

UCLA

Experiments in Linguistic Meaning

Title

Non-Implicature Sources of Exclusivity in Linguistic Disjunction

Permalink

<https://escholarship.org/uc/item/65q7t6wf>

Journal

Experiments in Linguistic Meaning, 3(1)

Authors

Felton, Casey

Jasbi, Masoud

Publication Date

2025-01-24

DOI

10.3765/elm.3.5806

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Non-Implicature Sources of Exclusivity in Linguistic Disjunction

Casey Felton & Masoud Jasbi*

Abstract. Disjunction in natural language alternates between an inclusive reading (*A or B or Both*) and an exclusive reading (*A or B but not Both*). Traditional accounts of this ambiguity focus on scalar implicature as the source of disjunction exclusivity, a process whereby Gricean reasoning over Horn scales strengthens the baseline inclusive reading to an implied exclusive reading (Grice, 1978; Horn, 1972; Gazdar, 1980). Despite nearly all theories acknowledging that other factors likely play a role in the generation of exclusivity implications, non-implicature factors have received comparatively little attention. Across four experiments we tested two such non implicature factors, prior compatibility and syntactic category, finding that both play a role in speaker interpretations of disjunctive sentences. Additionally, by drawing our stimuli in the first two experiments from the prior literature, we found evidence that previous research on disjunction, while accurately identifying the key role of scalar implicatures, may be overestimating the effect size thereof due to a failure to control for non-implicature factors.

Keywords. experimental pragmatics; prior probability, scalar implicature, disjunction

1. Introduction. When a listener encounters a natural language disjunction they must often determine whether the speaker intended an inclusive (*A or B or Both*) meaning or an exclusive (*A or B but not Both*) meaning. The most studied mechanism explaining the derivation of these implications is scalar implicature; a Gricean process wherein reasoning about the alternative words a speaker could have chosen, but did not chose, allows a listener to infer the logically stronger exclusive reading in many instances of natural language disjunction. That said, numerous other factors likely contribute to the interpretation of disjunction. One such factor that intuitively must contribute is prior compatibility, by which we mean a speaker's preconceived notions of the likelihood that two disjuncts would be true together. Cases exist wherein two disjuncts that are not logically incompatible are nonetheless extremely unlikely to co-occur, such as the sentence *John is singing or screaming*, which we sampled from the literature. While John may be doing both, it seems far more likely that someone would not be both singing and screaming at the same time. This prior incompatibility potentially limits the amount of extra exclusivity that can be introduced by scalar reasoning and other non-implicature factors. Extreme examples, such as those noted by Geurts (2006), may even lead to instances of exclusivity implications sans any scalar reasoning, e.g. in a sentence like *The ball is on the table or on the floor*; a speaker can reasonably derive an exclusive reading from their knowledge that objects can only be in a single location at a time. Another factor worth considering is the syntactic categories of the disjuncts themselves.

* Authors: Casey Felton, University of California, Davis (cdfelton@ucdavis.edu) & Masoud Jasbi, University of California, Davis (jasbi@ucdavis.edu).

For many disjunctions, the same semantics can be expressed with multiple syntactic frames, as in the following examples:

- (1) a. John likes coffee or tea.
b. John likes coffee or John likes tea. (Jasbi, 2018)

To some speakers, sentences formed by coordinating clauses lead to more exclusive interpretations than sentences formed by coordinating NPs. Due to the flexibility of English coordination, it is typically possible to generate object position disjuncts through the coordination of clauses, NPs, and VPs, but to date no experimental work has been done investigating whether speakers differ in their interpretation of these semantically identical but syntactically distinct sentences.

In this paper we present four experiments that test the role of two non-implicature factors on the generation of exclusivity implications. In Experiments 1 and 2 the role of prior compatibility, as well as the degree to which the literature on exclusivity implications in natural language disjunction controls for it, was assessed. In Experiments 3 and 4 the role of the varying syntactic categories of disjuncts coordinated by “or” was assessed while controlling for disjunct length.

2. Prior Compatibility. Experiments 1 and 2 tested the extent to which prior compatibility of disjuncts may have contributed to the exclusivity of examples and stimuli used in the literature on scalar implicatures. While the literature on scalar implicatures has acknowledged the role prior compatibility of disjuncts may play in generating exclusivity implications, there has been no study that measures it systematically. In Experiment 1, we sampled stimuli from previously published theoretical and experimental work on exclusivity implications and asked participants to rate disjuncts independently and without the presence of the connective “or” with respect to their compatibility. If prior compatibility of disjuncts had no contribution to exclusivity in these examples, we expected participants to rate the disjuncts separately as compatible. In Experiment 2, we asked a different group of participants to use the same scale to rate the disjunctive examples, this time with the word “or” present in the sentences. By the design of the two studies, the stimuli differed mainly by the addition of the word “or”, providing a measure of the exact amount of exclusivity, if any, that is attributable to scalar reasoning over and above the effect of prior compatibility measured in Experiment 1. We selected a between subjects design due to concerns that exposure to disjuncts without “or” in trials measuring compatibility might bias participants when exposed to the same disjuncts in the exclusivity rating task, either by causing them to match their responses or by artificially emphasizing or and thus inflating scalar implicature rate.

