodological issues, and more specifically issues of measurement contribute to the differences found between adults and children in pragmatic capacity. First, Study 1 showed that even for adults, the estimates of adult infelicity rates may differ based on the number of alternatives in the forced choice task. A 2AFC task underestimated adults' sensitivity to pragmatic infelicity. In fact, in a follow up study, we systematically varied the number of response options and replicated the results presented here (see Jasbi, Waldon, and Degen in press). Second, children's open-ended linguistic feedback in the experimental context better reflected their sensitivity to pragmatic nuances than the forced-choice judgment tasks. Third, children showed a higher rate of "wrong"

judgments for cases of exhaustive inferences (simple guesses with two animals on the card) than adults did. While a difference in sensitivity to ad-hoc vs. scalar implicatures has been reported and argued for before (Horowitz et al., 2017; Stiller, Goodman, & Frank, 2015), a higher sensitivity than adults is not predicted by any of the current accounts.

Figure 16 shows a summary of the factors that are proposed to affect pragmatic computations. As Pouscoulous & Noveck (2009) and Katsos (2014) have suggested, the central issue is "the rate" at which children and adults manifest pragmatic reasoning in the experimental setting. No one doubts children's capacity to perform such computations. At issue is the extent to which children and adults compute specific implicatures. As Katsos (2014) pointed out, it seems reasonable to assume that all these factors play some part here. What matters is the degree to which each contributes to the outcome.

The results of the studies reported here suggest that it is important to distinguish between factors that affect pragmatic computations and those that affect the observed "rate" in an experimental setting. As we showed in Study 1, given the number of alternatives in the forced choice task (2AFC vs. 3AFC), we may get different estimates of adults' rate of infelicity judgments, but we cannot assume that there is a difference in adults' pragmatic capacities in these two tasks. A similar situation exists when we compare children's forced choice measures of infelicity and their open-ended feedback.

In order to better understand the differences between adults and children's semantic and pragmatic capacities, it is necessary to have a good understanding of how our measurements affect estimates of adults and children's performance in the experimental tasks. Children may be no more capable of making exhaustive inferences than adults and no less capable of making scalar inferences either. They may simply have a different construal of the wrong-right scale and of what the forced-choice task is about. The concepts "right" and "wrong" are as much subject to developmental change and differences between adults and children as are scalar items that constitute the focus of our studies. Relying on a single type of measurement increases the risk of measurement-specific conclusions. Using multiple

measurements in the same task can provide converging evidence for felicity/infelicity or presence/absence of specific inferences. Ultimately, in order to capture semantic and pragmatic competences of adults and (especially) children, we need to develop methods that can reliably tap into specific dimensions of meaning.

Conclusion

We provided three studies that tested adults and children's comprehension of disjunction in existential sentences using three different measures: binary forced-choice truth value judgments (2AFC task), ternary forced-choice truth value judgments (3AFC task), and free-form verbal feedback. The results suggested that for each population, different measures were sensitive to different aspects of meaning. The binary measure captured children and adults intutions about truth values well: it showed that they considered a disjunction as inclusive in existential sentences of the guessing game. Ternary judgments provided evidence for adults' pragmatic inferences: adults often considered a disjunction when both disjuncts were true as "kinda right" and not completely right. For children, the ternary judgments did not register such an effect, but their free-form verbal feedback did. When both disjuncts were true, children verbally corrected the puppet and suggested that he should have said *and* instead of *or*. The combination of children's truth valued judgments and their verbal feedback suggested that on average, children in our sample understood that when both propositions were true, their conjunction and disjunction were true, yet conjunction made a more appropriate and felicitous utterance.

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 aspected of langauge and connected to distinct processes that generate them. Therefore, while the inclusive interpretation is hypothesized to be part of *or*'s semantics, 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 of experimental study, we also need to develop measures especially suited to capture the aspect of meaning under investigation.

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.

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.

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., 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., 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. (2012). *The emergence of meaning*. Cambridge: Cambridge University Press.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

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

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.

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, MA: Cascadilla Press.

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

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

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

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

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

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

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

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

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

Nitta, N., & Nagano, S. (1966). Basic logical operations and their verbal expressions: Child's 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. (2012). 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. (2012). Children's interpretation of disjunction

974 in the scope of "before": A comparison of English and Mandarin. *Journal of Child*
975 *Language*, *39*(03), 482–522.

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

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

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

982 Pouscoulous, N., & Noveck, I. A. (2009). Going beyond semantics: The development of
983 pragmatic enrichment. In S. Foster-Cohen (Ed.), *Language acquisition* (pp. 196–215).
984 Berlin: Springer.

985 Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental
986 investigation of processing costs in implicature production. *Language Acquisition*,
987 *14*(4), 347–375.

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

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

993 Skordos, D., Feiman, R., Bale, A., & Barner, D. (2018, July). *Do children interpret "or"*
994 *conjunctively?* Retrieved from https://osf.io/2srxk/

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

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

999 Su, Y., & Crain, S. (2013). Disjunction and universal quantification in child mandarin.

Language and Linguistics, *14*(3), 599–631.

