Developmental Psychology

Adults’ and Children’s Comprehension of Linguistic Disjunction

Masoud Jasbi, Michael C. Frank

Linguistics, University of California, Davis, CA, US, Psychology, Stanford University, Stanford, CA, US

Keywords: disjunction, semantics, pragmatics, logic, language acquisition

Disjunction has played a major role in advancing theories of logic, language, and cognition, featuring as the centerpiece of debates on the origins and development of logical thought. Recent studies have argued that due to non-adult-like pragmatic reasoning, preschool children’s comprehension of linguistic disjunction differs from adults in two ways. First, children are more likely to interpret “or” as “and” (conjunctive interpretations); Second, children are more likely to consider a disjunction as inclusive (lack of exclusivity implicatures). We tested adults and children’s comprehension of disjunction in existential sentences using two and three-alternative forced choice tasks, and analyzed children’s spontaneous verbal reactions prior to their forced-choice judgments. Overall our results are compatible with studies that suggest children understand the basic truth-conditional semantics of disjunction. Children did not interpret “or” as “and”, supporting studies that argue conjunctive interpretations are due to task demands. In addition, even though our forced-choice tasks suggest children interpreted disjunction as inclusive, spontaneous verbal reactions showed that children were sensitive to the adult-like pragmatics of disjunction. Theoretically, these studies provide evidence against previous developmental accounts, and lend themselves to two alternative hypotheses. First, that preschool children’s pragmatic knowledge is more adult-like than previously assumed, but forced-choice judgments are not sensitive enough to capture this knowledge. Second, children may have the knowledge of the relevant lexical scale themselves, but be uncertain whether a new speaker also has this knowledge (mutual knowledge of the scale).

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

**SPEAKER IGNORANCE** is the label we use today for the implication that the speaker does not know which disjunction is true. Tarski also noted that a disjunction has at least two other implications: exclusivity and inclusivity. Suppose, “a child has asked to be taken on a hike in the morning and to a theater in the afternoon, and we reply: No, we shall go on a hike or we shall go to the theater” (Tarski, 1941, p. 20). Tarski explained that disjunction in this example is **EXCLUSIVE** because “we intend to comply with only one of the two requests” and not both. However, a disjunction may also have an **INCLUSIVE** implication like the following example: “Customers who are teachers or college students are entitled to a special reduction”. Tarski explained that *or* in this example is inclusive “since it is not intended to refuse reduction to a teacher who is at the same time a college student.”

Grice (1975) argued that different implications of disjunction have different sources, strengths, and theoretical status. The inclusive implication is the literal meaning encoded by the word *or* (i.e. its semantics). It is in fact a strong entailment of disjunction, whose denial would result in a contradiction (e.g. "Bob drank tea or coffee. He drank neither in fact!"). On the other hand, exclusivity and ignorance are weaker inferences that enrich the literal meaning of *or* in context (i.e. its pragmatics). Grice called these pragmatic inferences **IMPICATURES** and considered them deniable or defeasible (e.g. "Bob drank tea or coffee. He drank both in
forced-choice vs. free-form spontaneous responses. Finally in General Discussion, we discuss the implications of our studies for theories of semantic and pragmatic development.

Previous Research

Children’s comprehension of logical connectives and or have been studied within two research programs. The first program, starting in the 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 the late 1990s and was inspired by the Gricean theory of meaning. Rather than emphasizing 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.

Within the Piagetian program, researchers hypothesized that the abstract and logical notion of disjunction (i.e. inclusive disjunction) is constructed from the more concrete concept of “choice between two options”. The prediction was that until the age of 11 (concrete operational stage), children understand a disjunction like "A or B" as "one of the two options". This is similar to the exclusive implication of 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 & Nagan, 1966). Participants were presented with pictures of objects and asked to circle those described by a statement such as "not bird", "bird and white", "bird or white". These studies concluded that the majority of the participants understood negation and conjunction, but only college students correctly answered statements with disjunction. They reported that participants made two types of “errors”. First across all ages, some participants interpreted disjunction as conjunction. Second, some participants interpreted disjunction as exclusive. Based on these results, Neimark (1970) concluded that a “correct” (i.e. inclusive) understanding of disjunction only develops in the high school years and depends on the attainment of formal operations as defined in the Piagetian theory.1

Further investigations suggested that the conjunctive errors may be due to the task design of in-class tests. Paris (1973) reported that in his in-class truth-judgment task, even a fifth of college students did not differentiate or from and, interpreting both as conjunction. He attributed these conjunctive interpretations of or to the application of non-linguistic strategies when the task is difficult or confusing (See Clark, 1973 for a discussion of non-linguistic strategies in child language acquisition). He explained that children in his task (as well as some adults) were probably “comparing visual and auditory information with little regard for the implied logical relationship in the verbal description.” In a

---

1 Following the convention in the language acquisition literature, we indicate children’s age using the "Year:Month" format.
disjunction such as “A or B”, participants responded with “true” if the individual disjuncts (A, B) matched the pictures and false otherwise. Such a non-linguistic “label-matching” strategy would yield correct answers for conjunction but incorrect (conjunctive) answers for disjunction. This account also explains why in Paris (1973)’s study, conjunctive readings reduced with age and why using the word either along with or helped reduce conjunctive interpretations further (presumably by adding an additional linguistic cue that further differentiated conjunction and disjunction).

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

Braine and Rumain (1981) tested participants with both a simplified replication of Suppes and Feldman (1969)’s “give-item” task and a version of what is today known as the truth value judgment task. For their replication of Suppes and Feldman (1969), they reported that both children and adults provided a “choose-one” (i.e. exclusive) interpretation of disjunction. They did not find any conjunctive interpretations, providing even further support for the role of task design. However, this was not the case in the truth value judgment task. In this task, a puppet described the contents of four boxes, each containing four animal toys. For example, the puppet said “Either there is a horse or a duck in the box.” The first box had both animals, the second had only a horse, the third only a duck, and the last had neither. Participants were asked if the puppet was right. The results showed that adults were split between an inclusive and an exclusive interpretation of disjunction. The 7 to 10 year-olds were more likely to consider the disjunction as inclusive. However, the youngest group (5-6 years old) was most likely to interpret a disjunction similar to a conjunction: they said the puppet was right when both animals were in the box and not right or partly right if only one of the animals was in the box. Following Paris (1973), Braine and Rumain (1981) argued that in this task, younger children do not take the contribution of the connective or into account. Instead, they use a non-linguistic strategy in which the disjunction is right if both propositions are true, partly right if only one is true, and wrong if neither is true. Braine and Rumain (1981) concluded that children’s ability to interpret a disjunction in a command develops earlier than their ability to judge its truth value. In Braine and Rumain (1981)’s judgment task, the puppet uttered a disjunction even though the content of the box was known to both the puppet and the participant (i.e. the speaker lacked ignorance). As Tarski (1941) noted, such uses of disjunction sound odd and infelicitous. This feature may have contributed to the application of a non-linguistic strategy and resulted in conjunctive readings. Later truth value judgment studies such as Chierchia, Crain, Guasti, and Thornton (1998) controlled for this effect of disjunction by making the puppet utter disjunction as a prediction or guess for an unknown event, and let participants judge the prediction after they see the outcome. Chierchia et al. (1998) argued that in order to truly capture children’s semantic competence with or, experiments need to test its comprehension in contexts that do not invite exclusivity implicatures. These contexts include embedding or under linguistic operators such as negation or conditionals.