2.1. SHARED STIMULI. We sampled 47 examples from prior literature using the following criteria. First, a disjunction had to be about something that it is reasonable for speakers to have world knowledge about. For instance, the following sentence from Quelhas & Johnson-Laird (2017) was excluded: *Teresa is writing a poem or Joaquim is painting a picture*. This was not suitable for our experiment because there is no reason to expect speakers to have any prior knowledge about the characters Teresa and Joaquim. Similar exclusions were made for items with disjuncts that are unrelated to one another, such as *The bird is in the nest or the shoe is on the foot* (Paris, 1973). Second, disjunctions could not be interrogatives, as the design of the study was based

around declarative or imperative sentences. Finally, the paper the disjunction occurred in had to discuss the alternation between inclusive and exclusive readings of *or*, broadly construed. We did not exclude examples of embedded disjunction or those with varying entailment environments which potentially introduced more by-item variability to the results of Experiment 2 (See our web resources at <https://doi.org/10.17605/OSF.IO/S2J9D> for a full list of stimuli and their sources).

The final set of stimuli, which were shared across the first two experiments, consisted of 47 disjunctions taken from 20 published papers. Most items were originally presented in English, but those taken from Katsos, Breheny & Williams (2005) were translated from Greek. Minor changes to the original sentences were sometimes necessary to integrate these items into our experimental designs. Alongside our experimental items, six control items were also created, three with disjuncts that cannot co-occur and three with disjuncts that must co-occur. These control items served as attention checks because their compatibility or incompatibility was unambiguous. This meant a total of 53 items, 47 experimental items and 6 controls (See (2) for examples).

- (2) a. Example Experimental Item: John brought pizza or pasta to the party.
- b. Example Control Item (Incompatible): The defendant is innocent of all charges or guilty of all charges.
- c. Example Control Item (Highly Compatible): Lauren is alive or not dead.

All items that were hypothesized in their source articles to be more likely to elicit an exclusive or inclusive interpretation (widely construed) were annotated as either Hypothesized Exclusive or Hypothesized Inclusive. Twenty-five items were Hypothesized Exclusive, fifteen were Hypothesized Inclusive, six were Controls, and seven did not have a predicted reading in their source, but were nonetheless judged to be quality stimuli and thus were included.

2.2. EXPERIMENT 1 METHODS. In Experiment 1 we conducted an online behavioral study to collect data on subjects' prior beliefs regarding the compatibility of the disjuncts in the stimuli. In this experiment participants did not see any instance of linguistic disjunction; they were presented only with the disjuncts themselves without *or*.

Participants. We recruited 57 participants from the Prolific participant pool. All participants were over the age of 18, and self reported as being both first language English speakers and American nationals. Of the 57 initially recruited, six declined to participate and one timed out automatically, leaving 50 participants who submitted responses and received \$3.75 in compensation. Forty-two responses were included in the analysis after excluding those who answered too many attention checking control trials incorrectly.

Stimuli. For the Experiment 1 stimuli, we removed *or* from the examples of disjunction we had collected from the literature, making sure that number agreement was adjusted as needed to maintain grammaticality. Next, we generated a question for each item that asked how likely it was that both of the two disjuncts were true together, that is, asking how compatible the items were. The example experimental item in (3) below explains our process. For example, the dis-

junction *John brought pizza or pasta to the party* would be broken down into the two disjuncts (3a) and (3b) and paired with the question (3c).

- (3) a. Disjunct A: John brought pizza to the party.
- b. Disjunct B: John brought pasta to the party.
- c. Question: How likely is it that someone brought both pizza and pasta to the party?

Procedure. Experiment 1 consisted of 53 randomly ordered trials, each consisting of 2 paired disjunct sentences, the corresponding likelihood question, and a slider bar to respond to the question. After consenting to participate in an online behavioral experiment, participants read instructions and completed 6 practice trials with feedback before beginning the real experiment. The experiment was self paced and took 12 minutes on average to complete.

In each trial participants were instructed to read the two disjunct sentences and then respond to the question using the slider. The slider ranged from 0% (it is impossible for the two disjuncts to co-occur) to 100% (the two disjuncts must co-occur). To reduce rushed responses, the slider had to be moved in order for a participant to proceed to the next trial. The feedback given on practice trials was designed to make sure participants had a good grasp of the scale and were aware that they could rate trials exactly 50% by moving the slider back to its starting position.

2.3. EXPERIMENT 2 METHODS. In Experiment 2 we aimed to collect exclusivity judgements for the same 47 disjuncts tested in Experiment 1. Unlike Experiment 1, here we presented the full examples of disjunction with the connective *or*.