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

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

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

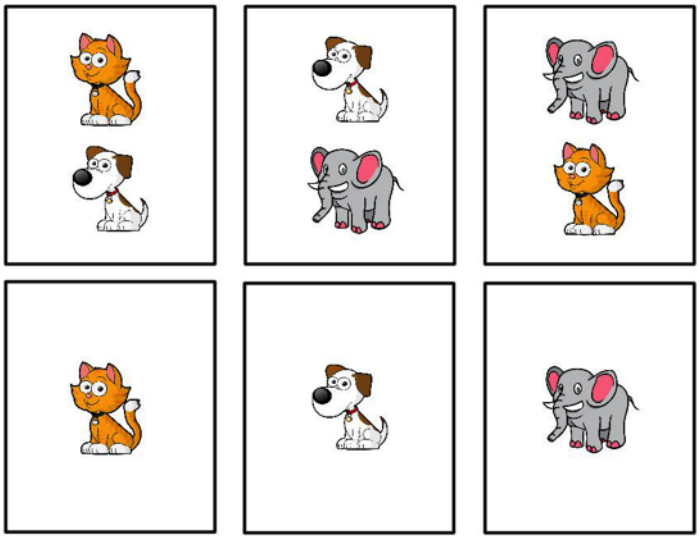

Figure 1. Cards used in the connective guessing game.

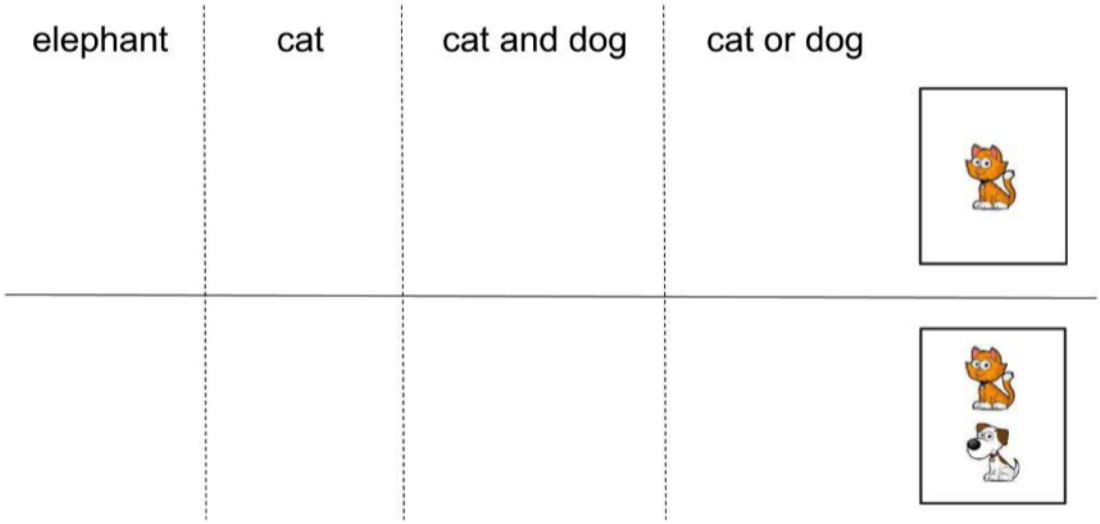

Figure 2. Trial types represented by example cards and example guesses.

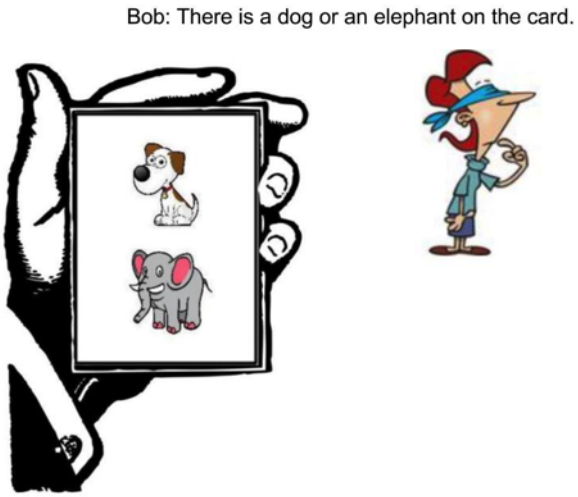

Figure 3. An example trial in Study 1.

