Insights into pragmatic strengthening from a new training paradigm

Nadine Bade* WooJin Chung† Léo Picat‡ Rachel Dudley§
Salvador Mascarenhas¶

February 2024

Abstract

We explore the role of alternatives in pragmatic strengthening using a novel training-with-feedback paradigm. In two experiments, we investigate whether training with inferences over contextual alternatives affects pragmatic strengthening with lexical alternatives, and the other way around. We find that training that encouraged (or discouraged) pragmatic strengthening of simple disjunctions carried over to complex disjunctions of an unfamiliar kind to our experimental participants. This shows that our novel methodology is effective in training general mechanisms for deriving alternatives. In a followup, we showed that this methodology can be made to work *across* different kinds of alternatives, if certain salience conditions are met. We argue in favor of a pluralist view of the mechanisms that generate particularized vs. generalized conversational implicatures.

Keywords: implicature; lexical alternatives; contextual alternatives; training with feedback; disjunction

1 Introduction

In conversation, listeners often arrive at meanings for sentences that go beyond their literal interpretation. This *utterance* or *pragmatic* meaning can be the result of the listener considering what sentences the speaker could have used but did not, i.e. *alternative utterances*. For example, given the context in (1), the listener might consider the alternative in (1b) when hearing (1a). Given that this alternative was not used, the

*Department of Linguistics, University of Potsdam
†Department of Linguistics, Seoul National University
‡Université Paris Cité, UFR de médecine, F-75006 Paris, France
§Cognitive Development Center, Central European University
¶Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University

stronger, pragmatic interpretation the listener might arrive at is (1c), often referred to as the *exhaustive* reading.

(1) CONTEXT: Yolanda had spinach and eggs for lunch.
   a. UTTERANCE: Yolanda liked the spinach.
   b. ALTERNATIVE: Yolanda liked the spinach and the eggs.
   c. STRENGTHENED MEANING: Yolanda liked the spinach but not the eggs.

Traditionally, the strong, exhaustive reading in (1c) is not taken to be part of the *literal* meaning of (1a). That is, the proposition that 'Yolanda did not like the eggs' is not *entailed* by (1a). This is easy to see when considering a minimally different context where it is common knowledge that Yolanda only had spinach for lunch. In that context, an inference that she did not like the eggs does not arise for (1a). However, given the context in (1) and assuming that the speaker is as informative as possible (following the *co-operative principle* and *maxim of quantity*, Grice 1989), the listener can infer (1c) from (1a): since (1b) would have been more informative but was not used by the speaker, they must have meant to convey that (1b) is not true. The result of this reasoning process is the pragmatically stronger, exhaustive meaning in (1c), which consists of the literal meaning in (1a), plus the inference that (1b) is false. This inferred non-literal meaning component is a type of *conversational implicature*.

There is an ongoing debate in the theoretical and experimental literature regarding the exact nature of the strengthening mechanism behind quantity-based *conversational implicatures* (referred to as quantity implicatures in what follows) (see in particular Breheny, Katsos, and Williams 2006; Chierchia 2013; Chierchia, Fox, and Spector 2012; Franke 2011; Geurts 2010; Grodner et al. 2010; Huang and Snedeker 2009; Magri 2009; Spector 2016).[1] A central issue in this debate is whether different kinds of alternatives influence pragmatic strengthening, and how (Bott and Chemla 2016; Breheny, Ferguson, and Katsos 2013; Fox and Katzir 2011; Horn 1972; Levinson 2000; Rees and Bott 2018; Waldon and Degen 2020).

Two sources of alternatives for quantity implicatures have been identified in the literature: the *lexicon* and the *context*. Example (1) above illustrates a case where alternatives are provided by the context of utterance. But lexically determined alternatives also appear to play an important role in the derivation of quantity implicatures, and were in fact at the origin of theoretical and experimental investigations into the phenomenon. For example, since the disjunction 'or' forms a lexical scale with the logically stronger 'and', (2b) is a salient alternative to (2a), even in the absence of a rich context (Horn 1972). Consequently, the disjunctive sentence in (2a) *competes* with the lexically derived alternative in (2b). The stronger reading in (2c) emerges, often referred to as the *exclusive* reading of disjunction.

[1] The term *conversational implicature* more generally refers to inferences resulting from all maxims of conversation introduced by Grice (1989): quantity, quality, relevance, and manner. We are exclusively concerned with inferences based on quantity in this article. We also refer to them simply as *implicatures* on occasion, when no confusion could arise.

(2) a. UTTERANCE: She liked the spinach or the eggs. DISJUNCTION
b. ALTERNATIVE: She liked the spinach and the eggs. CONJUNCTION
c. STRENGTHENED MEANING: She liked the spinach or the eggs but not both.

Quantity implicatures involving lexical alternatives have been shown to be far less context dependent. They have therefore been subsumed under the term *generalized conversational implicatures* (Levinson 2000). That is, (2c) is a salient reading of (2a) even without specifying a context of utterance as we did for (1). Due to this dependency of the the latter form of quantity implicature on a particular context they have been referred to as *particularized conversational implicatures* in the literature.

The two types of implicatures just reviewed share a high-level description in terms of reasoning about a speaker's communicative intentions as a function of statements they conspicuously did not make. Yet, the underlying *mechanisms* for accessing and excluding *alternatives*, that is the relevant sentences that the speaker did not utter, may be distinct. Specifically, whereas one requires lexical access (what alternatives is the sentence lexically associated with?), the other requires a context search (what alternatives are salient in the context?). This article addresses the question whether these two cases can or should be treated uniformly. To do so, we report on two studies testing whether the strengthening mechanism involved in *particularized* and *generalized conversational implicatures* influence each other. To look more into the nature of the alternatives involved, we included more complex cases involving conjunction within a complex disjunction as in (3a). The sentence has two readings beyond its literal meaning: a relatively weak one in (3b), and a much stronger reading in (3c).

(3) CONTEXT: Yolanda had spinach, eggs, and potatoes for lunch.
a. She liked the spinach and the eggs, or she liked the potatoes.
COMPLEX DISJUNCTION
b. She did not like all three at once (the spinach and the eggs and the potatoes).
c. Either she liked the spinach and the eggs but not the potatoes, or she liked the potatoes but not the spinach and not the eggs.

The simple exclusive inference in (3b) is well-known, and can be derived by any theory that derives exclusive interpretations for simple disjunctions, including a naive theory that simply states that natural-language 'or' is ambiguous between an inclusive interpretation and an exclusive one, the latter amounting to the logician's Xor (*a* or *b* but not both). The implicature in (3c) seems perhaps more exotic, but it has been observed in the literature on the basis of introspective judgments (Spector 2007), and studied in some detail in the context of reasoning problems (Koralus and Mascarenhas 2013; Mascarenhas 2014; Picat 2019; Sablé-Meyer and Mascarenhas 2021).

Importantly, unlike the cases in (2c) or in (3b), the strengthening in (3c) *cannot* be derived by simply taking it that English 'or' sometimes behaves like the logician's Xor. Take a schematic representation of (3a), for ease of exposition: [*a* and *b*] or *c*. A simple Xor analysis of 'or' in this schema yields [*a* and *b* and not *c*] or [*c* and not [*a* and *b*]]. But notice that (3c) is much stronger than this, it corresponds to the schema [*a* and *b*

and not *c*] or [*c* and not *a* and not *b*].