Since Chierchia et al. (1998)’s arguments, numerous studies within the Gricean program have tested preschool children’s comprehension of disjunction in embedded contexts as varied as negative sentences (Crain et al., 2000), conditional sentences (Gualmini et al., 2000), restriction and nuclear scope of the universal quantifier every (Chierchia et al., 2001, 2004), nuclear scope of the negative quantifier none (Gualmini & Crain, 2002), restriction and nuclear scope of not every (Notley, Thornton, et al., 2012), and prepositional phrases headed by before (Notley, Zhou, et al., 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 preschool children understand the semantics of disjunction and that the inclusive implication emerges earlier than the exclusive implication. However, the lack of exclusivity in early childhood stands in sharp contrast to the earlier conclusions from the give-item tasks. Since under the Gricean account, exclusivity is the result of pragmatic (scalar) implicatures, the lack of exclusivity is considered consistent with evidence from the development of quantifier implicatures some and all and the overall conclusion that implicatures are slow to develop in children (Barner et al., 2011; Noveck, 2001; Papafragou & Musolino, 2003). Methodological issues qualify this conclusion, however. As mentioned earlier, Braine and Rumain (1981) found that the same children were more likely to interpret a disjunction as exclusive in a give-item task and inclusive/conjunctive in a truth value judgment task. Therefore, truth value judgment tasks may not reveal the full picture regarding children’s knowledge of exclusivity implicatures.

More recently, two truth value judgment studies reported that the majority of preschool children in their sample interpreted a disjunction similar to a conjunction (Singh et al., 2016; Tieu et al., 2017). To control for ignorance, Tieu et al. (2017) used the “prediction mode” of the Truth Value Judgment Task, in which the puppet provides a prediction or guess. Then an event occurs and participants are asked if the prediction was right. For example, there was a chicken on the screen and two toy objects, a bus and a plane. The puppet appeared on the screen and stated that “the chicken pushed the bus or the plane”. Then the chicken pushed either one or both of the objects. Participants stamped under
a happy face or a sad face on a scorecard to show whether the puppet’s guess or recollection was right or wrong. They reported that unlike adults, preschool children were more likely to consider a disjunction as “right” when both disjunctions were true, rather than only one. They concluded that the majority of children in their sample (age range: 6;07 - 6;06) interpreted disjunction as conjunction. They hypothesized that this conjunctive interpretation of disjunction is due to children’s non-adult-like pragmatic enrichment.

However, a recent replication of Tieu et al. (2017) by Skordos, Feiman, Bale, and Barner (2020) 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. (2017). 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. (2017)’s findings in the replication condition, but showed that conjunctive interpretations of disjunction disappeared almost completely in the three-alternatives condition. Skordos et al. (2020) 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.

To summarize, our review of previous literature suggests that the design of an experimental task can have a big impact on our conclusions about children’s comprehension of disjunction (Table 1). First, different tasks may be more or less suitable for capturing different implications of disjunction. For example, a “Give-item” task can successfully capture exclusive implications, while a TVJT task with plausible context for speaker ignorance is more successful in capturing inclusive implications. Second, regardless of task type, increased task demands or infelicitous use of disjunction can result in increased conjunctive interpretations of disjunction. With the give-item task, Suppes and Feldman (1969) found a considerable rate of conjunctive interpretations. No evidence for conjunctive interpretations. Skordos et al. (2020)

Table 1. Summary of tasks used in previous studies and their conclusions

<table>
<thead>
<tr>
<th>Task</th>
<th>Conclusions</th>
<th>Representative Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>School test (Imparative) (e.g. Circle all that are bird or black!)</td>
<td>Comprehension of or develops in high school. Before that children often interpret it as conjunction.</td>
<td>Neimark and Slotnick (1970)</td>
</tr>
<tr>
<td>School test (Truth Value Judgment) (e.g. The bird is in the nest or the shoe is on the foot.)</td>
<td>Children (8-14 years) as well as some adults interpret or as a conjunction. This is likely due to task demands and application of nonlinguistic strategies.</td>
<td>Paris (1973)</td>
</tr>
<tr>
<td>Give-item (e.g. Put up the car or the doll.)</td>
<td>Children (4-7 years) interpret disjunction as exclusive (choose-one). The inclusive (logical) concept of disjunction develops later.</td>
<td>Johansson and Sjolin (1975)</td>
</tr>
<tr>
<td>Give-item + Truth Value Judgment (e.g. Give me all the green things or give me all the round things.)</td>
<td>Children (5-6 years) interpret or as exclusive in commands but ignore its contribution in truth value judgments and interpret it as a conjunction. Interpretation of disjunction in commands develops earlier than the knowledge of its truth conditions.</td>
<td>Braine and Rumain (1981)</td>
</tr>
<tr>
<td>Truth Value Judgment (controlling for Speaker Ignorance) (e.g. A troll ate a piece of pizza or an ice cream.)</td>
<td>Children (4-6 years) understand or as inclusive disjunction. Two studies report majority conjunctive interpretations too.</td>
<td>Chierchia et al. (2004), Tieu et al. (2017)</td>
</tr>
<tr>
<td>Truth Value Judgment (controlling for Speaker ignorance and Number of alternatives)</td>
<td>Children (4-6 years) understand the truth conditions of or similar to inclusive disjunction. No evidence for conjunctive interpretations.</td>
<td>Skordos et al. (2020)</td>
</tr>
</tbody>
</table>

2 Following the convention in the language acquisition literature, we indicate children’s age using the “Year;Month” format.
More specifically and with respect to conjunctive interpretations of disjunction, previous studies provide substantial evidence linking them to task design. While it is plausible to consider non-adult-like pragmatic computations as a cause of conjunctive readings of disjunction in children, it is important to first conclusively rule out the influence of task design.

In the studies reported here we use an experimental paradigm that avoids potential issues discovered and discussed in the previous literature. First, our experiments are presented as a guessing game in which a character guesses the animal on a card without seeing it and participants decide if the guess is right. This way the experimental context eliminates the issue of speaker ignorance by making it part of the game. Second, the card game involves three alternative animals (cat, dog, and elephant) which allows for a disjunction guess such as “cat or dog” to be plausibly false when the card has an elephant. Third, our paradigm uses declarative sentences as guesses and avoids imperatives which can lead to more complex implications. Fourth, our studies explicitly test children’s strategies using both binary and ternary forced choice tasks, which allows children to provide more nuanced responses to linguistic stimuli. Katsos and Bishop (2011) was the first study to use binary as well as ternary tasks for assessing scalar implicatures in children’s comprehension. They reported that when the quantifier some was used infelicitously in contexts where all would have been more appropriate, the ternary task reflected this infelicity judgment better (children were more likely to pick the intermediate option) than the binary task. We similarly included a ternary forced-choice task to see if the intermediate response option can be successful in capturing the exclusivity implicature of disjunction. Previous studies avoided one or the other of these potential issues but to our knowledge the experiments presented here are the first to avoid them all, providing a better assessment of children’s underlying competence with linguistic disjunction.

### Present Studies

The goal of this study was to further simplify task design and measure children’s comprehension of disjunction in multiple ways. We used existential sentences (e.g. there is a cat or a dog) in the context of a simple card game. The game controlled for the effect of speaker ignorance/knowledge by making the speaker guess what was on a card without seeing it. The study included trials with the word and to control for the comprehension of conjunction in the same task. The study also had adult participants as controls for children’s performance in the task. Children’s comprehension was measured in three different ways: a binary forced-choice task, a ternary forced choice task, and the analysis of children’s free-form spontaneous verbal reactions before they made their forced-choice judgments. Table 2 provides the summary of methods used in Experiments 1, 2, and 3.

### Experiment 1: Adult Binary and Ternary Judgments

This study examines adults’ comprehension of or, and uses it as a benchmark for children’s comprehension in Experiments 2 and 3. We tested adults in both binary and ternary forced-choice tasks.

### Methods

#### Participants

109 English speaking adults participated via Amazon Mechanical Turk (MTurk). 57 of them were assigned to a binary judgment task and 52 to a ternary judgment task. In the binary task, participants had to judge using the options “wrong” and “right”. In the ternary task they had to choose between “wrong”, “kinda right”, and “right”.

The two conditions were otherwise identical. Participants were randomly assigned to these conditions. The task took about 5 minutes on average to complete. At the end of the study, participants received $0.4 as compensation. All participants provided informed consent before taking the experiment.

---

3 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 et al., 2019). We expect similar behavior from labels that refer to non-maximal degrees of being “right” such as “a bit right” or “a little right”.