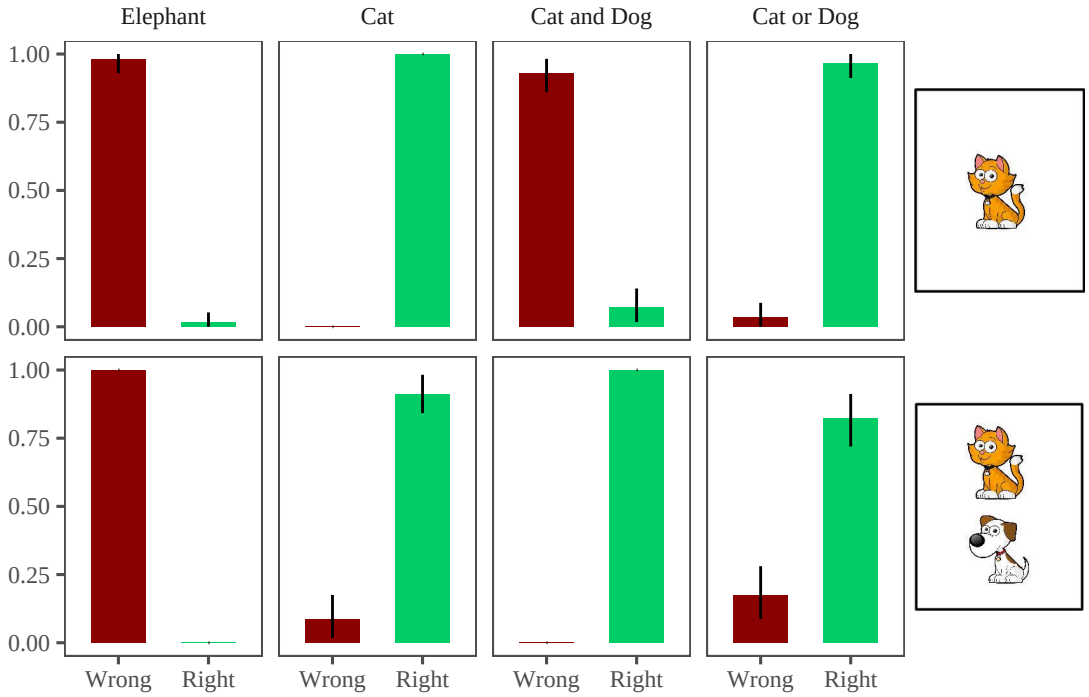

Figure 4. Adults' two-alternative forced choice judgments.

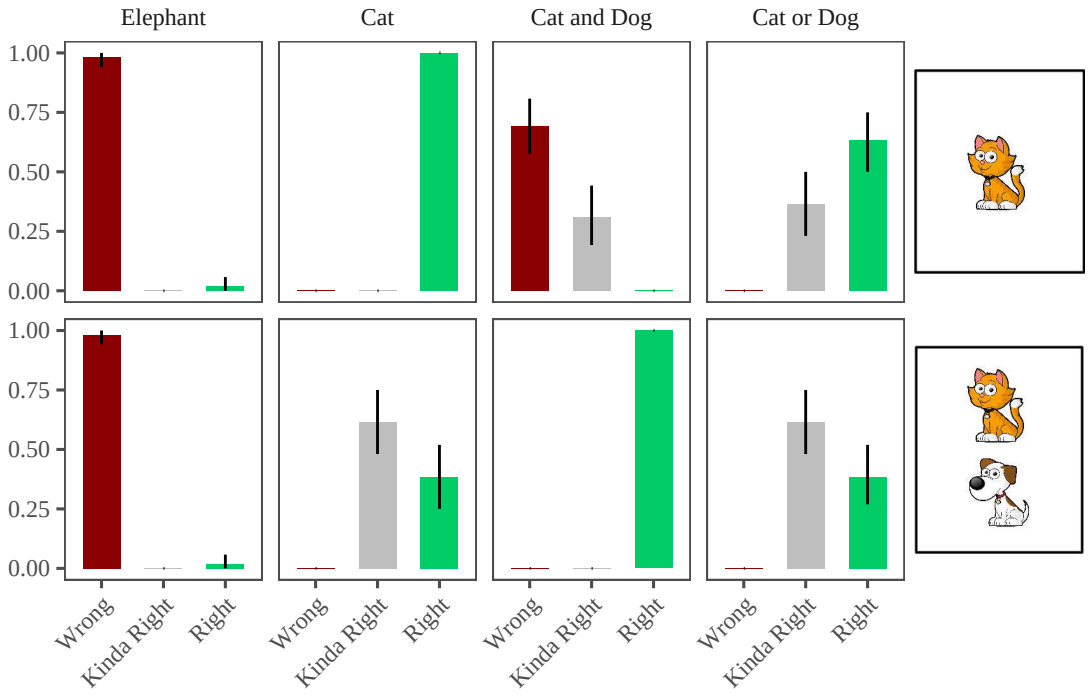

Figure 5. Adults' three-alternative forced choice judgments in the connective guessing game.

Figure 6. The puppet, Jazzy, with and without the sleeping mask.