Thus, the example in (3c) works as a rather refined test case of what strategies humans use when deriving implicatures of this sort. Our experimental investigation into the two processes illustrated in (1) and (2) involved a paradigm of training-with-feedback, where we trained participants on sentences with 'or' and manipulated the contexts of utterance. We gave them feedback meant to push them toward literal or strengthened interpretations of those sentences. Testing participants with sentences with the structure in (3c) allowed us to check precisely what participants were trained for: a particular strategy for dealing with the word 'or', which could not plausibly be generalized to derive the strong meaning in (3c), or more abstract mechanisms for deriving and excluding alternatives proposed in the theoretical literature, which would be able to generate the strong interpretation in (3c). Specifically, the implicature in (3c) requires an *exhaustification* analysis, essentially the view that there exist unpronounced occurrences of an operator much like English 'only' at play in these sentences. Continuing with the schematic meaning for simplicity of discussion, this amounts to 'ONLY [*a* and *b*] or ONLY *c*.'[2] The unpronounced exhaustive operator ONLY is sensitive to contextual alternatives by design, so that this schema can be paraphrased more intuitively as 'either *a* and *b* and nothing else that is relevant, or *c* and nothing else that is relevant.' The relevant propositions here are *a*, *b*, and *c*, so this reduces to 'either *a* and *b* but not *c*, or *c* but not *a* and not *b*.' Complex disjunctions as in (3c) are thus an extremely telling test case for determining what underlying mechanisms are shared by implicatures involving different alternative types.

Armed with these two methodological ingredients—a new paradigm based on training and a new test case in the form of complex disjunctions—we conducted two experiments meant to shed light on the differences and similarities between the two kinds of alternatives and two kinds of mechanisms discussed in the literature. Our Experiment 1 established that, after training which encouraged (or discouraged) the computation of simple implicatures, participants were significantly more likely (respectively less likely) to compute implicatures at a later testing phase which crucially involved the novel *complex* disjunctions. Experiment 2 replicated these results, and demonstrated additionally that our training methodology can be made to work *across* different kinds of implicatures, as long as certain conditions of salience obtain for the alternatives of the kind not directly trained on. We conclude that our methodology successfully trained participants in complex, abstract strategies for pragmatic enrichment, as demonstrated by their ready extension to complex disjunctions. More tentatively, we submit that our results are most compatible with the view that different mechanisms are at play in the derivation of contextual and lexical alternatives.

[2]For ease of exposition, we give here a version of exhaustivity where the unpronounced operator can apply liberally anywhere in the structure (Chierchia, Fox, and Spector 2012), but this particular implicature can also be derived under a globalist perspective, where essentially the same operator is only applied at the very top of the structure (Sauerland 2004), as demonstrated by Mascarenhas (2014, pp. 66–79).

2 Background

A central question in the theoretical and experimental literature on quantity implicatures concerns the role of alternatives in pragmatic strengthening. On the theoretical side, different alternative-generating mechanisms have been shown to produce diverging results for strengthening (Franke 2011; Spector 2016). On the experimental side, alternatives have been shown to determine when these processes are activated (Degen and Tanenhaus 2015; Rees and Bott 2018; Waldon and Degen 2020).

Theories of pragmatic strengthening can be categorized across two dimensions of interest for present purposes: whether they assume uniformity of alternatives, and whether they assume uniformity of a mechanism for their exclusion.

(4) **Uniformity of alternatives**: Alternatives are derived in a unified manner for quantity implicatures involving lexical versus contextual alternatives.

(5) **Uniformity of mechanism/operator**: There is a single exclusion mechanism for contextual or lexical alternatives.

The aim of the novel experiments in this paper is to study these two dimensions in combination, rather than in isolation, thereby informing theories of quantity implicature and models of their computation alike.

2.1 Theoretical background

Most theories of the phenomenon hold that the same exclusion mechanism for alternatives is at play in the derivation of all quantity implicatures. They differ as to whether the alternatives are derived in a uniform manner for lexical and contextual cases. The expressions evoking lexical alternatives that we focus on in this article are simple and complex disjunctions. Thus we will zoom in on theories that discuss disjunction in contrast to quantity implicatures using contextual alternatives.

According to the classical view founded by Grice (1975), the mechanism behind pragmatic strengthening is abductive reasoning over more informative things the speaker could have said. What these more informative things—the alternative utterances—are is taken to be highly context dependent and not formally determined. However, differences between contextual and lexical alternatives had been observed early on (Horn 1972; Levinson 2000). Concretely, lexical alternatives ordered on a scale by entailment (e.g. 'all' $>$ 'most' $>$ 'some') seem to be involved in quantity implicatures with logical words, such as quantifiers ('some', 'most') and propositional connectives ('or', 'if…then') (Horn 1972). Implicatures involving these lexical scales have been subsumed under the term *generalized conversational implicatures*, given that their derivation more generally depends on a lexical scale, not the specifics of the context of utterance. They contrast with *particularized conversational implicatures* based on contextually given alternatives (e.g. {'spinach', 'eggs'}) (Levinson 2000). There is no underlying lexical scale, or order imposed on these alternatives based on entailment.[3]

[3]The alternatives we label as *contextual* here are sometimes referred to as *ad hoc* scales/alternatives as they are built *ad hoc* from what the context provides. We will continue to use the term *contextual* to make explicit what we assume the source of these alternatives to be. We also refrain from using the term *scale* for

That is, under a traditional view, the two cases can be considered as distinct in terms of the alternatives involved.

Under a more recent alternative view, (the *grammatical view*) a covert syntactic operator is the (uniform) source of "pragmatic" strengthening.[4] Simplifying somewhat, this operator takes a set of formally defined alternatives *Alt* and a proposition *p*, and excludes those propositions in *Alt* that are not entailed by *p*. Its operation is very similar to that of the exclusive particle 'only' (*modulo* presuppositions). It can occur locally, embedded in the syntax of the utterance, and it can be obligatory (Chierchia 2013; Chierchia, Fox, and Spector 2012; Magri 2009). The idea behind postulating such an operator is that speakers, when describing a state of affairs, are as *exhaustive* as possible, a variation of the original Gricean observation that people are usually trying to be maximally informative.

There exist two notable versions of this operator, one working with *minimal models* (Groenendijk and Stokhof 1984; Schulz and Van Rooij 2006; Spector 2007; Van Rooij and Schulz 2004) and one working with *innocent exclusion* (Fox 2007). Spector (2016) shows that they differ in what alternatives they need to consider. A theory working with EXH_{mm} distinguishes between lexical and contextual alternatives insofar as lexical scales are built from what is given in the sentence (the alternatives for [*a* or *b*] are the disjuncts *a* and *b*). To build models for the contextual case, the context must be considered (the alternative for *a* is *b*, if *b* is contextually salient). Crucially, the 'and' alternative is not needed to derive the implicature in either case.[5] The operator EXH_{ie} is associated with another view on alternatives, where a limited set of operations—substitution, replacement, deletion—derive alternatives from a given structure, while ensuring that no alternative is ever strictly more complex than the structure of the original sentence (Fox and Katzir 2011; Katzir 2007). The theory of course generates different sets of alternatives for *generalized* and *particularized conversational implicatures* (our lexical vs. contextual cases), but the mechanism is argued to be the same, inasmuch as the same operations (substitution, replacement, deletion) are available in both cases. More specifically, 'or' is substituted by the lexical alternative 'and', while in contextual cases content words are replaced by contextually given alternatives.