---

Collabra: Psychology
Stimuli

We used six images of cards, each with one or two cartoon animals on them. Three cards had one animal and three cards had two (Figure 15 in Appendix). We represent these six cards with animal names in small caps: CAT, DOG, ELE, CAT+DOG, CAT+ELE, DOG+ELE (ELE stands for elephant). In each trial, a card was shown to the participant and a blindfolded cartoon character guessed what animal was on the card. The guess was either a simple existential sentence (There is a cat, There is a dog, There is an elephant), a conjunction (e.g. There is a cat and a dog, There is a cat and an elephant, There is a dog and an elephant), or a disjunction (There is a cat or a dog, There is a cat or an elephant, There is a dog or an elephant). Crossing different cards and guesses results in 54 different possible trials. However, not all these trials are equally informative for our purposes so we created 8 trial types that balanced the number of animals on the cards (one vs. two), types of guesses (simple, conjunction, disjunction), as well as true vs. false guesses. Figure 1 shows our trial types using example cards as rows and example utterances as columns.

Control trials consisted of simple guesses (e.g. elephant, cat) with cards that had only one animal (e.g. CAT) or two animals (e.g. CAT+DOG). In half of these trials the description was true and in half it was false. When two animals were on the card (e.g. CAT+DOG) and one was guessed (e.g. cat), the guess could be infelicitous or even false if interpreted exhaustively (e.g. only cat). In addition to acting as a control, such trials could show how often children derive exhaustive implicatures. Conjunction trials (e.g. cat and dog) were controls for disjunction trials. Conjunction trials were false when only one animal was on the card and true when both were. Finally, disjunction trials constituted the critical trials of our experiments. When only one animal was on the card (e.g. CAT) the disjunction guess (e.g. cat or dog) was true. When two animals were on the card (e.g. CAT+DOG), the disjunction guess (e.g. cat or dog) could be judged as true but infelicitous or even false. Such disjunction trials help us understand whether participants interpreted disjunction as inclusive or exclusive.

We should emphasize that throughout this paper we use example cards (CAT and CAT+DOG) as well as example utterances (elephant, cat, cat and dog, cat or dog) to represent trial types. For example, the "simple true" trial-type represented by the card CAT and the utterance cat, includes similar trials with the card DOG and the guess dog, as well as the card ELE and the guess elephant. Participants saw trials instantiated with all of the different cards and utterances, however, and trials were created randomly for each participant. We also use the short forms cat, cat and dog, and cat or dog without the existential carrier phrase there is to represent the guesses in this paper.

Procedure

The experiment had three phases: introduction, instruction, and test. In the introduction, participants saw the six cards and read that they would play a guessing game. Then a blindfolded cartoon character named Bob appeared on the screen. Participants were told that in each round of the game, they would see a card and Bob was going to guess what animal was on the card. The study emphasized that Bob could not see anything. Participants were asked to judge whether Bob’s guess was right. In the instruction phase, participants saw an example trial where a card with the image of a dog was shown with the following sentence written above Bob’s head: There is a cat on the card. All participants correctly responded with "wrong" and proceeded to the test phase. In the test phase, participants saw one trial per trial type. Within each trial type, the specific card and guess were chosen at random. The order of trial types was also randomized. Figure 15 in the appendix shows an example test trial.

Results

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

Figure 3 shows the results for the ternary judgment task. The addition of an intermediate response option did not affect false simple, true simple, and true conjunction trials. In false simple trials, the animal mentioned (e.g. elephant) was not on the card, and participants judged the guess "wrong". In true simple trials the animal mentioned (e.g. cat) was the only animal on the card and participant considered the guess "right". This was similar to true conjunction trials in which two animals were on the card (e.g. CAT+DOG) and the guess mentioned both (e.g. cat and dog). Participants judged true conjunction trials as "right" in both binary and ternary tasks.

Four trial types showed different patterns of judgments in the binary and ternary tasks. In true-but-incomplete simple trials, one animal was mentioned (e.g. cat) but two ani-
adults' binary judgments in Experiment 1. Columns represent example guesses and rows example cards. Each cell represents a trial-type as explained in Figure 1.

Figure 3. Adults' ternary judgments in Experiment 1.

Table show the proportion of correct guesses for each card and guess combination. The ternary judgments include an intermediate option, which might be used to express partial truth of the guess because one of the guessed animals was on the card. With true disjunction and true-but-infelicitous disjunction guesses, responses were split between "kinda right" and "right". It is likely that participants had different reasons for choosing "kinda right" in each disjunction trial type. In true disjunction trials, participants may have considered a simple guess (e.g. *there is a cat*) as more appropriate. In true-but-infelicitous trials, participants may have expected the connector *and* instead of *or*. As we shall see in the next two experiments, children explicitly mention these alternatives in their open-ended (free-form) responses. Since we are mainly interested in the differences between adults and children, we defer statistical analysis to Experiment 2 where we compare children and adults responses.
Table 3. Truth table for logical conjunction \( \wedge \), inclusive disjunction \( \vee \), and exclusive disjunction \( \oplus \). Letters “p” and “q” represent atomic propositions.

<table>
<thead>
<tr>
<th>p</th>
<th>q</th>
<th>p ( \wedge ) q</th>
<th>p ( \lor ) q</th>
<th>p ( \oplus ) q</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
</tbody>
</table>

Discussion

Consider the truth conditions for conjunction and disjunction in classical logic shown in Table 3. A conjunction is true when both propositions are true and false otherwise. An inclusive disjunction is true when at least one proposition is true, and false otherwise. An exclusive disjunction is true when only one proposition is true and false otherwise. Let’s also assume a simple linking function in which false statements map to “wrong” and true statements to “right.”\(^4\)

In the binary task, judgments for and matched logical conjunction and or inclusive disjunction. If adults in our task interpreted or as exclusive, we expected majority “wrong” responses when both disjuncts were true. This is not what we found.

If truth conditions were all that mattered, the addition of the intermediate option (kinda right) in the ternary task should not have substantially affected the judgments. In fact it did not in false simple trials, true simple trials, and true conjunction trials. These cases showed unequivocal “wrong” and “right” judgments. But in four other trial types, the intermediate option (kinda right) reflected more graded judgments. These trial-types represented utterances that were false, had false implications, or were sub-optimal as guesses.

First, responses in the false conjunction trials were split between “wrong” and “kinda right” responses. In such trials, even though the guess was false, it was not completely incorrect; one of the animals was guessed correctly. Therefore, choosing the intermediate option could reflect the judgment that such guesses are better than those that fail to name any animal on the card. In the other three trial types, responses were split between “kinda right” and “right” responses. In true-but-incomplete simple trials (e.g. cat as guess and CAT+DOG as card), as well as true-but-infelicitous disjunction trials (e.g. cat or dog as guess and CAT+DOG as card), the utterances were literally true but carried defeasible false implicatures; more specifically, in true-but-incomplete simple trials the exhaustive implicature (e.g. “only a cat is on the card”) and in true-but-infelicitous disjunction trials the exclusivity implicature (e.g. “there is a cat or a dog, but not both”). In both these cases, a conjunction (e.g. cat and dog) was the optimal guess and participants’ choice of the middle option “kinda right” likely reflected this judgment as reaction towards the defeasible false implicature. Finally, when only one animal was on the card (e.g. cat), a disjunction guess (e.g. cat or dog) was literally true but sub-optimal as a guess. A simple guess (e.g. cat) would have been better. Therefore, disjunction guesses (with either one or both disjuncts being true) had intermediate acceptability.