Participants. We recruited 52 participants from the Prolific participant pool. All participants were over the age of 18, and self reported as being both native speakers of English and American nationals. Of the 52 initially recruited, one declined to participate and one timed out due to inactivity, leaving 50 participants who submitted complete responses and received \$3.75 in compensation. Forty-one responses were included in the analysis after excluding those who answered too many attention checking control trials incorrectly.

Stimuli. The stimuli for Experiment 2 were the examples of disjunction collected from the literature on the semantics and pragmatics of disjunction. Four examples had the word *either* in addition to the disjunction word *or*. Since the role of *either* was not in the scope of this study, it was removed from these four stimuli. Unlike in Experiment 1 where number agreement was adjusted to maintain grammaticality, here number agreement was left as it appeared in the literature. Like Experiment 1, each item was paired with a question, this time designed to assess the overall exclusivity of a linguistic disjunction. For example:

- (4) a. Full Disjunction: John brought pizza or pasta to the party.
- b. Question: Based on the sentence, how possible is it that John brought both pizza and pasta to the party?

In order to draw participants' attention to the linguistic disjunction, all questions began with "Based on the sentence, . . ." and the word "possible" was swapped for "likely." These changes

in the wording of the question aimed to help participants focus on the linguistic assessment of disjunction rather than their prior beliefs about the compatibility of the events described.

Procedure. Experiment two consisted of 53 randomly ordered trials, each consisting of a full disjunction sentence, the corresponding question, and a slider bar to respond to the question. After consenting to participate in an online behavioral experiment, participants read instructions and completed 3 practice trials with feedback before beginning the real experiment. The experiment was self paced and took 10 minutes on average to complete.

In each trial participants were instructed to imagine that someone said the disjunction sentence and then respond to the question using the slider. A slight change in instructions to “imagine someone said the sentence” was made to further draw participants’ attention to the linguistic nature of the task. The task was otherwise identical to that in Experiment 1.

2.4. EXPERIMENT 1 RESULTS. **Figure 1** shows the compatibility ratings provided by the participants in Experiment 1 on the Y axis, and is sorted by item on the X axis. It is clearly visible by eye that the items predicted to generate exclusivity implicatures in the literature are grouped around the lower end of the ratings, with the opposite being the case for those predicted to be biased towards an inclusive interpretation (note that blue items are grouped towards the right half of **Figure 1** while green items are grouped towards the left).

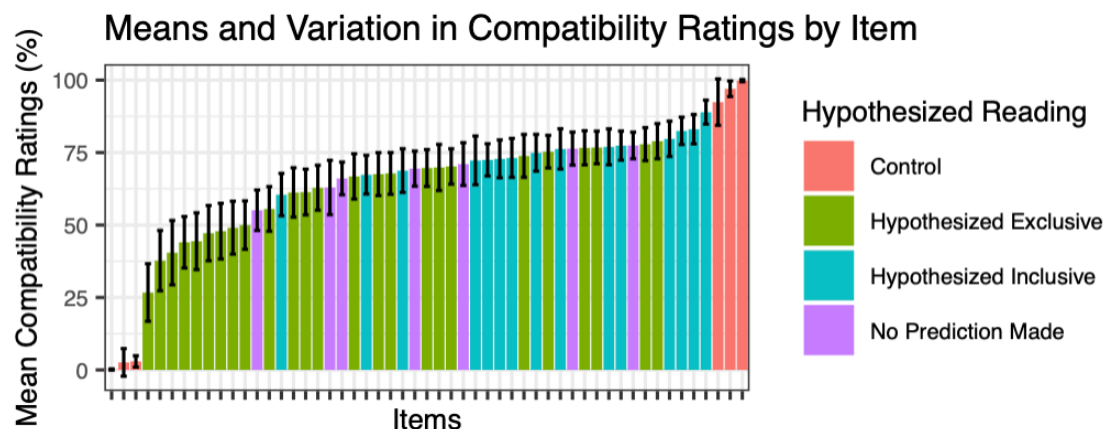


Figure 1: Experiment 1 Ratings (Lowest to Highest)

While both a linear regression or a beta regression are reasonable models to statistically assess whether the items with different hypothesized readings differ in mean compatibility, we selected a Bayesian beta regression to allow for a more accurate modeling of the boundedness of the data. After inflating 0s and deflating 1s to 0.0001 and 0.9999 respectively, both the central tendency and spread of the responses were modeled as a function of an item’s claimed exclusivity or inclusivity in the literature and a random intercept by participant.¹ All R-hats were equal to 1, and inspection of tranq plots suggested no issues with mixing. The model showed a higher mean value for items predicted by the literature to be inclusively biased ($\beta = 0.88$, 95% Credibility Interval = [0.77, 1]), which confirms that the items predicted to generate exclusivity

¹ Because no qualities of the raw data or theoretical considerations suggested that participants treated or ought to treat ratings of 0 or 1 as qualitatively distinct from ratings of 0.01 or 0.99, we decided to eschew a 0 & 1 inflated beta regression in order to reduce model complexity.

implicatures tended to be rated as less compatible while items predicted not to generate implicatures were rated as more compatible, despite no disjunction word being used and therefore no scalar reasoning being possible. This suggests that items in the literature were not being controlled for prior compatibility, and that any claims based off of them may be confounded by it.