The two grammatical views also differ in what alternatives they need to assume for complex disjunctions. Whereas in a minimal model view the alternatives remain minimal as well—that is the disjunction '*a* and *b*, or *c*' has the alternatives *a*, *b*, *c*—the structural view of alternatives has to assume that the alternatives involve structurally less complex alternatives including 'or' or 'and' (*a*, *a* or *b*, *a* and *b*, ...) but also equally complex alternatives where one or both of the logical operators is replaced by its lexical alternatives ('*a* or *b* or *c*', '*a* or *b* and *c*', ...). For details on how the

these cases as it seems to suggest ordering, which not all theories assume.

[4]Scare quotes around "pragmatic" are in order, for in this view the strengthening observed in quantity implicatures is not directly a matter of pragmatic enrichment via considerations about speaker communicative intentions, but rather an in-principle ambiguity as to what is the *literal* (semantic) meaning of the utterance. Pragmatics will of course still be involved in this process, but it will be in its much more general function of ambiguity resolution: is there good independent reason to assume that the speaker intended her utterance to be interpreted with the covert operator in question?

[5]Spector (2016) shows that the EXH_{ie} version of the operator will be vacuous for disjunction under the assumption of minimal alternatives. His main point is that the two operators are equivalent if alternatives are closed under conjunction.

two views introduced derive stronger readings of simple and complex disjunctions as well as exhaustive readings of simple sentences involving contextual alternatives, see Appendix A.

In sum, the two grammatical views just sketched leave room for distinguishing between contextual and lexical alternatives, because *how* they are derived differs. Furthermore, the two views differ in what they assume are the relevant alternatives for simple and complex disjunctions. For the minimal model view, no substitution with the lexical alternatives (other logical operators) is required.

Another view is offered by Geurts (2010), who proposes a Gricean reasoning mechanism overall in quantity implicatures, with disjunction as a relatively special case. Against the traditional Gricean account, Geurts argues that, since '*a* or *b*' communicates independently that 'the speaker does not know *a* and the speaker does not know *b*', it makes little sense to think about the speaker uttering the *conjunction* of claims *a* and *b*, a necessary step in the lexical account reviewed above. The resulting inference, in this view, is not an implicature but an inference based on the probability of both disjuncts being true in a given context. Reasoning about communicative intentions is not required. Thus, not only is the mechanism assumed to be different for the disjunction case illustrated in (2) as opposed to the contextual case in (1), alternatives of an entirely different form are considered to be at play as well.

2.2 Experimental background

The general importance and relevance of alternatives in generating quantity implicatures has previously been established with experimental methods (Bott and Chemla 2016; Breheny, Ferguson, and Katsos 2013; Chemla and Bott 2014; Degen and Tanenhaus 2015; Gotzner, Wartenburger, and Spalek 2016; Rees and Bott 2018; Van Tiel and Schaeken 2017; Waldon and Degen 2020).

Degen and Tanenhaus (2015) report three different experiments using sentence verification tasks on the quantity implicature associated with 'some' ('some but not all'). Their results show that both the size of the domain quantified over and the presence of numerals as alternative expressions offered in the experiment influence the rate of implicatures associated with 'some'. They propose a constraint-based model which predicts that the process of deriving the implicature can sometimes be delayed and sometimes immediate, depending on whether the right contextual conditions are met. They identify the presence and relevance of the right alternative expressions in the experiment as one such condition. They suggest that the difference between contextual and lexical alternatives need not be rooted in an entirely different mechanism for their exclusion to be at play, but rather in different constraints on activation of the alternatives. This model reflects both the more complex theoretical issue of activation of the mechanism and its interaction with alternatives outlined above, as well as the conflicting empirical findings on delays.

Chemla and Bott (2014) present results from reaction-time studies with sentence-verification tasks showing that quantity implicatures associated with 'some' and free choice inferences associated with disjunction ('you may have cake or ice-cream' implies 'you may have cake and you may have ice-cream') display different signatures. They argue that this does not speak against the same underlying mechanism of exclu-

sion for both alternative types, but that access to alternatives works differently for the two cases. More specifically, to derive the implicature of sentences with 'some', the lexical alternative 'all' has to be accessed. However, in the case of disjunction the alternatives are found within the utterance itself: the two disjuncts. In a priming experiment, Bott and Chemla (2016) show that there is priming across lexical and contextual domains, suggesting that they share at least an important part of the strengthening mechanism. They consider different explanations for this: one based on the search for proper alternatives being shared (or not), the other being that the exclusion mechanism itself is primed (or not).

Van Tiel and Schaeken (2017) find differences between lexical ('some') and contextual cases in a picture-verification task with abstract shapes. They argue that this supports a *lexical access* view of scalar implicatures, where it is the accessing of the lexicalized scalar alternative that causes delay in decision times.

Rees and Bott (2018) used the same priming paradigm as Bott and Chemla (2016) to test the role of alternative expressions in implicature computation for existential constructions involving contextual alternatives ('there is a star' when there is also a heart) versus lexical cases involving 'some' and numerals ('some of the hearts are red' when all of them are). They found that the presence of the lexical alternative expression 'all' in the priming phase of the experiment—what they labeled *alternative priming*—was sufficient to prime an exclusion mechanism. That is, even when priming did not force participants to assume the stronger meaning of 'some', but just made them aware of the existence of 'all', they derived implicatures to a higher degree in the probing phase. Based on these findings, they argue for a *salience model* of pragmatic strengthening, in contrast to a *combined model*. The latter is a two-step model which proposes that activating alternatives and activating a mechanism for their exclusion are discrete steps. Both are triggered independently and a certain activation threshold has to be met for each. The salience model is a simpler, one-step model which takes the activation of the alternative to be the threshold for activating exclusion. As soon as the threshold for activating the alternative is met, the exclusion of alternatives will be triggered.

In view of their results, Rees and Bott (2018) revisit the theoretical options discussed by Bott and Chemla (2016) and argue that the priming observed across domains was probably due to a search for alternatives activated in both processes. They find that contextual alternatives differ slightly from lexical alternatives, and suggest that this may be due to higher activation thresholds. This contrasts in part with the processing results reported by Van Tiel and Schaeken (2017), which suggest that exhaustive readings lead to lower reaction times than implicatures associated with lexical scales ('some'). More recently, Waldon and Degen (2020) partially replicated the findings of Bott and Chemla (2016) and Rees and Bott (2018). Employing the same priming paradigm, they again find evidence for priming of quantity implicatures across different expressions (numerals, existential construction 'there is', and 'some'). Like Rees and Bott (2018), they find that exposure to their respective 'canonical' alternatives ('and' for contextual cases and 'all' for 'some') modulates inferences. However, unlike Rees and Bott (2018), they do find differences between priming with strong readings and priming only with the presence of alternative expressions in the experiment. They do also show that if 'only' is offered as an alternative expression in the experiment

to the participants for both sentences with 'some' and existential constructions ('only some…'/'there is only a…'), the rate of implicatures decreases in the probing phase. This suggests that in a experimental setting participants consider alternative expressions that are not part of lexical scales, and are more complex, contra Katzir (2007).