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

We should add here that we are aware of one previous study similar to the one presented here on adults’ comprehension of disjunction in simple existential sentences using pictures and forced-choice judgments. Chevallier et al. (2008) presented participants with words (e.g. TABLE), pseudo-words (e.g. IAMIS), or non-words (e.g. RSOU) as well as statements about their spelling containing disjunctions such as “there is an A or a B” (in the word). Participants responded with “true”/”false” in a binary task. Even though they used different response options to ours and they were more concerned with response times and process-

\(^4\) see Jasbi et al. (2019) for a discussion of linking assumptions in forced-choice truth-value judgment tasks.
ing of implicatures, they reported similar results to the ones presented in our study. Most importantly, participants accepted a disjunction when both disjuncts were true 50-80% of the time depending on how much time they were given to respond. The higher end of this range is consistent with the findings of our binary task and the lower end is consistent with the results of our ternary task which showed more sensitivity to pragmatic inferences. We should emphasize that the most important difference between Chevallier et al. (2008)’s study and ours is that we controlled for the role of speaker ignorance in our paradigm by presenting the task as a guessing game.

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

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

Methods

Participants

We recruited 42 English speaking children from the Bing Nursery School at Stanford University. Children were between 3;1 and 5;2 years old (mean = 4;3). Parents of children had provided written informed consent for this experiment.

Materials

We used the same set of cards and linguistic stimuli as the ones in Experiment 1. There were 8 trial types and 2 trials per trial type for a total of 16 trials. We made two changes to make the experiment more suitable for children. First, instead of the fictional character Bob, a puppet named Jazzy played the guessing game with them. Jazzy wore a sleeping mask over his eyes during the game (Figure 15). Second, a pilot study showed that a scale with three alternatives is better understood and used by children if it is presented in the form of rewards to the puppet rather than verbal responses such as “wrong”, “a little bit right”, and “right”, or even hand gestures such as thumbs up, middle, and down. Therefore, we placed a set of red circles, small blue stars, and big blue stars in front of the children. These tokens were used to reward the puppet after each guess. During the introduction, the experimenter explained that if the puppet was right, the child should give him a big star; if the puppet was a little bit right, a little star, and if he was not right, a red circle.

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, ELK, CAT+DOG (Table 4). The experimenter put the sleeping mask on the puppet’s eyes and explained that the puppet is going to guess what animal is on the cards. He then picked the first card and asked the puppet: "What do you think is on this card?" The puppet replied with "There is a dog." The experimenter showed the CAT-card to the child and explained that when the puppet is "not right" he gets a circle. He then asked the child to give the puppet a circle. Rewards were collected by the experimenter and placed under the table to not distract the child. The second trial followed the same pattern except that the puppet guessed "right" and the experimenter invited the child to give the puppet a big star. In the final trial of the instruction, the puppet guessed that "there is a cat" on the card when the card was CAT+DOG. The experimenter said that the puppet was "a little right" and asked the child to give him a little star.

In the test phase, the experimenter removed the three instruction cards and placed a deck of 16 randomized cards on the table. He explained that it was the child’s turn to play with the puppet. For each card, the puppet provided a guess and the child provided the puppet with a reward. The guesses were paired with each card in a way that allowed two trials per 8 trial types.\(^5\)

Offline Annotations

While playing the game, children often provided spontaneous verbal feedback or reactions to the puppet’s guesses. For example if the puppet guessed "there is a cat" and they saw the DOG card, children said (even shouted): "No! Dog!". These reactions happened naturally before children pro-

---

\(^5\) The pilot study had shown that some children struggle with understanding the word "wrong", so "not right" was used instead.

\(^6\) 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.
vided their forced-choice responses by giving the puppet a circle, a little star, or a big star. These verbal responses were categorized into four types based on what words children produced: 1. None, 2. Judgments, 3. Descriptions, and 4. Corrections. The first category (none) referred to cases where children did not say anything (only rewarded the puppet). The second category (judgments) referred to positive/negative words that did not include the name of the animals on the card, for example: "you are right!", "yes", "nope", or "you wonned!". In the third category (descriptions), children mentioned the name of the animal(s) on the card: "cat!", "dog and elephant!", "there is a cat and a dog!" etc. Finally, with correction, children added extra focus and functional elements such as focus words just, only, or stressed the connective word AND. Examples include: "just a cat!", "both!", "the two are!", "only cat!", "cat and dog" (with emphasis placed on and). In trials where the child provided both judgments as well as descriptions or corrections (e.g. "yes! Cat!"), we placed the feedback into the more informative categories, namely description or correction. We should emphasize that the annotation of children's spontaneous verbal feedback was independent of whether the guess was normatively considered "right" or "wrong" and relied only on the words they produced.

Results

Ternary Judgments

Figure 4 shows the results for children's ternary judgments. Starting with the leftmost column in Figure 4, children judged false simple trials as "wrong". In these trials the mentioned animal (e.g. elephant) was not on the card. Moving to the second column, children judged true simple trials as "right". In these trials the mentioned animal (e.g. cat) was the only animal on the card. Here we ignore the results for true-but-incomplete trials in which the animal mentioned (e.g. cat) was only one of the animals on the card (e.g. cat+dog). The reason is that such trials were used in the instruction phase to introduce the "little bit right" option, and the results are probably biased by the instructions. However, it is important to note that the instruction was successful and the majority of children considered such guesses as "kinda right".

Moving to the third column, children judged false conjunction trials as "wrong" or "a little right". In these trials, only one animal was on the card (e.g. cat), but two were mentioned (e.g. cat and dog). In true conjunction trials, both mentioned animals were on the card and children judged the guess as "right". Finally, in true disjunction trials only one animal was on the card and children considered the guess (e.g. cat or dog) as either "right" or "kinda right". In true-but-inelicitious disjunction trials both animals were on the card and children judged the disjunction "right".

Figure 5 compares the results for children and adults' ternary judgments in the conjunction and disjunction trials. The major difference seems to be the cases of disjunction (e.g. cat or dog) when both disjuncts were true (e.g. cat+dog). Children were more likely than adults to consider such utterances as "right". To quantify possible differences between adults and children more precisely and model both our ternary task as well as the subject-level clustering of data, we decided to fit ordinal mixed-effects logistic models. Since ordinal and multinomial logistic models with complex random effects structures are not easily fit in standard frequentist packages, we adopted the Bayesian framework and used the R package "brms" (Bürkner, 2017).

First, we fit separate ordinal mixed-effects logistic models for adults and children. The models included the fixed effect of trial-type and maximal random-effects structures (Barr et al., 2013), i.e. random intercepts and slopes for participants and items (cards). Second, we fit an ordinal mixed-effects model to the combined dataset of adults and children with the added interaction effect of "age category" (adult vs. child), with "adults" set as the intercept. Third, to understand the role of age in children's responses, we fit an ordinal mixed-effects model to children's data with "child age" as an interaction term. For all models, the response variable had three ordered levels: "wrong", "kinda right", and "right". All were cumulative logit models. The trial types "T,Con" (true conjunction), "T,Dis" (true-but-inelicitious disjunction), and "F,Con" (false conjunction) constituted the (dummy-coded) fixed effects of the models, with "T,Dis" (true disjunction) set as the intercept. The priors over trial types were set to $\mathcal{N}(0, 10)$. For other parameters, default weakly informative priors – Student-t $(3, 0, 10)$ and Cholesky LKJ Correlation $(1)$ – were used as endorsed in "brms" documentation. All four chains converged after 4000 samples (with a burn-in period of 2000 samples).

We did not find any effect of children's age on their ternary responses. Therefore, the remainder of this section focuses on the effect of trial-types and the comparison of children and adults' responses. Figure 6 shows the means and the 95% highest posterior density intervals (HPDIs) for the coefficients of these models. The left panel of Figure 6 shows the results from separate ordinal models for adults and children. It helps us understand how adults and children interpreted conjunction and disjunction separately. Because predictors were dummy-coded, it is possible to examine contrasts of interest by computing the difference between coefficients for pairs of conditions. The x-axis shows three contrasts of interest. First, both adults and children rated false conjunction trials lower than true disjunction trials (F,Con - T,Dis) [children's 95% HDI: -4.99, -0.16]. Second, both adults and children judged true conjunction trials better than true disjunction trials (T,Con - T,Dis). Nevertheless, the 95% credible intervals for both groups

---

7 response - trial type + (1 + trial typesid) + (1 + trial typecard)
8 response - trial type * age category + (1 + trial typesid) + (1 + trial typecard)
9 response - trial type * child age + (1 + trial typesid) + (1 + trial typecard)
contained zero. Finally, adults judged true disjunction trials slightly better than true-but-infelicitous ones while children judged true-but-infelicitous disjunction trials slightly better. However, the 95% credible intervals for both groups contained zero. The means and credible intervals computed separately for adults and children match truth conditions of conjunction and disjunction: false conjunction trials were judged negatively and differently from true conjunction and disjunction trials.