2.5. EXPERIMENT 2 RESULTS. Figure 2 shows participants' ratings of exclusivity for the linguistic items in Experiment 2. While there is clear variability among the items, it is easy to see that overall the majority of items are rated lower in Experiment 2 than in Experiment 1 despite the only major difference between tasks being the addition of the disjunction word *or*. In addition, the predicted interpretations from the theoretical literature generally line up with the interpretations provided by participants naive to the linguistic theory. Finally, while some items showed an increase in ratings from Experiment 1 to Experiment 2, this is likely attributable to between group variation in prior compatibility norms.

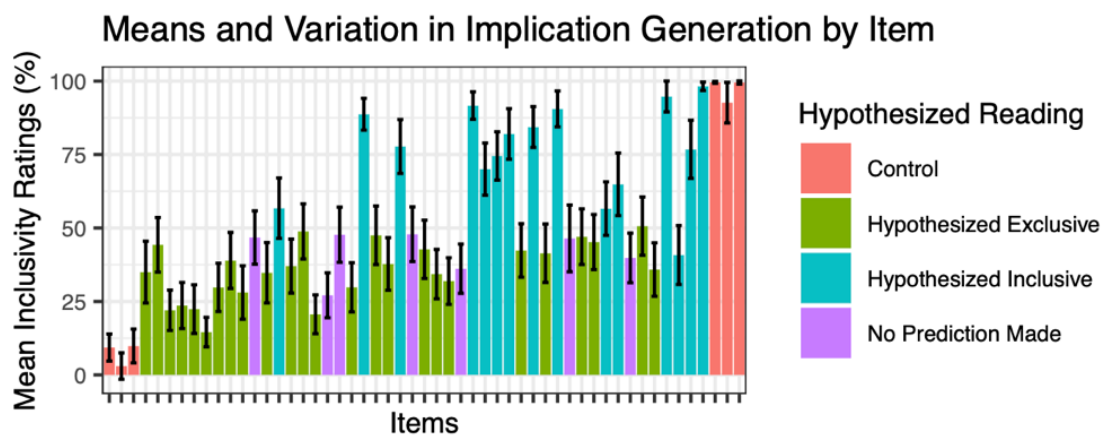


Figure 2: Experiment 2 Ratings (ordering identical to Figure 1)

By regressing the mean compatibility ratings on the mean inclusivity ratings (analogous to implication rate) we were able to directly assess the role of compatibility on exclusivity implication generation. As Figure 3 shows, items that were rated as more compatible were subsequently rated as more inclusive; or less exclusive. Traditional statistics lack straightforward ways to compare the results of the two experiments, but hierarchical bootstrapping as outlined by Saravanan, Berman, & Sober (2020) allows an estimation of the correlation between item means in each experiment. Ten-thousand bootstrapped samples were conducted, returning a mean correlation of 0.527 with a 95% CI of [0.439, 0.608]. This supports the hypothesis that prior beliefs regarding the compatibility of the two disjuncts contribute significantly to the overall interpretation of a disjunction as exclusive. It is important to note here that our stimuli come from the theoretical literature on scalar implicature, where researchers ostensibly select examples that are not *a-priori* exclusive. Nevertheless we see that prior beliefs of compatibility do have a significant effect. It is quite possible that the actual role of prior beliefs on judgments of exclusivity is even more prominent in every-day examples. Going beyond the predictions made by Geurts (2006) that mutually incompatible disjuncts are interpreted exclusively sans implicature, our results suggest not only that disjuncts judged less likely to co-occur often lead to a more exclusive interpretation, but also that disjuncts judged highly likely to co-occur often lead to a more inclusive interpretation. Additionally, the effect does not seem to be limited to completely mu-

tually incompatible disjuncts, but extends to differences in prior compatibility across a range of values.

Even though a large portion of the variance in disjunction exclusivity was explained by variance in prior beliefs about disjunct compatibility, there remained a large amount of variance attributable to other factors, such as pragmatic scalar exclusivity implicatures. For most items, especially those hypothesized to be interpreted as exclusive by the literature, the ratings tended to be lower (that is, more exclusive) in Experiment 2. These differences cannot be attributed to compatibility, and are likely attributable to scalar implicatures given that the two experiments mainly differed with respect to their inclusion of linguistic disjunction, though we cannot rule out other non-implicatures factors as contributors.