2.3 Summary

Given the theoretical positions and extant experimental evidence laid out above, there are three possible hypotheses regarding the involvement of different types of alternatives and the mechanism for their exclusion in pragmatic strengthening. Under what we will call H_1, implicatures based on contextual versus lexical alternatives are completely parallel in the derivation of alternatives as well as the mechanism for excluding them. According to our H_2, these two cases are completely distinct. H_3 proposes partial overlap between the cases.

Hypothesis 1 ***Uniformity*** There is a single mechanism for generating lexical and contextual alternatives and a single mechanism behind their exclusion.

Hypothesis 2 ***Non-Uniformity*** The alternative generating mechanism and their exclusion mechanism are different for contextual and lexical alternatives.

Hypothesis 3 ***Partial Uniformity*** The alternative generating mechanism is different for contextual and lexical alternatives but the exclusion mechanism is the same (or vice versa).

3 Experiments

The aim of our two experiments is to test the three hypotheses stated above. We look at exclusive and strong readings of simple and complex disjunctions involving lexical alternatives, and compare them to exhaustive readings of simple sentences and conjunctions with contextual alternatives. We designed two training-based experiments investigating the role of visually presented, contextual alternatives versus alternatives provided in the sentence in deriving pragmatically strengthened meanings.

3.1 Experiment 1

Experiment 1 tested the influence of lexical and contextual alternatives on pragmatic strengthening using different sentence types. The goal was to see whether training participants with strong or weak readings of one sentence type would affect interpretation of the other.

3.1.1 Participants

We recruited 199 participants via Prolific. They received what Prolific labeled "good" pay for taking the experiment (GBP 7.50/h). The actual average pay exceeded this amount, as most participants were quicker than the estimated time they were allotted

(GBP 11/h). We excluded 5 participants who answered fewer than 85% of controls in the training phase correctly. Participants were randomly assigned to one of four different training groups (51 participants received weak contextual training, 50 strong contextual training, 48 were in the weak lexical group and 50 in the strong lexical group), see more on the training groups below.

3.1.2 Task and Procedure

The experiment proceeded in two steps: a training phase and a testing phase. During the first phase, participants in one condition received feedback-based training to accept either weak or strong readings of sentences with disjunction involving lexical alternatives ('*a* or *b* is red'). Another group of participants were similarly trained to accept weak or strong readings of sentences involving contextual alternatives ('*a* is red', with objects in the scene other than *a*). In what follows, we refer to training with lexical alternatives as "lexical training," and to training with contextual alternatives as "contextual training." We will go over the different conditions, training regimes, and relevant readings in detail shortly.

In the testing phase of the experiment, participants in the two groups were asked to judge sentence-picture pairings they did not encounter in the training phase. These included inclusive disjunctions if they were previously trained with exhaustive and non-exhaustive readings of simple sentences and conjunctions. They included non-exhaustive readings of simple sentences and conjunctions if they were trained with exclusive and inclusive disjunctions. In addition, they were asked to judge more complex sentences involving disjunction and conjunction ('*a* and *b*, or *c* is red') in scenarios that made their weaker and stronger reading false. Participants did not receive feedback on their responses during the testing phase.

For the training phase, participants were instructed that the experiment was about a guessing game in which someone predicts which, if any, of the shapes that will be displayed are red. Participants' task was to decide whether the prediction was accurate given the picture they saw. They were told that they would be given feedback on their decisions in the first half of the experiment. Phrasing instructions and framing the experiment in terms of a guessing game served two purposes. First, it made clear that sentences were about what is red, thereby determining what is at issue. Second, introducing sentences as guesses made disjunctions felicitous descriptions. As is well known, disjunctions come with the inference that the speaker is ignorant with regard to which of the two disjuncts is true. If the person uttering a sentence had visual access to the picture it would be clear to that person which of the shapes is red. As a result, uttering a disjunctive statement would always be under-informative (and possibly even uncooperative). This would make such an utterance highly infelicitous. However, in a guessing scenario the ignorance with regard to which of the shapes is red becomes perfectly reasonable. Thus, we expected participants to not make additional assumptions regarding how uncooperative or unreliable a speaker is. To remind participants of this general setup, target sentences were always preceded by 'Someone predicts:' The picture showing different shapes appeared after a delay of 1500ms. The delay was added to reduce complexity for participants, as the order in which the shapes were mentioned did not necessarily match the order of visual presentation of the shapes. We wanted

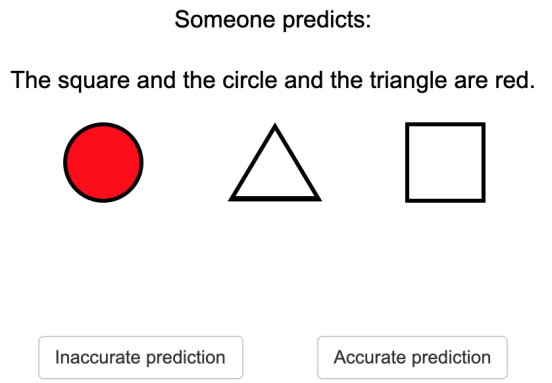


Figure 1: Screenshot of a training trial (condition = false control).

to avoid any attempts of linear mapping that could affect the parsing and interpretation of sentences, especially in the case of more complex disjunctions. A sample of what a trial looked like is given in Figure 1.

3.1.3 Materials

Training phase Participants were confronted with different sentence types in the training phase, depending on which training group they were in. Participants in groups 1 and 2 were trained with sentences involving lexical alternatives, in this case strong and weak readings of simple disjunction (TRAINING TYPE = LEXICAL). That is, they saw critical sentences such as (6) below. The weaker, inclusive, reading of disjunction is paraphrased in (6a), and its stronger, exclusive, reading in (6b).

(6) The triangle or the square is red. SIMPLE DISJUNCTION
   a. *The triangle or the square is red, possibly both.* INCLUSIVE READING
   b. *The triangle or the square is red but not both.* EXCLUSIVE READING

These sentences were paired with pictures which falsified the exclusive reading of disjunction 6 times in the training phase. Furthermore, participants saw 3 pictures with disjunctions that verified their exclusive reading and with 3 pictures that falsified any reading of disjunction, see Table 1.

Within the participants that were trained with lexical cases (disjunction), there was a strong and weak training group (TRAINING STRENGTH = WEAK/STRONG). The feedback differed for these groups. Participants in the WEAK group got positive feedback if they said 'accurate' to pictures that falsified the exclusive reading of a disjunction (first row of Table 1). Participants in the STRONG group got negative feedback if they said 'accurate' in the same situation. Both groups should say 'accurate' in pictures that falsified no reading (second row of Table 1) and 'inaccurate' to pictures that falsified all readings (third row of Table 1). They received suitable feedback in these cases. The

Sentence	Picture	False under	Type
The triangle or the square is red. (6)		exclusive reading	critical simple disjunction
The triangle or the square is red. (3)		no reading	true simple disjunction
The triangle or the square is red. (3)		any reading	false simple disjunction

Table 1: Critical pairings of disjunctive sentences and pictures in the training phase for groups 1 and 2 (LEXICAL priming type). Numbers in brackets indicate number of occurrences in the training phase of the experiment.

full response-feedback matrix can be found in Appendix B.