To provide a precise estimate of the differences between adults’ and children’s judgments, we look at the means and the 95% HPDIs of the interaction coefficients in the combined adult-child dataset (Figure 6, Right). For false conjunction and true-but-infelicitous disjunction trials, the 95% credible intervals do not contain zero. This suggests that children’s and adults’ judgments differed in these two trial types. In both trial types, children’s judgments were higher than adults’ judgments. This is compatible with the possible effect of two previously discussed hypotheses. First that children are more lenient than adults, and second that children’s judgments are affected by how many labels match the animals on the card (Paris, 1973). However, these hypotheses do not explain children’s responses across all trial-types because children were not always more lenient than adults or were not always affected by label-animal matches in every trial-type. Finally, higher ratings in true-but-infelicitous trials are also consistent with the hypothesis that children compute exclusivity implicatures at a lower
Figure 6. Left: The means and 95% highest posterior density intervals for the coefficients estimated for each trial type in separate ordinal logistic regressions for adults and children. "F,Con - T,Dis" on the x-axis shows the difference between false conjunction and true disjunction trial types; "T,Con - T,Dis" the difference between true conjunction and true disjunction trial types; "T.in,Dis - T,Dis" the difference between true-but-infelicitous disjunction and true disjunction trials. Right: The means and 95% highest posterior density intervals for the interaction coefficients (age category, adults as intercept) in the adult-child combined dataset. The x-axis labels represent False Conjunction (F,Con), True Conjunction (T,Con), True Disjunction (T,Dis), and True but Infelicitous Disjunction (T.in,Dis) trials types. A one unit increase in a parameter value of a model increases the log odds of observing a category or lower ones on the scale over higher ones (Right vs. Kinda Right + Wrong; Right + Kinda Right vs. Wrong). The diagram in the middle shows critical trial-types with lines representing contrasts of interest.

rate than adults (Barner et al., 2011; Noveck, 2001; Papafragou & Musolino, 2003).

Open-ended Verbal Feedback

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

Starting with the leftmost column of Figure 7, in false simple trials the guessed animal was not on the card (e.g. elephant) and children either provided judgments like "No!" or descriptions like "cat" or "cat and dog". Moving to the second column, in true simple trials the guessed animal (e.g. cat) was the only animal on the card and most children provided positive judgments like "Yes". In true-but-incomplete trials, the animal guessed (e.g. cat) was only one of the two animals on the card (e.g. CAT+DOG) and children provided a description of the the card, for example "cat and dog".

In false conjunction trials, only one of the animals was on the card when two were guessed (e.g. cat and dog). In such trials, children provided a high number of corrections and descriptions. The category "corrections" here means that children used the focus particles just and only as in "just a cat" or "only a cat". In true conjunction trials, both animals were on the card and children predominantly provided positive judgments like "Yes!". With true disjunction trials, only one of the guessed animals was on the card and most children simply provided a description of what was on the card (e.g. "cat"). However, in true-but-infelicitous disjunction trials, both animals were on the card yet children provided corrections like "Both!" or emphasizing and as in "cat AND dog!".

To quantify and compare the distribution of children’s spontaneous feedback in different trial-types, we used a Bayesian mixed-effects multinomial regression model with the fixed effect of trial-type as well as random intercepts and slopes for participants and items (cards). The dependent measure was children’s feedback categories of judgment, description, and correction, with correction set as the reference category. The trial types "T,Dis", "F,Con", and "T,Con" constituted the (dummy-coded) fixed effects of the model with "T.in,Dis" set as the intercept. To test the effect of children’s age on their corrective feedback, we used a

feedback ~ trial type + (1 + trial typesid) + (1 + trial typecard)
similar model but added the interaction term "child age". Priors and convergence information were identical to those reported for our previous models.

We did not find any effect of children’s age on their verbal feedback. Therefore, the remainder of this section focuses on the model without the effect of age. Figure 8 shows the means and 95% credible intervals of the multinomial model coefficients, with the x-axis separating trial types. Starting from the left, the credible intervals for judgments and descriptions over corrections for the "F,Con" trial-type included zero. This suggests that the feedback distribution was similar in false conjunction trials and true-but-infelicitous disjunction trials. Both trial types received a relatively high number of corrections. With "T,Con" trials, the credible interval for descriptions over corrections covers zero while that of judgments over corrections stays above zero. This suggests that with true conjunctions children provided more affirmative judgments like "yes" than corrections. Finally with "T,Dis" trials, even though children provided more descriptions, the credible intervals for judgments and descriptions over corrections included zero. As we will see in the next experiment where we register and replicate children’s verbal feedback, children do provide more descriptions than corrections in true disjunction trials.

Discussion

In Experiment 2, we used a ternary judgment task to test children’s comprehension of logical connectives and or. We compared these results to those found in the ternary judgment task of Experiment 1 with adults. The general comparison showed that adults and children had similar patterns of judgments with respect to the truth conditional semantics of the connectives. Both groups had negative judgments for false conjunction statements and positive judgments for true conjunction and disjunction trials. Furthermore, we did not find any effect of children’s age on their forced-choice judgments. This suggests that 3-to-5-year-old children understood the semantics of linguistic conjunction and disjunction in an adult-like manner. However, the results also showed that children’s judgments differed from adults in two ways. First, children were more likely to consider a conjunction guess "kinda right" (and reward the puppet) when it was false. Second, children were more likely to consider a disjunction guess "right" (and reward the puppet) when the guess was infelicitous and potentially carried an exclusivity implicature. This second difference is consistent with the hypothesis that children compute implicatures at a lower rate than adults. However, our analysis of children’s spontaneous verbal feedback provided evidence that children are sensitive to this infelicity and the exclusivity implicature.

Children often volunteered open-ended verbal feedback to the puppet’s guesses before they made their forced-choice judgments. Our analysis of this feedback showed that children’s feedback was sensitive to the semantics and pragmatics of the experimental utterances. Children provided more instances of "correction" when the utterance was false or carried a defeasible false implicature (exhaustive or scalar). As expected from an adult-like understanding of connectives, children corrected the puppet most often when a conjunction was false (i.e. only one proposition was true), or when a disjunction was infelicitous (i.e. both propositions were true) and could carry a false exclusivity implicature. Children’s spontaneous verbal feedback showed that they might be sensitive to the pragmatic infe-
Figure 8. The means and 95% highest posterior density intervals for the coefficients of interest in experiment 2's mixed-effects multinomial logistic model on children's feedback.

The diagram on the right shows critical trial types with lines between trial-types representing contrasts of interest. The category "Correction" was set as the reference category and "True but Infelicitous Disjunction" (T.in,Dis) was set as the intercept of the model. The x-axis labels represent False Conjunction (F,Con), True Conjunction (T,Con), and True Disjunction (T,Dis) trial types. One unit increase/decrease in parameter values corresponds to increase/decrease in log odds of producing the other feedback category over Correction.

licity of a disjunction when both disjuncts were true. Children often explicitly mentioned and as the connective that should have been used in such contexts instead of or. Finally, we did not find evidence for any age effect on children’s verbal feedback. In the next Experiment, we follow up on these finding and replicate the results of Experiment 2 in a binary judgment task.