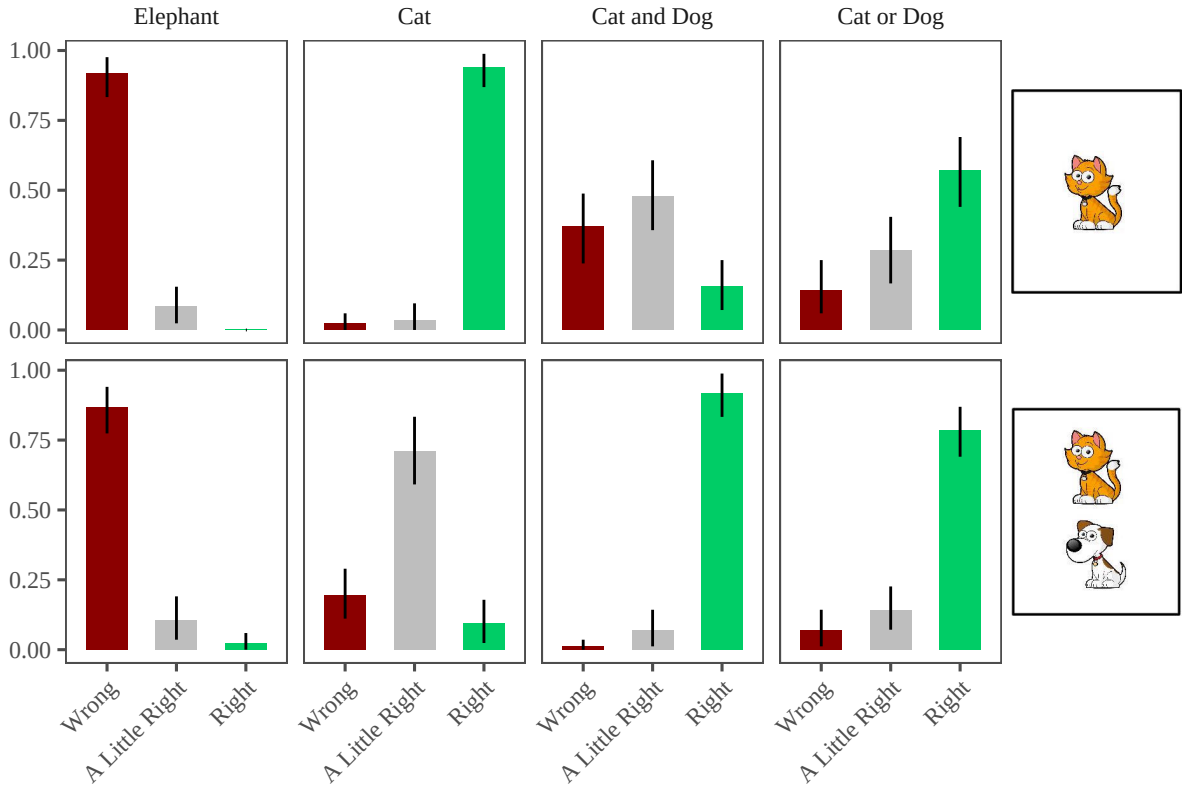

Figure 7. Children's 3AFC judgments in the connective guessing game.

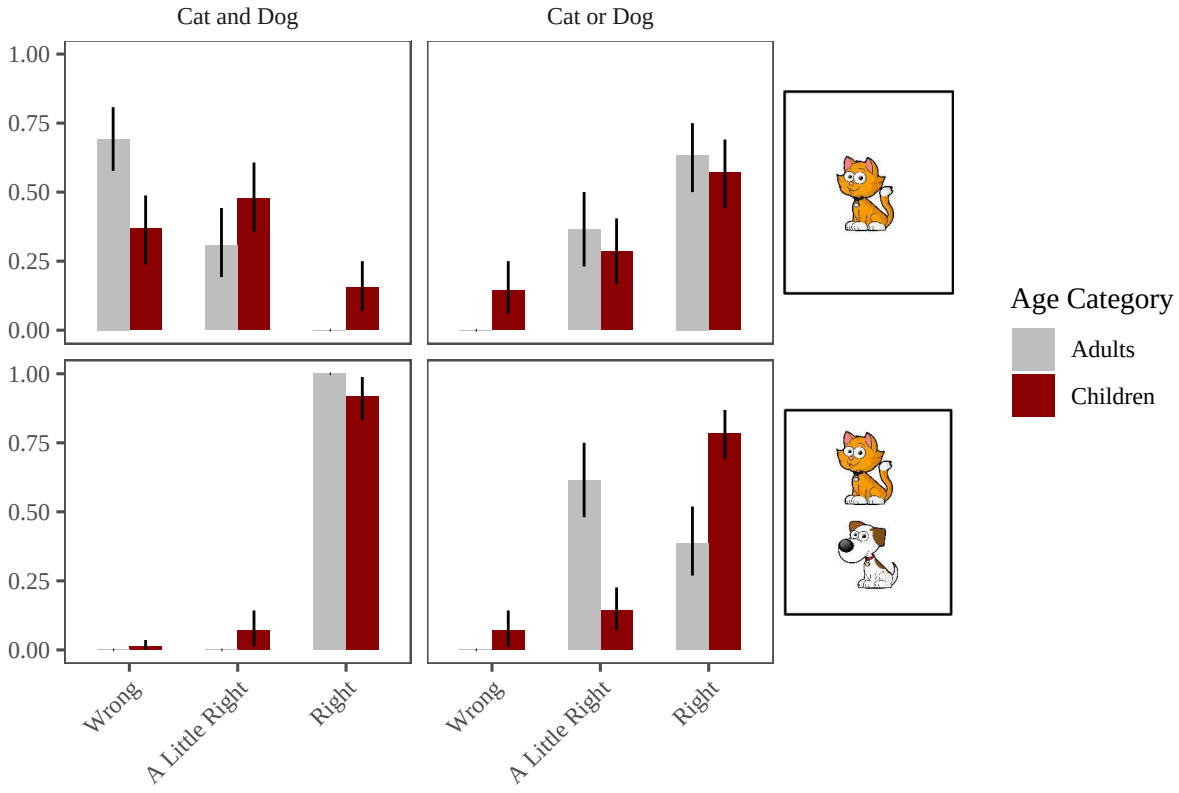

Figure 8. Comparison of Adults' and Children's 3AFC judgments.

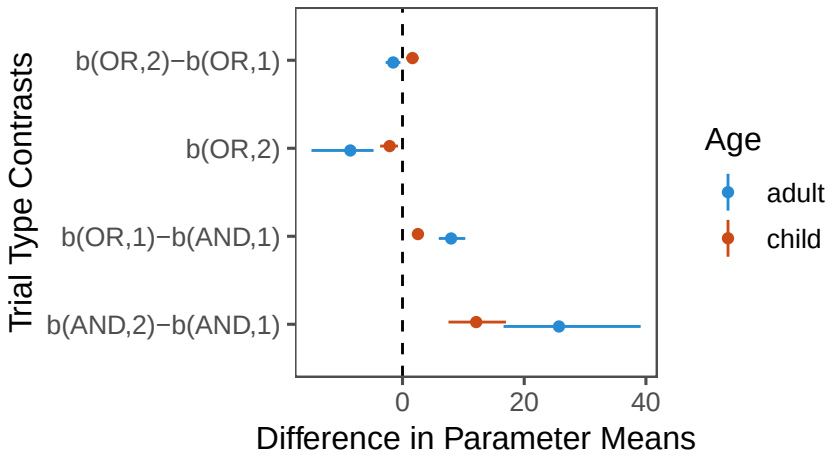

Figure 9. Coefficients capturing the relevant comparisons across conditions in 3AFC judgments in Study 1 and 2. In naming the coefficients like b(OR,2), OR/AND represents the connective used and the number 1/2 represents the number of animals on the card. Error bars represent 99% regions of highest posterior density.