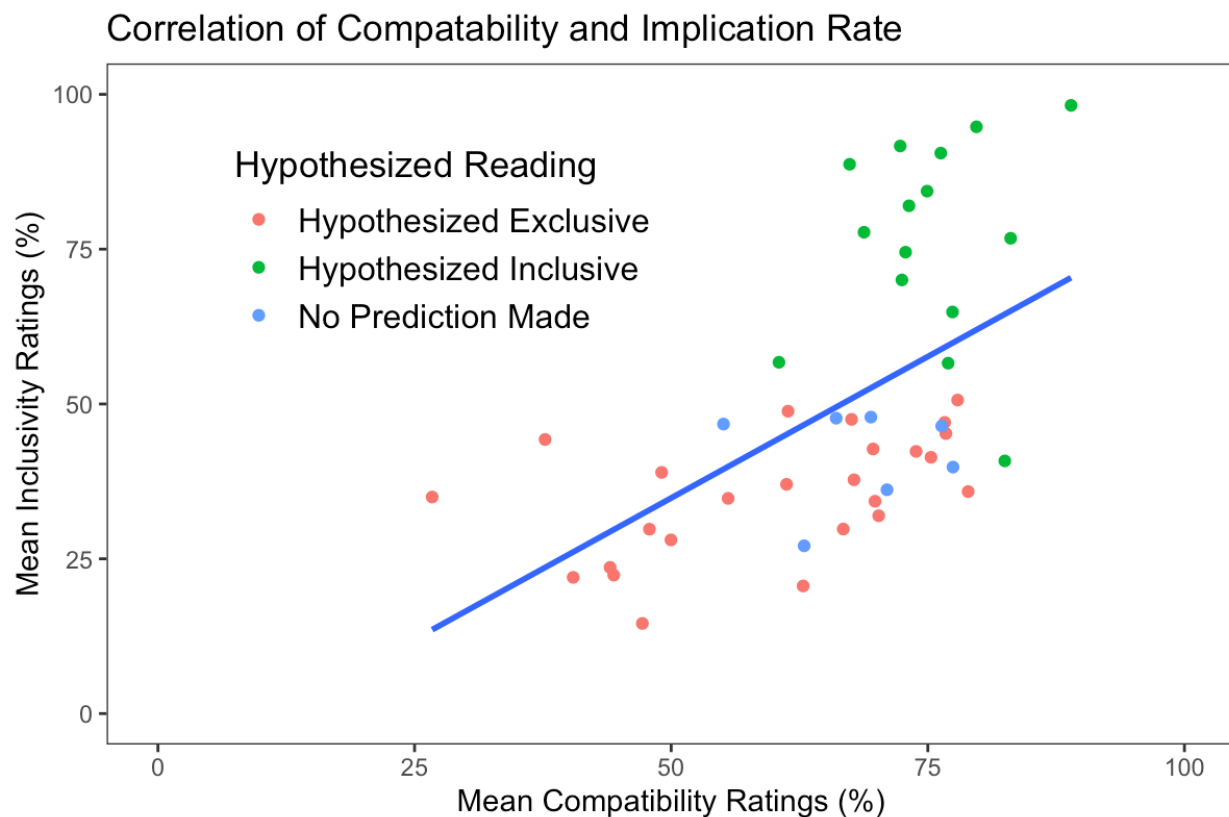


Figure 3: Correlation of Experiment 1 and 2 Ratings

3. Syntactic Category.

3.1. EXPERIMENT 3 METHODS. In Experiment 3 we aimed to assess the effects of syntactic category and disjunct length while controlling for prior compatibility. Additionally, we included an exploratory condition for whether the use of a pronoun rather than repeating the proper name of the subject impacted the interpretation of coordinated clauses, *e.g. John sings or he screams*, versus *John sings or John screams*. Experiment 3 uses a similar procedure to Experiment 2, but with stimuli that control for disjunct length and syntactic category.

Participants. We initially recruited 64 participants from the Prolific participant pool, but later recruited an additional 11 to increase power (since we use a bayesian analysis this does not violate the assumptions of our statistics). All participants were over the age of 18, and self reported as being both native speakers of English and American nationals. Of the 75 total participants recruited, 3 declined to participate and 2 timed out due to inactivity, leaving 70 participants who submitted complete responses and received \$3.75 in compensation. 52 responses were included in the analysis after excluding those who answered too many attention checks and control trials incorrectly.

Stimuli. The stimuli for Experiment 3 were not pulled from previous literature, but instead were created for the explicit purpose of controlling for syntactic category and disjunct length. 32 sentence frames were created that can appear in any of the 4 conditions shown in the table below, but which varied in NP length from 1 to 8 words. Frames were sorted into 4 groups to ensure that each participant only saw each frame in a single permutation. An additional 18 filler items were created that used “and” or “but not” as connectives. These fillers also served as attention checks due to their objective answers. Example experimental stimuli are shown in (5).

- (5) a. Clauses + Proper Name: John likes coffee or John likes tea.
 b. Clauses + Pronoun: John likes coffee or he likes tea.
 c. Verb Phrases: John likes coffee or likes tea.
 d. Noun Phrases: John likes coffee or tea.