In addition to sentences containing disjunction, participants saw 3 true and 6 false control sentences, which were either simple sentences or conjunctions with two or three conjuncts, see Table 2. The uneven amount of true and false controls was to avoid creating a general 'accurate'-bias through weak training (assuming that initially participants want to say 'inaccurate' to disjunctions falsifying the exclusive reading). It did create a 'inaccurate'-bias for the strong groups, however, which we took into consideration in the analysis. All training groups received the same feedback for control sentences. The feedback was positive when they said 'inaccurate' to false controls and 'accurate' to true controls, otherwise feedback was negative.

Participants in groups 3 and 4 were trained with strong and weak readings of sentences such as (7a) and (7b) (TRAINING TYPE = CONTEXTUAL), which involved contextual alternatives. The context was provided by the picture displayed, which for critical trials contained more shapes than the ones mentioned in the target sentence. For example, the picture with respect to which (7a) would be evaluated might show a square alongside the mentioned triangle, and the picture for (7b) might show a circle alongside the mentioned triangle and square.

(7) a. The triangle is red. SIMPLE
b. The triangle and the square are red. CONJUNCTION

For these sentences, there are two possible readings: the weaker non-exhaustive readings in (8a) and (9a), which allow for other (contextually given) things to be red, and

Sentence	Picture	False under	Type
The triangle is red. (1)		no reading	true control (simple)
The triangle is red. (2)		any reading	false control (simple)
The triangle and the square are red. (1)		no reading	true control (two conjuncts)
The triangle and the square are red. (2)		any reading	false control (two conjuncts)
The triangle and the square and the circle are red. (1)		no reading	true control (three conjuncts)
The triangle and the square and the circle are red. (2)		any reading	false control (three conjuncts)

Table 2: True and false control sentences used for training groups 1 and 2. Numbers in brackets indicate number of occurrences in the training phase of the experiment (same as for the testing phase).

the stronger exhaustive readings paraphrased in (8b) and (9b).

(8) The triangle is red. SIMPLE

a. *The triangle is red and possibly something else is.* NON-EXHAUSTIVE READING

b. *The triangle is red and nothing else is.* EXHAUSTIVE READING

(9) The triangle and the square are red. CONJUNCTION

a. *The triangle and square are red and possibly something else is.* NON-EXHAUSTIVE READING

b. *The triangle and square are red and nothing else is.* EXHAUSTIVE READING

In these two contextual-training groups, sentences were paired with pictures making the strong (exhaustive) reading false 6 times. To prevent participants from developing strategies based on specific picture types, we varied how many (non-)red shapes there were in the picture (at most 2). Participants saw 3 pictures that verified the exhaustive reading, and 3 pictures that falsified any reading of simple/conjunctive sentences. The critical sentence-picture pairings are given in Table 3.

As with groups 1 and 2 discussed above, groups 3 and 4 differed on whether the training reading was weak or strong (PRIMING STRENGTH = WEAK/STRONG). The strong training group received negative feedback when saying 'accurate' to picture conditions that falsified the exhaustive reading. The weak group received positive feedback when saying 'accurate' to these sentence-picture pairings. Participants in all groups should say 'accurate' in trials that verified any reading and 'inaccurate' to sentence-picture pairings that falsified any reading (and were given feedback accordingly). In addition to the critical sentence types, participants in group 3 and 4 saw the same 9 true and false control sentences as groups 1 and 2. They also received the same type of feedback as groups 1 and 2 for control sentences. The full response-feedback matrix can be found in Appendix B.

Testing phase Test items were the same for all groups of participants. The first kind of test items included the same sentence types as described above, repeated in (10a) to (10c) below.

(10) a. The triangle is red. SIMPLE

b. The triangle and the square are red. CONJUNCTION

c. The triangle or the square is red. DISJUNCTION

These sentences were only paired with pictures that falsified their strong readings. Simple disjunctions appeared 3 times with pictures falsifying the exclusive reading. Simple sentences/conjunctions appeared 3 times with pictures falsifying the exhaustive reading.

Crucially, the testing phase also included a sentence type that was new to all participants. This sentence type included both a conjunction and disjunction. These complex 'and-or' sentences are associated with three types of readings, the weak or inclusive

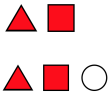
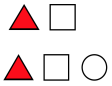
Sentence	Picture(s)	False under	Type
The triangle is red. (4)		exhaustive reading	critical simple
The triangle and the square are red. (2)		exhaustive reading	critical conjunction
The triangle is red. (2)		no reading	true simple
The triangle and the square are red. (1)		no reading	true conjunction
The triangle is red. (1)		any reading	false simple
The square is red. (1)		any reading	false simple
The triangle and the square are red. (1)		any reading	false conjunction

Table 3: Pairings of sentences and pictures in the training phase for groups 3 and 4 (CONTEXTUAL training). Numbers in brackets indicate number of occurrences in the design. The "exhaustive reading" is the one whereby no object is allowed to be red other than the one(s) explicitly characterized as such in the target sentence.

reading (11a), the intermediate reading (11b), and the strong reading (11c).[6]

(11) The triangle and the circle are red, or else the square is. COMPLEX DISJUNCTION

a. *The triangle and the circle are red, or the square is, or possibly all three of them are.* WEAK READING

b. *The triangle and the circle are red, or else the square is but not all three are.* INTERMEDIATE READING

c. *Either the triangle and circle are red but the not the square, or the square is red but not the triangle and not the circle.* STRONG READING

The complex 'and-or' sentences appeared in five different picture conditions during the testing phase (4 times per condition): pictures falsifying any reading (false complex disjunctions), pictures falsifying the intermediate reading and strong reading (complex disjunction (1)), pictures falsifying the strong reading (complex disjunction (2)), and pictures verifying any reading (true complex disjunctions), see Table 4 (page 17). In addition to these test sentence types, the testing phase also contained the same 9 control sentence types illustrated in Table 2 above.

In total, there were 9 controls in the training phase plus 12 critical sentences per training group (21 trials in training phase). There were 22 critical sentences in the testing phase (31 total trials in testing phase). There were also 9 controls of the same type as used in training in the testing phase. Overall, there were 52 trials in the experiment. For all trials the order of symbols was randomized. That is, they did not necessarily match the order of symbols as they were mentioned in the sentence. Which of the items appeared as red was pseudo-randomized. The goal was to include as much variability as possible, so that nothing could be immediately predicted from the form of the sentence or the picture alone.

3.1.4 Design and Conditions

We manipulated two between-subjects factors in Experiment 1: TRAINING TYPE with two levels (contextual training vs. lexical training) and TRAINING STRENGTH with two levels (weak vs. strong). These levels were fully crossed to yield the following 4 conditions: LEXICAL-STRONG, LEXICAL-WEAK, CONTEXTUAL-STRONG, CONTEXTUAL-WEAK. We also manipulated the within-subject factor critical SENTENCE TYPE in the test phase, which had 4 levels (simple/conjunction, simple disjunction, complex disjunction (1), complex disjunction (2)). The dependent variable was rate of 'inaccurate'-responses to a given critical SENTENCE TYPE in the testing phase.