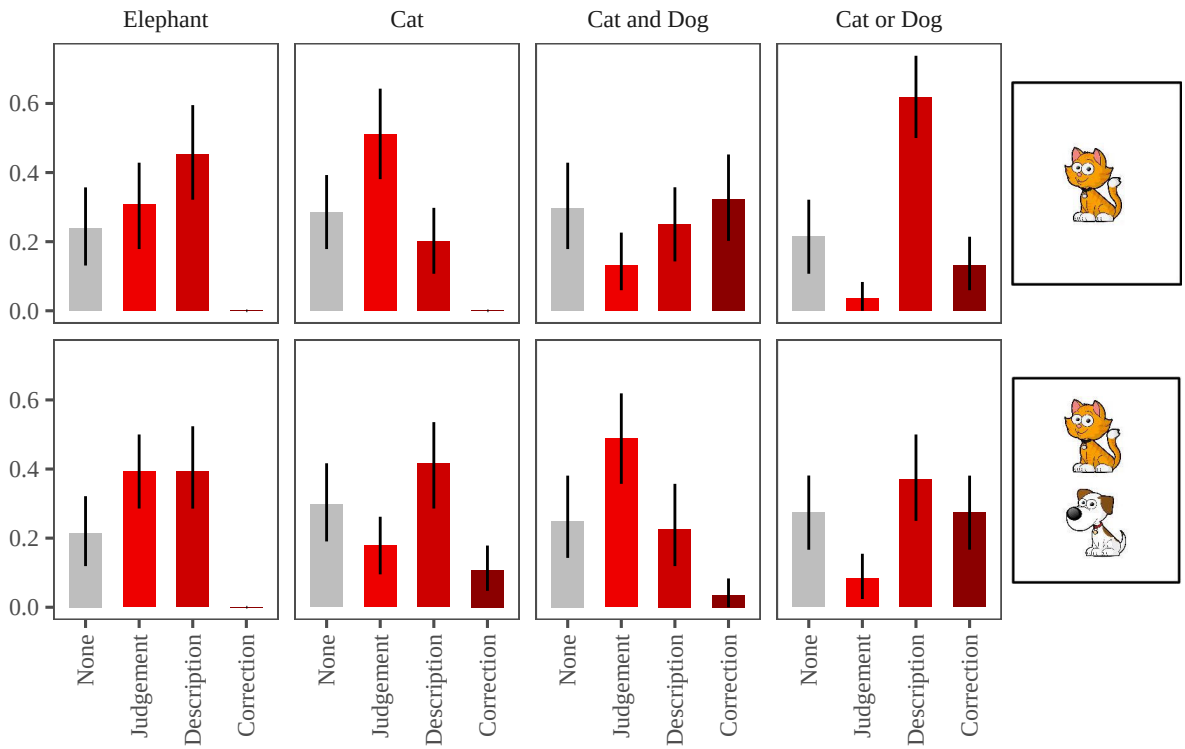

Figure 10. Children's open-ended Feedback. Error bars represent 95% confidence intervals.