Procedure. Experiment 3 consisted of 50 randomly ordered trials, each consisting of a full disjunction sentence (32 items) or filler (18 items), the corresponding question, and a slider bar to respond to the question. After consenting to participate in an online behavioral experiment, participants read instructions and completed 3 practice trials with feedback before beginning the real experiment. The experiment was self paced and took 10 minutes on average to complete.

In each trial participants were instructed to imagine that someone said the stimuli sentence and then respond to the question using the slider. The wording of the prompts were changed from experiment 2 to use “likely” instead of “possibly” to avoid any potential confusion stemming from a modal.² As in the previous two studies, the slider had to be moved in order for a participant to proceed to the next trial.

3.2. EXPERIMENT 3 RESULTS. The results of Experiment 3 were highly noisy, with most visual summaries providing no additional clarity. Analysis of mean ratings in Table 1 showed a slight trend in the hypothesized direction, with VPs and NPs being rated as slightly more inclusive on average than coordinated clauses, regardless of pronoun versus proper name use. No trends were observable across items that differed in disjunct length. The large amount of noise in the data, as well as the small effect size, made extracting clear conclusions difficult, but some of the effects visible in the means summary can be supported by modeling.

² This change was deemed appropriate because the new items are perfectly controlled for prior compatibility because they are semantically identical. Likelihood can thus stand in as a measure of exclusive interpretation, unlike in Experiment 2 where we wanted to specifically emphasize the disjunction while de-emphasizing compatibility.

	Mean Score	S.E.
Clauses (Proper Noun)	31.15	3.862320
Clauses (Pronoun)	29.06	3.756284
Verb Phrases	32.13	3.839423
Noun Phrases	34.02	4.119652

Table 1: Experiment 3 Mean Ratings by Syntactic Category

The modeling challenges in Experiment 3 run parallel to those in Experiments 1 and 2 due to the similarly unusual distribution of the results. In comparison to the previous experiments there is far less one inflation, but instead a much higher rate of zero inflation. Inspection of the data revealed that this zero inflation seems to be the result of some participants having a large bias towards “0%” responses with little variance in their data, nearly always responding that there was no chance both disjuncts were true. Due to this, we elected to use a zero inflated beta regression that models the “0%” responses separately from the non-zero responses, since many of them seemed to be stemming from a different process than the non-zero responses. Additionally, the differences in the shape of the distribution of data for each condition are minimal, suggesting that it is likely unnecessary to model the precision of the beta distribution as a function of our variables. These considerations led us to the following model which aimed to test the minimal aim of the experiment: are coordinated clauses interpreted more exclusively than coordinated VPs and NPs, and is this effect moderated by constituent length.

$$\begin{aligned}\mu &\sim \text{Constituent_Length} * \text{Syntactic_Category} + (1|\text{Item}) + (1|\text{participant}) \\ \phi &\sim 1 + (1|\text{Item}) + (1|\text{participant}) \\ z_i &\sim 1 + (1|\text{Item}) + (1|\text{participant})\end{aligned}$$

In the model, Syntactic Category was collapsed to compare just the two clause conditions (Proper Noun and Pronoun) to the two non-clause conditions (VPs and NPs), and Constituent Length was grand mean centered. Four chains of 4000 samples with 1000 warm up samples were used; all R-hats approached 1, and inspection of tranq plots suggested no issues with mixing. The model output suggests a small positive effect of constituent type ($\beta = 0.09$, 95% Credibility Interval = [0, 0.19]). Inspection of the posterior samples show that 96.9% of the samples estimated an effect of non-clausal syntactic category that was positive. No meaningful effect of any other predictor or interaction was observed, reinforcing what was visible by eye: Constituent Length both plays no observable role in exclusivity implications and does not confound the effect of Syntactic Category. That said, the observed effect of Syntactic Category is both quite small in magnitude and has a credibility interval that comes close to crossing zero, which motivated us to conduct Experiment 4 in order to determine whether the effect of syntax on implication generation is real but small in magnitude or spurious.

3.3. EXPERIMENT 4 METHODS. The results of Experiment 3 left in question whether there is a real effect of syntactic category with a small effect size or no effect of syntactic category at all. Experiment 4 directly tests this question in a modified picture selection task (Gerken & Shady, 1998) that striped away all potential confounds and directly assessed whether disjunctions with

coordinated NPs are interpreted more inclusively than semantically matched disjunctions with coordinated clauses.

Participants. We recruited 51 participants from the Prolific participant pool. All participants were over the age of 18, and self reported as being both native speakers of English and American nationals. Of the 51 initially recruited, 1 declined to participate and no participants timed out, leaving 50 participants who submitted complete responses and received \$3.75 in compensation. 49 responses were included in the analysis after excluding 1 participant who answered too many attention checks and control trials incorrectly.

Stimuli. The stimuli for Experiment 4 were highly simplified compared to those in Experiment 3, and were based on those used in Qiu et al. (2023). A total of twenty four sentences were created made up of 12 experimental disjunction sentences (half clauses, half NPs) and 12 control items. The control sentences had objectively correct answers, allowing them to serve as attention checks as well as distractors.