[6] 'Else' was added to the disjunction to make sure that the syntactic structure corresponds to the logical representation [[*a* and *b*] or *c*]. Notice that 'or else' by no means forces exclusive interpretations: in a conditional sentence like 'If John, or else Mary comes to the party, then Bill will be happy,' the consequent follows by *modus ponens* on the conditional and 'John *and* Mary came to the party.' This inference should not be there if the construction 'John, or else Mary' were interpreted exclusively as a matter of necessity. Moreover, since this formulation was kept stable across training conditions and experiments, our results will allow us to tell whether this construction pushed participants to get exclusive readings more strongly than in other cases. Finally, recall from the discussion in the introduction that the strong reading of complex disjunctions in (3c) (p. 3) cannot be arrived at with a simple exclusive interpretation of 'or'.

Sentence	Picture	False under	Type
The triangle is red. (3)		exhaustive	critical simple disjunction
The triangle or the square is red. (3)		exclusive	critical simple
The triangle and circle are red, or else the square is. (4)		strong reading	critical complex disjunction (1)
The triangle and the square are red, or else the circle is. (4)		strong/intermediate reading	critical complex disjunction (2)
The triangle and the square are red, or else the circle is. (2)		no reading	true complex disjunction
The square and the circle are red, or else the triangle is. (2)		no reading	true complex disjunction
The triangle and the square are red, or else the circle is. (4)		any reading	false complex disjunction

Table 4: Test items for all four training groups. The number in brackets indicates how often the condition appeared in the testing phase.

3.1.5 Predictions

The three hypotheses introduced above (page 9) correspond to different predictions for how strong training in each group will affect critical sentence types differently. To see whether different sentence types were affected differently by the type and strength of training, we will first look for a three-way interaction between TRAINING TYPE, TRAINING STRENGTH and SENTENCE TYPE. If the three-way interaction is not present, and we only find a two-way interaction of TRAINING STRENGTH and SENTENCE TYPE, our results are in line with hypothesis (H_1), *Uniformity*. If we find a three-way-interaction, we will investigate what is driving it by looking at pairwise comparisons between strong and weak training for each training and sentence type. If we find that the effect is only present within the same alternative type, we have more specific evidence for H_2 (*Non-Uniformity*). If we find the effect of TRAINING STRENGTH is present but asymmetric across training and sentence types we have evidence for the two cases being partially the same, as held by H_3 (*Partial Uniformity*). To show that the effect goes beyond a general 'inaccurate'-bias created by strong training or asymmetric number of true/false controls in testing, we will compare the effect of strong training on critical sentence types (simple/conjunction, simple and complex disjunction) to 'inaccurate'-responses to 'accurate'-baselines (controls in the test phase). If we find them to be different, we have evidence that the effect goes beyond a general bias created by training.

3.1.6 Results

All analyses were performed using `R` Statistical Software (R Core Team 2022). We analyzed the data using logistic regression and generalized linear mixed effect models for binomial data with the `lme4` package in `R`.[7] The dependent variable was the rate of 'inaccurate'-responses to critical test sentences, as saying 'inaccurate' indicated the presence of a specific reading, as summarized in Table 5. The results for critical sentence types and the 'accurate'-control sentences are summarized in Figure 2.

First, we were interested in the question whether 'inaccurate'-responses to different critical sentence types were affected differently by TRAINING STRENGTH but also TRAINING TYPE (lexical/contextual). To test for the presence of an interaction between all three fixed factors,[8] we used nested model comparisons via log likelihood ratio tests. The models included random intercepts for participants. We did not include items in the random effects structure as a given trial was only defined through the variable SENTENCE TYPE. The order in which the symbols appeared and which of them were red was random for each participant. To see whether the three-way interaction between SENTENCE TYPE, TRAINING TYPE, and TRAINING STRENGTH is justified we compared a model with this three-way interaction to a model without the three-way interaction, see Table 6. We find evidence for the presence of such an interaction through log likelihood ratio tests.

[7] For the full analysis, data, and link to the experiments see the supplementary materials: OSF project.

[8] We used 'dummy'-coding with the following reference levels: for SENTENCE TYPE = 'accurate'-control, for TRAINING STRENGTH = weak, and for TRAINING TYPE = contextual.

sentence (type)	picture	response	reading
simple disjunction			
'*a* or *b* is red'	*a b*	'accurate'	inclusive
'*a* or *b* is red'	*a b*	'inaccurate'	exclusive
simple/conjunction			
'*a* is red'/'*a* and *b* are red'	*a b* / *a b c*	'accurate'	non-exh reading
'*a* is red'/'*a* and *b* are red'	*a b* / *a b c*	'inaccurate'	exh reading
complex disjunction (1)			
'*a* and *b* are red, or else *c* is'	*a b c*	'accurate'	weak
'*a* and *b* are red, or else *c* is'	*a b c*	'inaccurate'	intermediate/strong
complex disjunction (2)			
'*a* and *b* are red, or else *c* is'	*a b c*	'accurate'	weak/intermediate
'*a* and *b* are red, or else *c* is'	*a b c*	'inaccurate'	strong

Table 5: Reading corresponding to different responses to the critical sentences in the testing phase.

model	npar	Chisq	Df	Pr(>Chisq)
resp $\sim$ (sentence type + training type + training strength)^2 + (1 | subjectId)	17			
resp $\sim$ sentence type * training type * training strength + (1 | subjectId)	21	81.545	4	$<$ 2.2e-16 ***

Table 6: Model comparisons between a model including the three-way interaction of all three fixed factors and a model without the three-way interaction but including two-way interactions between each fixed factor.

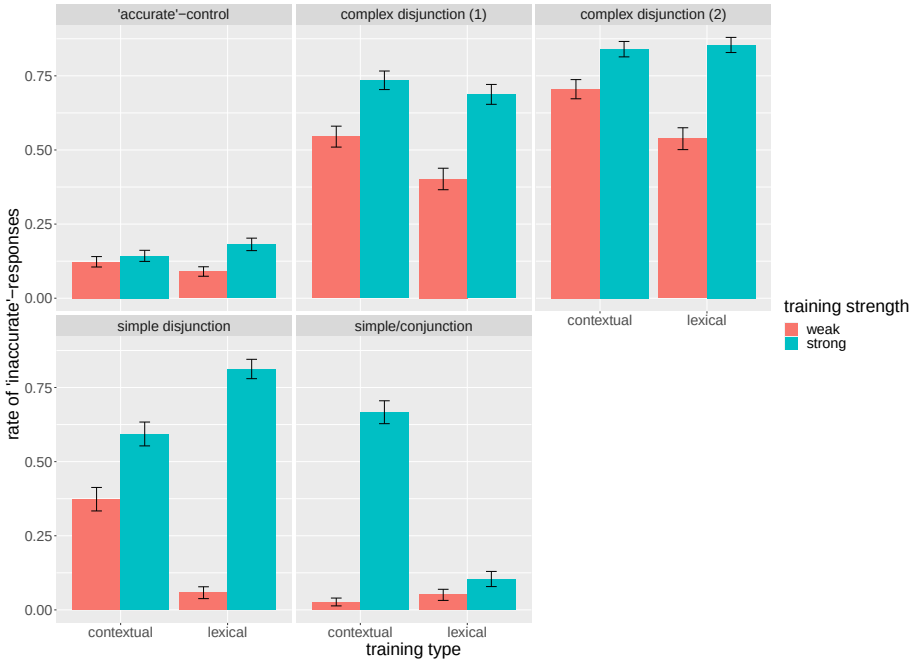


Figure 2: Rate of 'inaccurate'-responses in the testing phase by SENTENCE TYPE (simple/conjunction, simple disjunction, complex disjunction (1) and (2)), TRAINING TYPE (lexical/contextual), and TRAINING STRENGTH (weak/strong). Error bars indicate the standard error. Complex disjunction (1) refers to the intermediate reading of 'and-or' sentences, complex disjunction (2) refers to the strong reading.