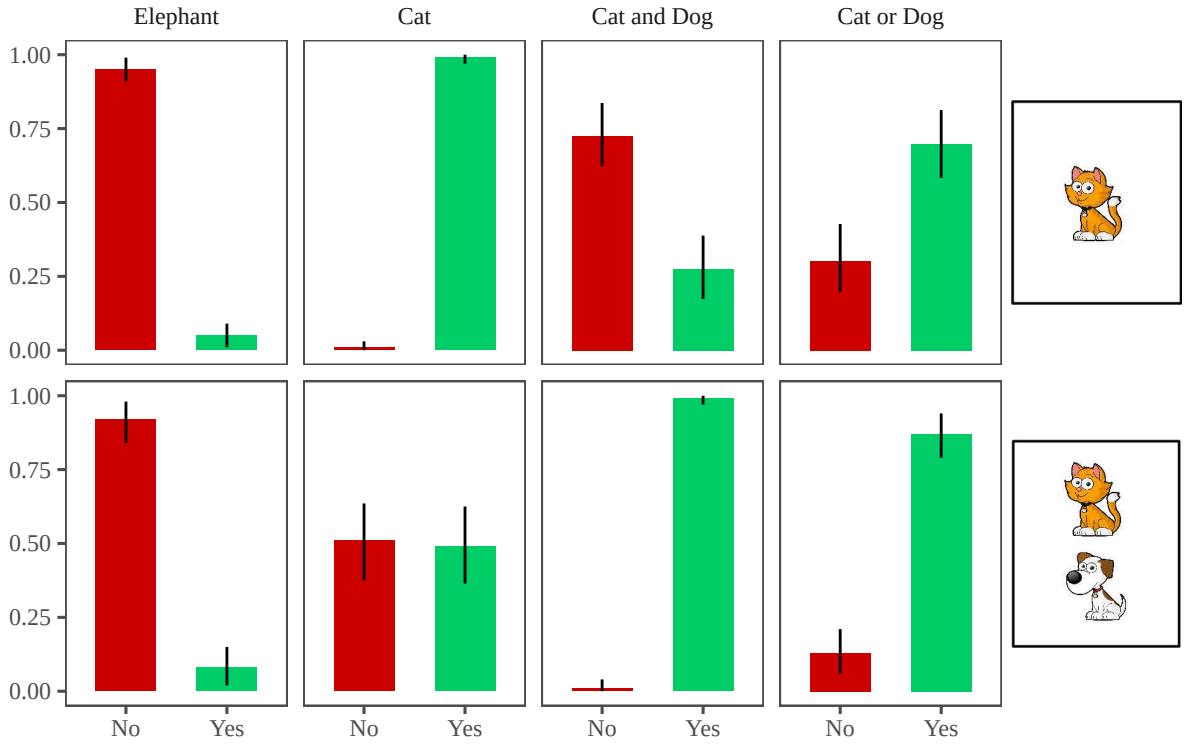

Figure 11. Children's binary truth value judgments.

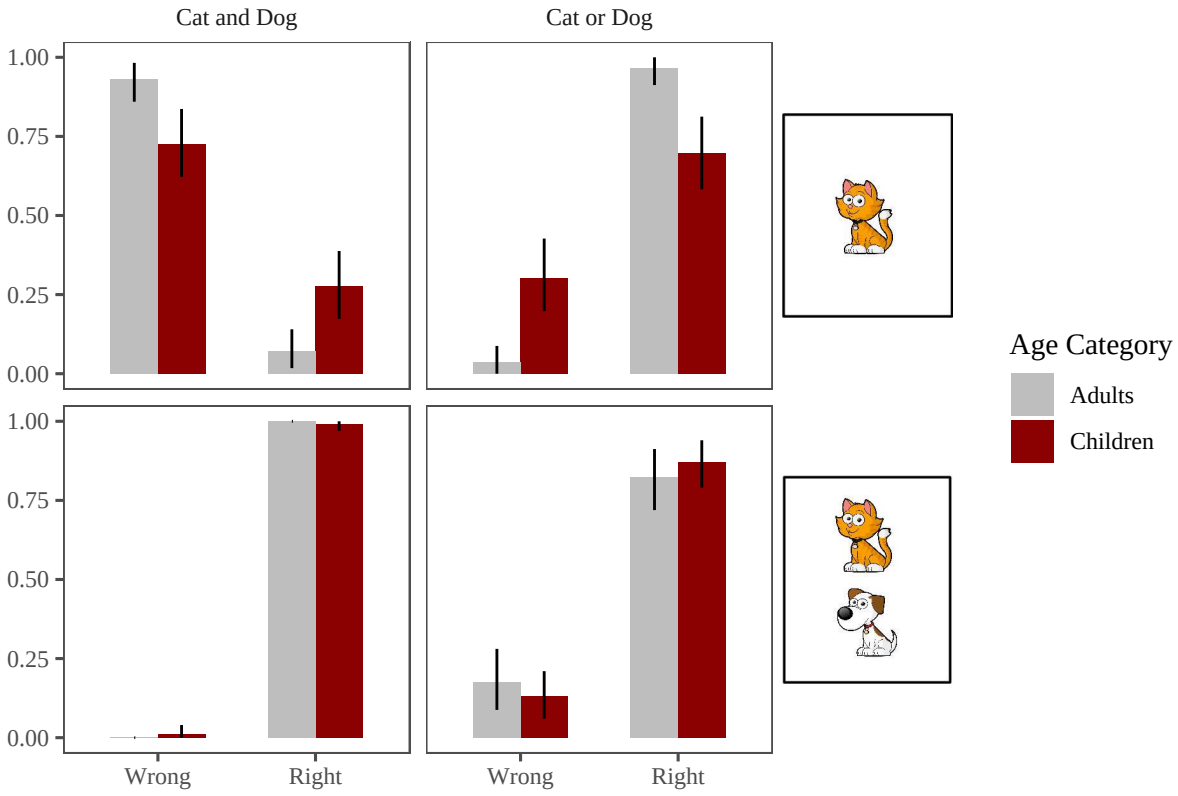

Figure 12. The comparison of the 2AFC judgment task for conjunction and disjunction trials in adults (study 1) and children (study 3).