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

This study used the same paradigm as Experiment 2 but measured children’s judgments using a binary forced-choice task. Similar to Experiment 2, children’s open-ended feedback was also analyzed. The main hypothesis was that preschool children provide corrective feedback if the disjunction is true but infelicitous. However, they do not consider this infelicity to be grave enough to render the guess “wrong” in a binary judgment task. The main hypothesis along with relevant analyses and predictions were preregistered in an "As Predicted" format. 11

Methods

Participants

We recruited 50 English speaking children from the Bing Nursery School at Stanford University. Children were between 3;6 and 5;9 years old (Mean = 4;7). Parents of children had provided written informed consent for this experiment.

Materials

Experiment 3 was similar to Experiment 2 but differed in how children provided their judgments. Based on the findings in Experiment 2, we focused on verbal feedback, instead of forced-choice responses. We used two different ways of measuring children’s judgments. First, we encouraged children to provide verbal feedback to the puppet. They were asked to say "yes" when the puppet was right and "no" when he was not right. Importantly, they were also asked to help him say it better. In each trial, after children were done with this initial open-ended feedback, we asked the classic truth value judgment question: "Was Jazzy (the puppet) right?". This question elicited a binary forced-choice response for each trial independent of children’s earlier open-ended response. These two measures allowed us to compare open-ended and binary forced-choice judgments in the same paradigm and for the same trials.

Procedure

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

As in Experiment 2, the experimenter put a sleeping mask over the puppet’s eyes and explained that Jazzy (the puppet) was going to guess what animal was on the cards. He then picked the first card and asked the puppet: "What do you think is on this card?" The puppet replied with

11 The As Predicted PDF document is accessible at https://aspredicted.org/x9ez2.pdf.
“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 similar simple trials (Table 5) before the test phase began. The test phase contained 16 randomized trials, half of which contained guesses with the words and or or and and or.

Results

In our pre-registration, we had planned for two primary analyses: first, a mixed-effects binary logistic regression on the forced-choice yes/no responses; and second, a mixed-effects binary logistic regression on the number of Corrections in children’s verbal feedback. In the analyses we present in this section we have carried out the first (binary) analysis as planned but deviated from the pre-registered analysis for the second (ternary) analysis. The reason is that a new software for fitting Bayesian multinomial regression models (BRMS) allows us to provide similar but more appropriate models for ternary tasks. Therefore, instead of separate binary logistic regressions on different response categories, we fit a single multinomial model that tests the same pre-registered comparisons. As far as we know, this deviation does not affect our conclusions but makes our analyses more appropriate to the nature of our tasks.

We had left it implicit in our pre-registration that adults serve as an important comparison group in our study. In retrospect, we should have made this explicit in our pre-registration, but given that comparing children’s performance to adults is best practice in such language acquisition tasks, we include it in all our analyses here. We have not carried out the secondary analyses that we had proposed in our pre-registration. These include looking at sub-categories of verbal feedback as well as the agreement between children’s spontaneous Yes/No judgments and their forced-choice responses. Since the discussion of the primary analyses required considerable time and space already, we think the secondary analyses add unnecessary information that is not directly relevant to the main hypotheses of this paper. Finally we have added analyses to check for the possible effect of age on children’s forced-choice and free-form responses. This was not included in our pre-registration but it is best practice in the field.

In the remainder of this section, we first look at the results of the binary judgment task for each trial type and compare them to those of the adults’ in Experiment 1. Then we analyze children’s open-ended verbal feedback and compare them to the forced-choice responses obtained in the same trial types. We should emphasize that similar to Experiment 2, the open-ended feedback was produced before the forced-choice responses and cannot count as verbal justifications for such responses. For the binary 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.

Binary Judgments

Figure 9 shows children’s binary judgments. Starting with the leftmost column and false simple trial types, the guessed animal (e.g. elephant) was not on the card and children considered the guess “wrong”. Moving to the next column and true simple trials, the guessed animal (e.g. cat) was the only animal on the card and children considered the guess “right”. In true-but-incomplete trials only one of the animals was guessed and children’s judgments were equally split between “wrong” and “right”. This is in contrast to adults who unanimously considered such guesses as “right” in their binary judgments (Figure 2). There are two possible explanations for this difference. First, some children may interpret a simple guess like “there is a cat” exhaustively as “there is only a cat”. Second, some children may consider leaving out an animal as a grave violation even though they do not interpret the guess as there is only one animal on the card. The first explanation is unexpected for a theory of acquisition that assumes children are overall more logical or literal interpreters than adults (Noveck, 2001).

In false conjunction trials, only one of the two guessed animals was on the card and most children considered the guess “wrong”. These binary judgments are similar to those of adults’, but different in extent: adults were more consistent and unanimous in rejecting such guesses. In true conjunction trials, children unanimously judged the guess “right”, similar to adults. In true disjunction trials, the card had only one of the guessed animals and most children considered the guess “right”. This is again similar to adults but differs from them in extent: adults more consistently and unanimously judged such guesses as “right”. Finally, with true-but-infelicitous disjunction trials, children considered the guess “right”.

Figure 10 provides a side-by-side comparison of adults’ and children’s binary judgments for conjunction and disjunction trials. The judgments are similar but differ somewhat in trial types where there is only one animal on the card. To quantify trial-type differences in adults and children, we fit separate Bayesian mixed-effects binomial logistic regressions for each group, with “trial-type” as a predictor. Similarly, to capture differences between adults and children, we fit a model to the combined dataset of adults and children and added age category (adult vs. child) as an interaction term. Finally, to check the effect of age on children’s forced-choice responses, we fit a similar model to children’s data with “child age” as an interaction term. These models mirror what we did in our analysis of Experiment 2 data. As in Experiment 2, the models included the fixed dummy coded effect of trial-type (Levels: “T,Dis” (ref-

12 The randomization code as well as the details of the methods are available on this paper’s online repository.
Figure 9. Children’s binary judgments in Experiment 3.

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

Similar to Experiment 2, we did not find any age effect in children’s forced-choice judgments. Therefore, the rest of this section focuses on the effects of trial types and comparison of children’s responses with those of adults. The left panel of Figure 11 shows the means and 95% HPDIs for three contrasts of interest shown on the x-axis, estimated from separate binomial models for adults and children. First, for both adults and children, the 95% credible intervals for "F,Con - T,Dis" do not contain zero. This suggests that for both groups, judgments of false conjunction trials were lower and different than true disjunction trials. Second, 95% credible intervals for "T,Con - T,Dis" estimated for adults and children contains zero. Therefore, adults and children had similar judgments for true conjunction and true disjunction trials. Third, the 95% credible intervals for "T,In,Dis - T,Dis" contains zero as well, suggesting that children and adults judged true-but-infelicitous disjunction trials similar to the true disjunction trials. Overall, the separate binomial models show that judgment patterns match the truth conditional semantics of conjunction and disjunction. False conjunction trials were judged negatively and differently from true conjunction and disjunction trials.

To estimate the extent to which adults’ and children’s judgments differed from each other, we looked at the means...
and the 95% credible intervals of the interaction coefficients computed in the combined binomial model (Figure 11, Right). Based on the 95% credible intervals, we can infer that judgments of adults and children differed in three ways. First, children judged false conjunction trials slightly more positively than adults (F,Con). Second, they judged true disjunction trials slightly more negatively than adults (T,Dis). Notice that these two differences between children and adults are compatible with a small effect of label-matching (Paris, 1973), given that for both of these trials there is a mismatch between the mentioned labels and the animal on the card. Non-adult-like pragmatic enrichment, however, cannot explain this pattern (Singh et al., 2016; Tieu et al., 2017). Pragmatic enrichment predicts that children would rate a true disjunction more negatively than adults, but does not predict more positive judgments for false conjunction trials. The label-matching account predicts both these outcomes, because it posits that in both cases, the partial match between animal labels and animal pictures affects children’s judgments. Nevertheless, this effect is small and label-matching cannot explain children’s responses in all trial-types given that with the same labels (e.g. cat, dog) and picture (e.g. CAT) the use of the connective (or vs. and) creates significant differences in judgments. Finally in the third difference, children judged true but infelicitous disjunction trials more positively than adults did. This is consistent with the hypothesis that children compute exclusivity implicatures at a lower rate than adults. However, as we will see in the next section, children’s spontaneous verbal feedback paints a more nuanced picture regarding exclusivity implicatures in children.