training type	sentence type	contrast	estimate	SE	p.value
lexical	simple disj.	weak - strong	-5.0462	0.503	$<$.0001 ***
lexical	compl. disj. (1)	weak - strong	-1.4164	0.325	0.0008 **
lexical	compl. disj. (2)	weak - strong	-1.8468	0.354	$<$.0001 ***
contextual	monocl./conj.	weak - strong	-4.8599	0.592	$<$.0001 ***

Table 7: Results of the pairwise comparisons between weak and strong training for critical sentence types and different training types (contextual/lexical)

We were interested in how strong training of the two different types affected different sentence types differently. To do so, we looked at pairwise comparisons (Bonferroni-corrected for multiple comparisons) between weak and strong training for different TRAINING TYPES (contextual/lexical) and different SENTENCE TYPES with the `emmeans` package (Lenth 2023), see Table 7.

We see that lexical training affects all types of sentences involving disjunction, i.e. it increases the rate of exclusive readings for simple disjunction, as well as the intermediate reading (complex disjunction (1)) and the strong reading of complex disjunction (complex disjunction (2)). We furthermore see that contextual training increases exhaustive readings of simple sentences and conjunctions. No other contrasts between weak and strong training were significant (see supplementary materials for the full results matrix).

To make sure that the effect goes beyond a general 'inaccurate'-bias created by training, we checked whether strong lexical and contextual training affect these sentence types differently from 'accurate'-baselines. We see that, for all sentence types where we found an effect of strong training, the increase in 'inaccurate'-responses exceeds the effect of that training on the rate of 'inaccurate'-responses to 'accurate'-controls. The relatively high error rates we observe with the 'accurate'-controls are mostly due to the complex control sentences containing both 'and' and 'or' (see the results for individual control types in the Supplementary Materials). The fact that the effects on critical sentence types are different from these errors suggests that they are pragmatic in nature and do not just reflect a low level strategy participants developed for these sentence-picture pairings.

To evaluate these findings fully, it is important to consider what the initial rate of exclusive and exhaustive readings is, and how training unfolded for the different sentence types. We looked at the rate of 'inaccurate'-responses to simple disjunctions and simple/conjunctive sentences by trial number during the training phase, see Figure 3. We see that exhaustive readings are initially a bit lower than exclusive readings of disjunctions. However, we also see that the development over time is relatively similar for lexical and contextual training, with lexical training reaching higher levels of 'inaccurate'-responses. Importantly, training works in both directions: weak training reduces the rate of 'inaccurate'-responses reflecting the relevant readings, whereas strong training increases this rate.

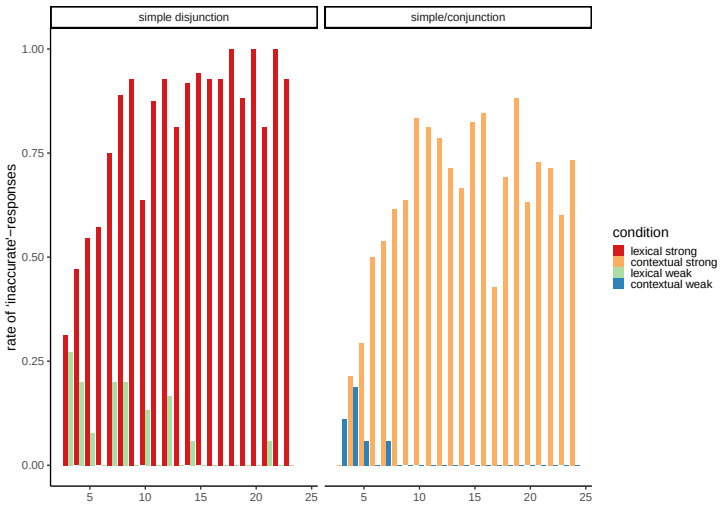


Figure 3: Rate of 'inaccurate'-responses in the training phase by training group and trial number for each critical sentence type (simple disjunction and simple/conjunction).

3.1.7 Discussion

We find that contextual training increases the rate of exhaustive readings of sentences with contextual alternatives. It did not increase the rate of stronger readings for simple or complex disjunctions involving lexical alternatives. Lexical training with exclusive disjunctions affected the rate of stronger readings of all types of disjunctions, simple and complex, but did not affect the rate of exhaustive readings of simple or conjunctive sentences significantly. Crucially, these results is not compatible with an account where participants simply learned a novel, exclusive meaning for 'or'. This is because strong readings of complex disjunctions cannot be generated by simply deploying an exclusive ('but not both') interpretation of disjunction. The finding suggests that the same mechanism for deriving and excluding alternatives is involved in these lexical cases. It is in line with both involving the activation of sentence internal, minimal alternatives (Spector 2016), or the activation of a substitution mechanism in both cases (Katzir 2007). Since neither lexical substitution nor activation of minimal alternatives within the sentence were activated with strong contextual training, our experiment was not successful at increasing the rate of exclusive readings of simple disjunctions, or the rate of strong readings of complex disjunction.

The fact that training in Experiment 1 was only successful within one alternative type lends support to H_2 (*Non-uniformity*): the alternative generating and the alternative excluding mechanisms behind implicatures involving lexical versus contextual alternatives are different. Given that this is at odds with previous results suggesting at least partial uniformity, we considered different explanations for why there is no training effect across alternative domains with the set-up of Experiment 1. We hypothesized

that participants may have developed specific strategies for the verification of different sentence types based on the visual set-up. More specifically, judging the accuracy of using disjunction did not require a close consultation of the context as the alternatives mentioned in the sentence always matched the symbols presented in the picture. That is, in order to verify the exclusive or inclusive reading, participants in the lexical training groups only had to check whether none, one or two of the shapes were red. They never had to think about *which* shapes were red. In other words, to derive the alternatives for disjunction the sentence itself gives sufficient information, as the sentence '*a* or *b* is red' was paired with a picture which always contained a shape *a* and a shape *b*. However, to determine what alternatives need to be excluded for the contextual case ('*a* is red'), the broader visual context needs to be considered. Since a context search was not needed for the verification of simple or complex disjunctions, this explanation also accounts for the fact that contextual training did not affect the rate of implicatures associated with them. The overall high rate of strong readings of complex disjunctions in contextual training groups suggest that a strategy based on sentence internal alternatives/lexical substitution was the default for these sentence types, and was not mediated through the exclusion of contextual alternatives in training. To test the hypothesis that a closer consultation of the visual context was required in training to activate contextual alternatives, we conducted a follow-up experiment.