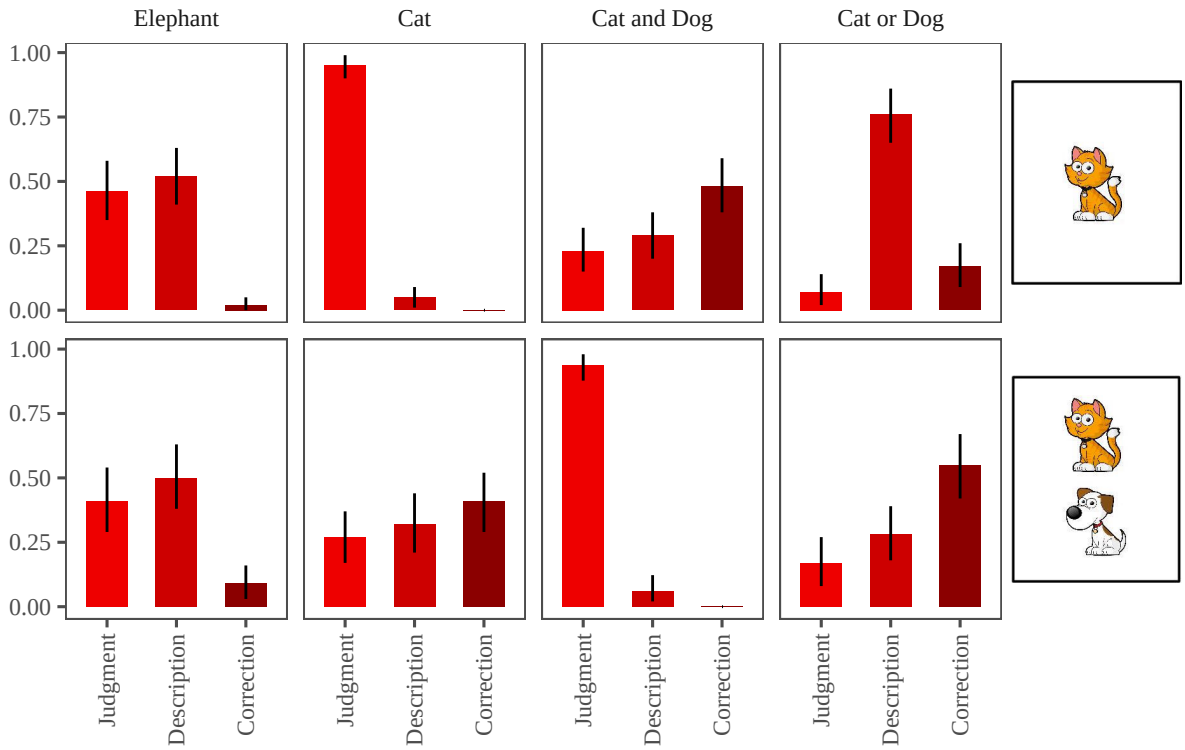

Figure 13. Children's Open-ended Feedback in Study 3. Error bars represent 95% confidence intervals.

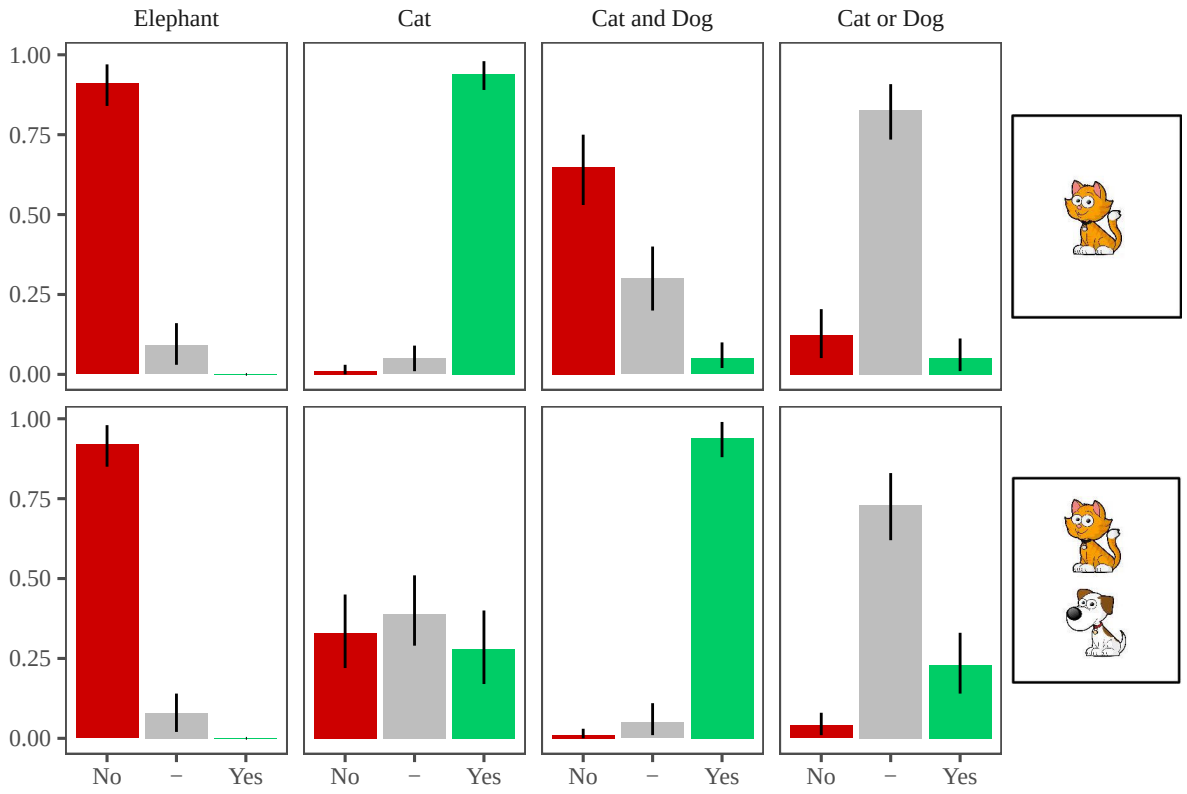

Figure 14. Children's open-ended feedback to the puppet's guesses. The x-axis shows whether children spontaneously provided a yes (green), no (red), or other response (grey).