Open-ended Verbal Feedback

Figure 12 shows the distribution of children’s feedback to the puppet in Experiment 3 (see Table 6 for the definitions and examples of feedback categories). Similar to Experiment 2, children’s feedback showed four main patterns. First in false simple trial types when the puppet guessed an animal not on the card (e.g. elephant), there was a split pattern between negative judgments like "No!" and descriptions like "Cat!". Second, almost all children responded with positive judgments like "Yes!" in true simple and true conjunction trial types. These are the trials where the puppet’s guess correctly and optimally matched what was on the card. Third, in true disjunction trials where the puppet used a disjunction (e.g. cat or dog) with only one of the animals on the card, almost all children named the animal on the card (e.g. "cat"). Fourth and most importantly, children provided corrections in trials where a guess was either false or infelicitous (could carry a defeasible false implicature). These included three trial types. First, true but incomplete simple trials in which two animals were on the card (e.g. CAT+DOG) but the puppet only guessed one (e.g. cat). Such trials can carry a defeasible false exhaustivity implicature. Second, false conjunction trials in which the puppet guessed two animals (e.g. cat and dog) but only one of them was on the card (e.g. CAT). Third, true but infelicitous disjunction trials in which two animals were on the card (e.g. CAT+DOG), and the puppet guessed both but used a disjunction (e.g. cat or dog). Such trials can carry a defeasible false exclusivity implicature.

To quantify and compare the distribution of children’s feedback in trial types with connectives, we used a Bayesian
mixed-effects multinomial regression model with the fixed
effect of trial-type as well as random intercepts and slopes
for participants and items (cards). Similar to our analysis in
Experiment 2, the dependent measure was children’s feed-
back categories of judgment, description, and correction,
with correction set as the reference category. The trial-
types “F,Con”, “T,Con”, “T,Dis” constituted the (dummy-
coded) fixed effects of the model with “T.in,Dis” set as the
intercept. Priors and convergence information were identi-
cal to those reported for our previous models.

Figure 13 shows the means and 95% credible intervals of
the multinomial model coefficients. These results replicate
the findings on children’s feedback reported in Experiment
2. Starting from the left, the credible intervals for judg-
ments over corrections as well as descriptions over correc-
tions for false conjunction trials (F,Con) include zero. This
suggests that the feedback distribution was similar in false
conjunction and true-but-infelicitous disjunction trials. In
true disjunction trials (T,Dis), the credible interval for judg-
ments over corrections includes zero but not that of de-
scriptions. Therefore, children provided more descriptions
than corrections in true disjunction trials (T,Dis). Finally
with true conjunction trials (T,Con), the credible interval for descriptions over corrections includes zero, but not judg-
ments over corrections. This suggests that with true con-
junctions, children provided more affirmative judgments
like "yes" than corrections. Overall, the results confirm the
findings reported in Experiment 2: children were more
likely to provide corrections in trial-types that were either
false or could carry a false defeasible implicature (exhaus-
tive or scalar).

To better appreciate the pattern of spontaneous correc-
tions provided by children, Figure 14 breaks down correc-
tions into two sub-categories: those using exclusive focus
words such as only and just (blue) and those using inclusive
focus elements such as both and emphasizing AND. Our
goal here is to focus on the trial types with corrective feed-
back (blue and red). The type of corrective feedback children
provided in these trial types matched the type of mistakes
made in the guesses. With conjunction guesses (e.g. cat and
a dog) when there was only one animal on the card (e.g.
cat), children provided exclusive corrections such as “just a
cat” or “only a cat!”, suggesting that the other animal in the
guess (e.g. dog) should have been excluded. When two an-
imals were on the card (e.g. CAT+DOG) and the puppet used
a disjunctive guess (e.g. cat or dog), or a simple guess (e.g.
cat), children provided inclusive feedback such as “cat AND
dog” or ”both”, suggesting that another animal should have
been included. This is particularly notable in the case of dis-
junction since both animals were mentioned, but children
still emphasized that the connective and should have been
used, or that both animals mentioned were actually on the
card. Such corrective comments hint at a good understand-
ing of differences between the meaning and usage of the
connectives.

Discussion

Experiment 3 measured children’s comprehension of
logical connectives in two ways: First, with analyzing their
open-ended feedback and second, with a binary forced-
choice task. The binary responses followed the predicted
pattern: a false conjunction was judged “wrong” and a true
conjunction “right”. Disjunction guesses were judged right
whether they were true or true-but-infelicitous. Children’s
open-ended feedback in Experiment 3 replicated the find-
ings of Experiment 2. Children’s feedback was sensitive to
the pragmatics of connective use. They provided more cor-
Figure 13. The means and 95% highest posterior density intervals of the coefficients of interest in experiment 3’s mixed-effects multinomial logistic model on children’s feedback. The diagram on the right shows critical trial types with lines between trial-types representing contrasts of interest. The category "Correction" was set as the reference category and "True but Infelicitous Disjunction Trial Type" (T,in,Dis) were set as the intercept of the model. The x-axis labels represent False Conjunction (F,Con), True Conjunction (T,Con), and True Disjunction (T,Dis) trial types. One unit increase/decrease in parameter values corresponds to increase/decrease in log odds of producing the other feedback category over Correction.

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

corrective feedback when the utterance was false or could carry a false defeasible implicature. Furthermore, the corrective feedback was tailored to the puppet’s mistakes. If the puppet used a conjunction when there was only one animal on the card, children pointed out that the other animal should have been excluded from the guess. They used the exclusive adverbials just and only in their feedback. If the puppet used a disjunction when both animals were on the card, children stressed and or both, implying that both animals should have been included. Taking both measures into account, we conclude the following: children’s binary judgments suggest that they understand the basic truth-conditional semantics of or in simple existential sentences as inclusive disjunction. Specifically, when both disjuncts are true, they do not consider an infelicitous disjunction “wrong”. In addition, children’s spontaneous verbal feedback suggests that they consider a conjunction to be a more appropriate utterance in such cases.

General Discussion

Almost a century has passed since Alfred Tarski observed
that disjunction gives rise to complex implications with important psychological effects. To Tarski, these implications appeared unsystematic and informal. Paul Grice, however, considered them a natural consequence of human rational and social interaction. Following Grice’s insights, research in formal semantics and pragmatics has discovered a great deal of systematicity in how we interpret linguistic disjunction. This theoretical progress has in turn led to experimentally testable predictions about the comprehension of disjunction and how it develops in children. Developmental studies in the past two decades have argued that preschool children understand the basic semantics of disjunction. Yet the pragmatic inferences they derive from the use of disjunction has been argued to differ from those derived by adults in two ways. First, due to non-adult-like reasoning, children are more likely to interpret or as and (conjunctive interpretations; Singh et al., 2016; Tieu et al., 2017); Second, children are more likely to consider a disjunction as inclusive (lack of exclusivity implicatures; Chierchia et al., 2001, 2004; Crain, 2008).

Using different types of measurement, this study investigated adults’ and children’s comprehension of linguistic disjunction in simple existential sentences. The results of our experiments confirmed previous findings that adults and preschool children understand the truth-conditional semantics of or as inclusive disjunction (Chierchia et al., 1998, 2004; Crain, 2012). With respect to children’s pragmatic reasoning, however, our experiments suggested that children’s pragmatic reasoning may be more adult-like than previously argued. First, our experiments did not provide evidence for conjunctive interpretations of disjunction, confirming previous studies that have argued such findings may be due to task characteristics and demands (Braine & Rumain, 1981; Paris, 1973; Skordos et al., 2020). Second and with respect to exclusivity implicatures, we found conflicting results based on the type of measurement used. Specifically, our ternary forced-choice task suggested that adults but not children, show sensitivity to the exclusivity implicature of disjunction. However, the quantitative analysis of children’s spontaneous verbal reactions showed that children were also sensitive to such exclusivity implicatures, or at least that they consider and as a better connective in such cases. More concretely, children explicitly mentioned and as the alternative that should have been used in contexts where both disjuncts were true and or was used instead.