Each sentence was paired with 4 image cards, like those in Figure 4. Two of the cards contained a single animal mentioned in the sentence, one of the cards contained both of the two animals mentioned in the sentence, and the final card was always blank. Participants were instructed to select one or more cards to reflect their interpretation of the stimuli sentence. This design allows participants to respond in a way that expresses the uncertainty that is central to the meaning of disjunction, unlike traditional picture selection tasks which do not.

There is cat and a dog.

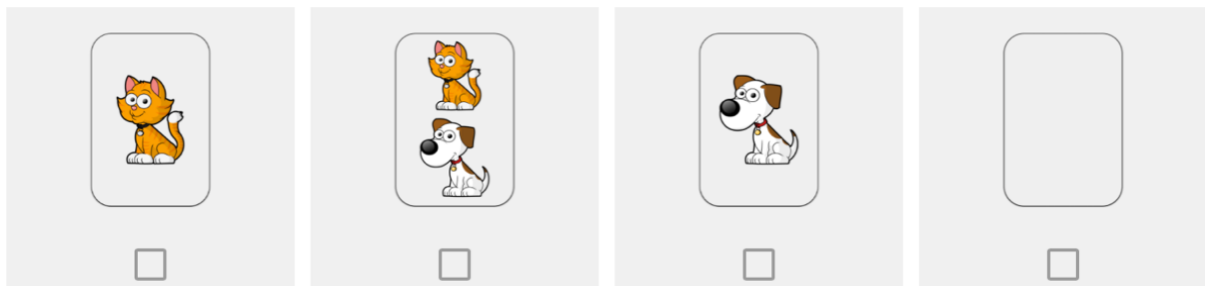


Figure 4: Experiment 4 Trial Example

Procedure. Experiment 4 consisted of 36 randomly ordered trials, each consisting of a disjunction item (24 items, each trial presented twice) or a control item (12 items) and the 4 images paired with that item. After consenting to participate in an online behavioral experiment, participants read instructions and completed 3 practice trials before beginning the real experiment. The practice items did not employ “or” in order to avoid bias, but did include examples that required the selection of multiple cards in order to ensure that participants were aware that they could select more than one card. Non-practice items were presented in a random order, and the linear order of the image cards were scrambled in each trial. In each trial, participants read a

stimuli sentence and then selected one or more cards based on their interpretation of it. The experiment was self paced and took less than 5 minutes to complete on average.

3.4. EXPERIMENT 4 RESULTS. After excluding participants with low accuracy on control items, the data was filtered to include just trials involving *and* and *or* for comparison. The responses were scored categorically into Conjunction (selecting just the card with 2 animals), Exclusive Disjunction (selecting just the two cards with a single animal each), Inclusive Disjunction (selecting all the cards except the blank one) and Other (any other response). A complete summary of the responses is shown in Figure 5, where it is clear by eye that the overwhelming majority of *and* trials were responded to correctly (participants provided conjunctive responses). The *or* trials were split between Inclusive and Exclusive Disjunctive responses, with slightly more Inclusive responses in the coordinated NP condition visible by eye.