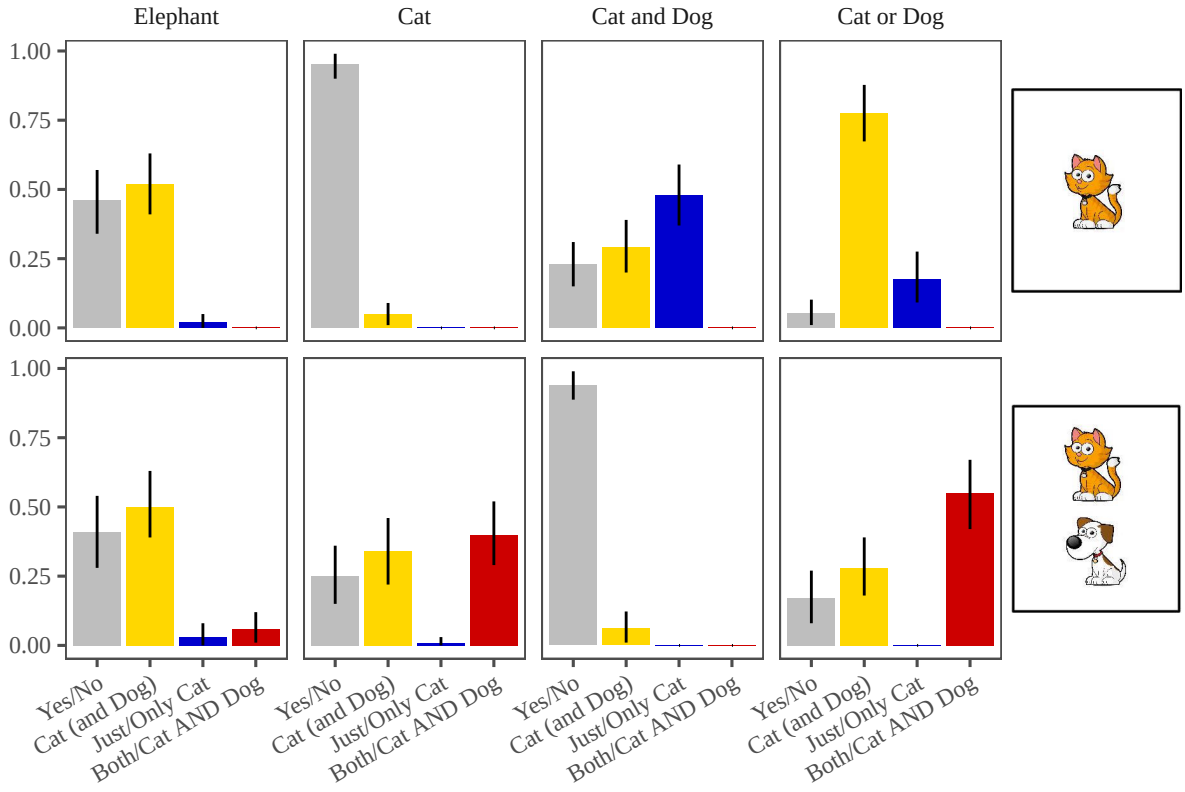

Figure 15. Children's feedback categories in disjunction trials.

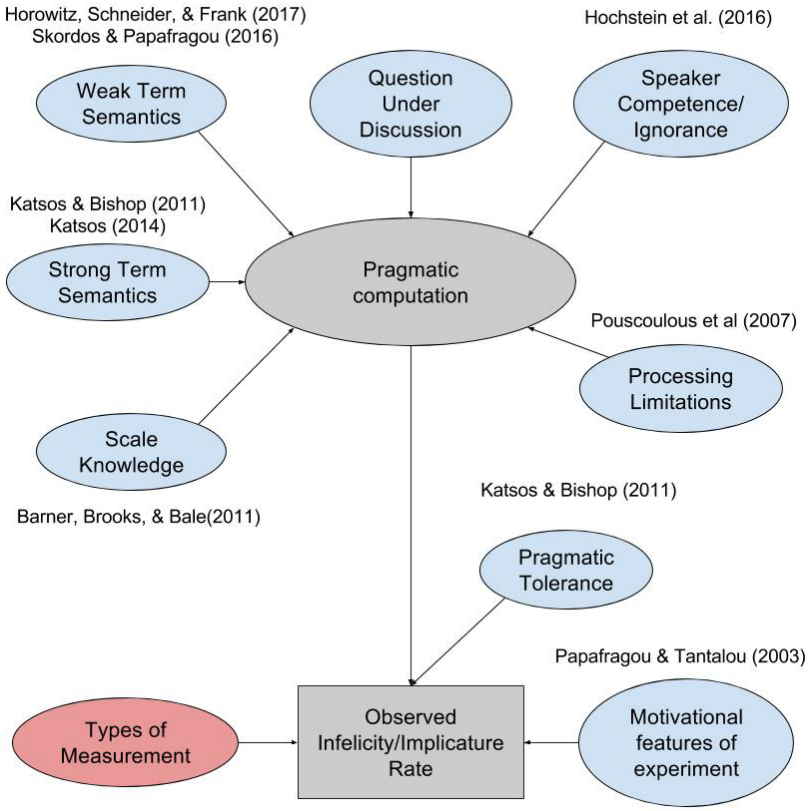

Figure 16. Factors that could affect pragmatic computations and the estimates of these computations in the experimental settings