The results reported here have two main implications for developmental semantics and pragmatics. First, children’s conjunctive interpretations of disjunction in some of the previous studies have been attributed to a particular theory of pragmatic implicatures (Fox, 2007) and a developmental account in which children differ from adults with respect to the set of alternatives they generate while computing such implicatures (Singh et al., 2016; Tieu et al., 2017). However, as explained in our literature review, there is substantial evidence that conjunctive interpretations, even when robustly observed, are likely due to task demands and application of non-linguistic strategies (Braine & Rumain, 1981; Neimark & Slotnick, 1970; Paris, 1973; Skordos et al., 2020). Therefore, in order to show instances of pragmatically enriched conjunctive readings in preschool children, it is crucial to first rule out conjunctive interpretations due to task demands and application of non-linguistic strategies. Advocates of pragmatically enriched conjunctive readings could achieve this goal by including trials in which the disjunction word (e.g. or) is replaced by a nonsense word. If it is truly the disjunction word that children enrich pragmatically via non-adult-like alternatives, then trials with the disjunction word should elicit higher conjunctive interpretations than control trials with the nonsense word.

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

The experimental results presented here do not fit the predictions of these previous accounts and show more nuance regarding the development of implicatures. We found that children were more likely than adults to judge a disjunction “right” when both propositions were true. This phenomenon is often referred to as “lack of scalar implicatures” in children. Yet, we also found that children were more likely than adults to judge a simple guess (e.g. cat) as “wrong” when there were two animals (e.g. cat+dog). In other words, children were more likely to interpret a simple guess (e.g. “there is a cat”) exhaustively (e.g. “there is only a cat”). Let’s call this pattern “surplus of exhaustivity implicatures” in children. Neither the processing account nor the tolerance account predict “lack of scalar (exclusivity) implicatures” as well as “surplus of exhaustivity implicatures” in preschool children. Whether children struggle with processing pragmatic inferences, or they are more tolerant of pragmatic violations, we should observe the lack of implicatures across the board.

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

Instead, we suggest that these experimental results are compatible with two alternative hypotheses. First, it is possible that even though children know the meaning of the connectives, and know that and is the stronger alternative to or (e.g. they know the scale), they still do not derive an exclusivity implicature. In other words, children know that when both propositions are true, the speaker should have used and instead of or, but they do not infer that the speaker believed the conjunction to be false. At the same time, when both propositions are true (e.g. cat and dog) and only one is asserted (e.g. cat), children know that a conjunction (e.g. cat and dog) is the stronger alternative, and they derive an exhaustivity implicature; in fact more frequently than adults according to our binary task.

What can explain children’s success with exhaustive implicatures but failure with scalar ones even though they show knowledge of the scale? One possibility is that children are uncertain about the “mutual knowledge of the scale”: that both them and the speaker know that and is the stronger alternative to or. Under this hypothesis, children would not punish the speaker by choosing "wrong" or "kinda right" because they are not sure whether the speaker has the knowledge that and is the stronger alternative to or. However, they know the scale themselves and they are happy to offer it verbally in their reactions or feedback. On the other hand, they might have an easier time assuming that ad-hoc nominal scales <cat, cat and dog> are mutually known given that such scales do not require knowledge of subtle function words. The common use of puppets in developmental studies like ours may have contributed to this phenomenon, given that children may have more uncertainty regarding linguistic abilities of puppets.

The second hypothesis is that children and adults differ in how they use the "right"-"wrong" response scale and what types of linguistic violations they consider "wrong". Even though the truth value judgment task commonly uses the "wrong"-"right" scale to measure semantic and pragmatic knowledge, it is not clear how different theoretical constructs such as entailment, presupposition, implicature, or infelicity link to this scale and affect our conclusions. More importantly, it is not clear whether this linking is the same for adults and children. Children may differ from adults with respect to which linguistic violations would make an utterance "wrong" or punishable. They might consider a violation in connective use (or vs. and) much less severe than an incomplete assertion (cat vs. cat and dog). Given that the notions "wrong" and "right" are subject to development with respect to other types of social behavior, it might not be surprising to see the same with respect to linguistic behavior. This issue is further compounded when we consider implicature not as a binary phenomenon (generated vs. not), but rather as an inference with degrees of strength and certainty (Frank & Goodman, 2012; Goodman & Frank, 2016). Future research should systematically explore different linking hypotheses between theoretical semantic/pragmatic constructs and experimental measurements, establishing which types of measurements are most suitable for particular constructs, especially in adults vs. children.

Since Tarski’s original observations on disjunction, research in semantics and pragmatics has shown that the variety of implications Tarski observed are in fact distinct types of meaning observed in many aspects of language and connected to distinct processes that generate them. Therefore, while the inclusive implication is hypothesized to be part of the semantics of linguistic disjunction, exclusivity and ignorance are analyzed as distinct pragmatic inferences generated separately. This theoretical insight has in turn lead developmental researchers to seek distinct developmental mechanisms for each type of meaning. The results of the studies reported here suggest that as more and more varieties of meaning become subject to experimental studies, we may also need to pay closer attention to which types of measurement may be more suitable to capture the specific aspect of meaning under investigation.

Data Statement

All materials, data, and code for this article are available in the Open Science Framework Repository [https://osf.io/jhw8s](https://osf.io/jhw8s)

Declaration of Competing Interest

The authors declare that there are no conflict of interest.

CRediT authorship contribution statement

Masoud Jasi: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Writing - original draft, Writing - review & editing. Michael C. Frank: Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

We would like to thank Eve V. Clark, Christopher Potts, Ellen Markman, and Susan Carey for their generous help and advice on this project. We would also like to thank the staff and children at Stanford’s Bing Nursery School. This work was supported by NSF #1456077 and the Stanford University fellowship to Masoud Jasbi.

Submitted: April 15, 2021 PDT, Accepted: July 25, 2021 PDT

13 We would like to thank Roman Feiman for encouraging us to develop this alternative hypothesis.
REFERENCES


### Table 4. Instruction Trials for Experiment 2.

<table>
<thead>
<tr>
<th>Card</th>
<th>Guess</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>There is a dog!</td>
<td>Circle</td>
</tr>
<tr>
<td>ELEPHANT</td>
<td>There is an elephant!</td>
<td>Big Star</td>
</tr>
<tr>
<td>CAT-DOG</td>
<td>There is a dog!</td>
<td>Little Star</td>
</tr>
</tbody>
</table>

### Table 5. Instruction Trials for Experiment 3.

<table>
<thead>
<tr>
<th>Card</th>
<th>Guess</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>there is a dog!</td>
<td>No!</td>
</tr>
<tr>
<td>ELEPHANT</td>
<td>there is an elephant!</td>
<td>Yes!</td>
</tr>
<tr>
<td>DOG-ELEPHANT</td>
<td>there is a cat!</td>
<td>No!</td>
</tr>
<tr>
<td>DOG</td>
<td>there is a dog!</td>
<td>Yes!</td>
</tr>
</tbody>
</table>

### Table 6. Definitions and Examples for the Feedback Categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>no verbal feedback</td>
<td></td>
</tr>
<tr>
<td>Judgment</td>
<td>provided verbal judgment mirroring the reward</td>
<td>“No!”, “Yes!”, “You are right!”</td>
</tr>
<tr>
<td>Description</td>
<td>mentioned the animal(s) on the card</td>
<td>“elephant”, “cat and dog”</td>
</tr>
<tr>
<td>Correction</td>
<td>used focus particles like only/just, emphasized and or used both</td>
<td>“only cat”, “just elephant”, “both!”, “cat AND dog!”</td>
</tr>
</tbody>
</table>
Figure 15. Cards used in the connective guessing game, sample scenario for adults in Experiment 1, and the puppet in children’s experiments with the sleeping mask on and off.

Bob: There is a dog or an elephant on the card.
SUPPLEMENTARY MATERIALS

Peer Review History