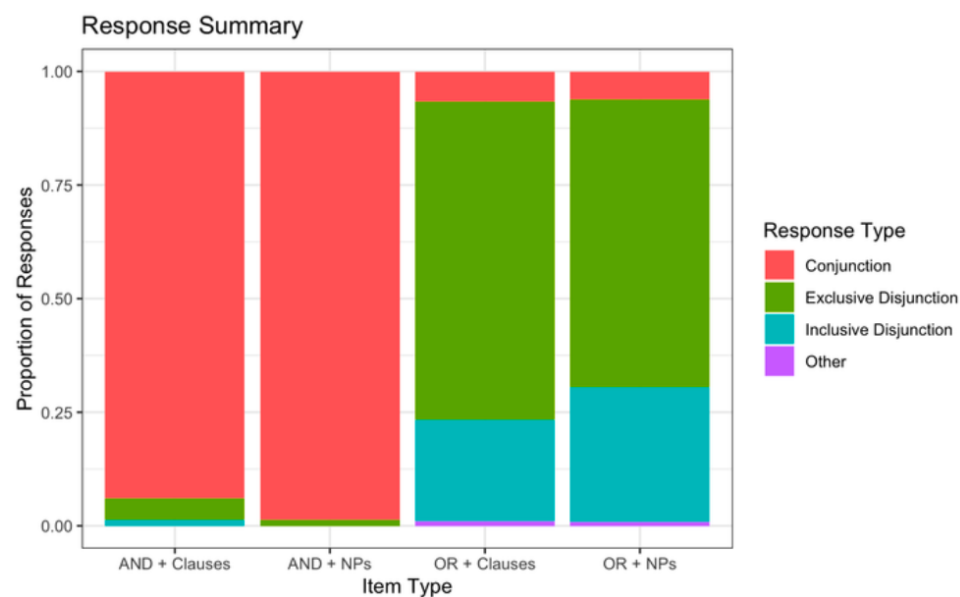


Figure 5: Summary of E4 Responses to AND & OR Items

A bayesian logistic mixed effects regression was fit to test the effect of syntactic category on participant responses to *or* items. Responses were modeled as a function of the syntactic category of their disjuncts, alongside a random intercept of both item and participant and a random slope of syntactic category by participant. Four chains of 4000 samples with 1000 warm up samples were used; all R-hats approached 1, and inspection of tranq plots suggested no issues with mixing. The model output suggests a non-zero effect of syntactic category ($\beta = -2.06$, 95% Credibility Interval = $[-4.14, -0.20]$), meaning that the exclusivity rate was lower for items with coordinated NPs than for items with coordinated clauses, confirming our initial hypotheses. That said, the effect size is small, the model predicting only an 11% change in exclusivity between item types. Taken together with the results of Experiment 3, our data suggests that there is a real, but small effect of syntactic category on exclusivity implication generation.

4. Conclusions. Across 4 experiments we investigated 2 non-implicature sources of exclusivity. In Experiment 1, we used examples from the literature on scalar implicatures to measure the

degree to which prior compatibility contributes to judgments of exclusivity. The results showed that sentences that were judged to be exclusive tended to have disjuncts that were judged to be less compatible and that those judged inclusive tended to have disjuncts that were judged to be more compatible. This finding suggests that informal judgments of exclusivity in the theoretical literature may risk being confounded, unless properly controlled in appropriate minimal-pair contrasts. The results of Experiment 2 confirmed that prior compatibility plays a role in exclusivity implication, explaining a large portion of the overall variance in participant ratings of full disjunctions. That said, the results also clearly demonstrated that the mere presence of the disjuncts with the word "or" in a disjunction introduces exclusivity implications over and beyond judgments of prior compatibility; an effect most likely attributable to scalar implicatures generated by the usage of disjunction word "or". Next, Experiment 3 tested the role of syntactic category and disjunct length, finding no evidence for an effect of disjunct length, but marginal evidence that coordinated NPs may be interpreted more inclusively than coordinated VPs or clauses. Finally, in Experiment 4 we re-tested the effect of syntactic category with a new task designed to reduce noise and detect a lower magnitude effect. The results suggested a consistent, but small, effect of syntactic category, with coordinated clauses being interpreted more exclusively than coordinated NPs. Further research should aim to clarify the theoretical model implied by our data, as well as investigate potential interactions between the sources of exclusivity implications in natural language.

References

- Breheny, Richard, Napoleon Katsos & John Williams. 2005. Interaction of structural and contextual constraints during the on-line generation of scalar inferences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 27. <https://escholarship.org/uc/item/73c75114>
- Gazdar, Gerald. 1980. Pragmatics and logical form. *Journal of Pragmatics* 4(1):1-13. [https://doi.org/10.1016/0378-2166\(80\)90014-4](https://doi.org/10.1016/0378-2166(80)90014-4)
- Geurts, Bart. 2006. Exclusive disjunction without implicature. Nijmegen: University of Nijmegen Unpublished Ms. Note.
- Geurts, Bart. 2010. *Quantity implicatures*. Cambridge: Cambridge University Press.
- Gerken, Louann, & Shady, Michelle E. 1998. The picture selection task. In Dana McDaniel, Cecil McKee, & Helen. S. Cairns (eds.), *Methods for Assessing Children's Syntax*. Cambridge: MIT Press.
- Grice, Herbert P. 1978. Further notes on logic and conversation. In *Pragmatics* (pp. 113-127). Leiden: Brill.
- Horn, Laurence R. 1972. On the semantic properties of logical operators in English. Los Angeles, CA: University of California Dissertation.
- Jasbi, Masoud. 2018. Learning Disjunction. Stanford, CA: Stanford University Dissertation.
- Paris, Scott G. 1973. Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, 16. 278-291. [https://doi.org/10.1016/0022-0965\(73\)90167-7](https://doi.org/10.1016/0022-0965(73)90167-7)
- Qiu, Zhuang, Casey D. Felton, Zachary N. Houghton & Masoud Jasbi. 2023. Task Variation in Scalar Implicature Computation. Poster presented at The Annual Meeting

- of the Cognitive Science Society (Cogsci).
- Quelhas, Ana Cristina & Phillip N. Johnson-Laird. 2017. The modulation of disjunctive assertions, *The Quarterly Journal of Experimental Psychology*, 70(4). 703-717. <https://doi.org/10.1080/17470218.2016.1154079>
- Saravanan, Varun, Gordon J. Berman & Samuel J. Sober. 2020. Application of the hierarchical bootstrap to multi-level data in neuroscience. *Neurons, Behavior, Data Analysis, and Theory*, 3. <https://nbd.scholasticahq.com/article/13927-application-of-the-hierarchical-bootstrap-to-multi-level-data-in-neuroscience>.