3.2 Experiment 2

The goal of Experiment 2 was to further test Hypotheses 1–3 regarding the role of alternatives and the mechanism for their exclusion in pragmatic strengthening of simple and complex disjunctions on the one hand, and simple sentences/conjunctions on the other. The results of Experiment 1 show that training people to interpret disjunctions exclusively did not increase rate of exhaustive readings, and training participants for exhaustive readings did not increase the rate of exclusive interpretations of simple disjunctions. They thus lent support to H_2: the two cases differ completely with regard to the generation and exclusion of alternatives. We speculated that this finding—which is at odds with previous findings suggesting at least partial overlap between the two types of implicatures—are due to certain strategies for the verification of exclusive readings having been developed which did not require a close consultation of the visual context. Specifically, a context search for which shapes are red was not required. As a result, the threshold for activating contextual alternatives might not have been met. Even if the mechanism was active, it might have failed due to the absence of the right alternative.

To test this new hypothesis that contextual alternatives need to be relevant for the verification of strong readings to be activated, the design of Experiment 2 highlighted the importance of which alternatives were present in the picture (as opposed to the sentence) during the training phase. Furthermore, the sentence type used for training was now the same for all training groups, which made the training more similar. More specifically, participants were now always only trained with simple disjunction ('*a* or *b*') but in the presence of a picture with 3 shapes (*abc*). Participants thus had to consult the visual context as to which of the shapes were red when they needed to verify exclusive readings of disjunction. We hypothesized that forcing participants to consider the *c* alternative would invoke context searches and affect contextual cases. In addi-

tion, some groups were trained to *exclude* this third alternative (trained for exhaustive readings) whereas other groups were only trained with exclusive disjunction but were visually exposed to a third shape not present in the disjunctive sentence. The goal was to see whether visual exposure to additional context alternatives was sufficient to activate a mechanism for their exclusion, or whether a specific mechanism for excluding contextual alternatives needed to be trained. If we find training to be effective across alternative types (support for H_2) this further distinction allows us to check whether activating the alternative is sufficient to trigger a mechanism for exclusion, or whether an additional threshold for excluding the alternatives has to be met (one- versus two-step model of computation).

3.2.1 Procedure

The procedure used for Experiment 2 was the same as the one used for Experiment 1.

3.2.2 Participants

We tested 197 participants via Prolific, with a final average remuneration of GBP 8.23/h. They are randomly assigned to one of four training groups. 50 participants received inclusive-exhaustive training, 49 participants exclusive-non-exhaustive training, 50 participants were trained with inclusive-exhaustive readings and the remaining 48 participants were assigned to the group that received exclusive-exhaustive training, see more on these training groups below. We excluded 20 participants due to them responding inaccurately to less than 85% of the control sentences in the training phase. We analyzed data from 177 participants.

3.2.3 Materials

Training phase In Experiment 2 participants were trained with simple disjunctive sentences such as (12). The sentences were always paired with a picture containing three shapes. These pictures varied with regard to two properties: first, whether one or both of the shapes mentioned by the disjunction were red (thus verifying or falsifying the exclusive reading of disjunction). Second, whether the third object (not mentioned in the disjunction) was red, thereby verifying or falsifying the contextual exhaustive reading (nothing else is red). Four different readings of these sentences were the object of training using these picture types. These four readings are the result of multiplying the options of (i) the disjunction being weak or strong (inclusive or exclusive) with the two options of (ii) '*a* or *b*' being interpreted as non-exhaustive ('at-least *a* or *b* is red') or exhaustive ('*a* or *a* is red but nothing else is red'). The four readings are paraphrased

in (12a) to (12d).

(12) The triangle or the square is red. SIMPLE DISJUNCTION

a. *The triangle or the square is red, possibly both, and possibly something else is red.* INCLUSIVE NON-EXHAUSTIVE

b. *Either the triangle or the square is red, though not both, and possibly something else is red.* EXCLUSIVE, NON-EXHAUSTIVE

c. *The triangle or the square is red, possibly both, and nothing else is red.* INCLUSIVE, EXHAUSTIVE

d. *Either the triangle or the square is red, though not both, and nothing else is red.* EXCLUSIVE, EXHAUSTIVE

In the training phase, participants were confronted with sentence-picture pairings that made each these readings either true or false. That is, they saw 6 sentence-picture pairings which were true under the exclusive and exhaustive reading, 6 which were true only under the exclusive reading, 6 which were true only under the exhaustive reading and 6 sentence-picture pairings which were true under any reading. In addition, there were 3 false controls which were not true under any reading of the disjunction (exhaustive/exclusive). Table 8 indicates which pictures falsified which of the readings of simple disjunction laid out in (12). Moreover, there were 12 control sentences of the same form as used in Experiment 1 (simple, conjunctions with 2 or 3 conjuncts) in the training phase. There were 6 true and 6 false controls. Given that the groups received an uneven amount of negative feedback, there were still different levels of 'inaccurate'-biases created by training. The 'accurate'-controls in the testing phase thus served as a baseline for 'inaccurate'-responses for the analysis, see more below.

Participants were given feedback on their judgments regarding how appropriate the sentences were as a prediction for the pictures they saw according to the training group they were assigned to. There were four different training groups. Group 1 was trained to accept only exclusive and exhaustive readings. Group 2 was trained to accept only exclusive but both exhaustive and non-exhaustive readings. Group 3 was trained to accept both inclusive and exclusive readings of disjunctions, and both non-exhaustive readings and exhaustive readings. Group 4 was trained to accept both inclusive and exclusive readings of disjunction but only exhaustive readings. The feedback matrix for responses and each of these groups is given in Appendix C.

Testing phase The relevant test sentences were the same as in Experiment 1. They appeared the same number of times as in Experiment 1 (as illustrated in Table 4 above). Additionally, there were 3 'accurate'-controls for simple sentences and conjunctions (making the exhaustive reading true), and 3 'accurate'-controls for simple disjunctions (making the exclusive reading true) in Experiment 2. The same type of control items as used in the training phase appeared in the testing phase 12 times (6 true/6 false). The 'accurate'-controls in the testing phase served as a baseline for a comparison of the effect of training on critical sentence types to make sure that it was not just creating a general 'inaccurate'-bias.

In total, there were 12 controls in the training phase plus 27 critical items per group

Sentence-picture(s)	False under
The triangle or the square is red. (6) 	exclusive reading
The triangle or the square is red. (6) 	exhaustive reading
The triangle or the square is red. (6) 	exclusive, exhaustive reading
The triangle or the square is red. (6) 	no reading
The circle or the square is red. (3) 	any reading

Table 8: Picture conditions under different readings.

(39 trials in training phase). The same 12 controls appeared in the testing phase plus 6 'accurate'-controls for simple disjunction (2), simple sentences (2) and conjunction (2). In addition, there were the same 22 items involving critical sentence types as used in Experiment 1 (40 trials in testing phase). In total, there were 79 trials.

3.2.4 Design and Conditions

The between-subjects factors we manipulated in Experiment 2 were CONTEXTUAL TRAINING with two levels (weak= non-exhaustive/strong= exhaustive) and LEXICAL TRAINING with two levels (weak = inclusive/strong=exclusive). There were thus 4 training groups in total:EXHAUSTIVE-EXCLUSIVE DISJUNCTION, EXHAUSTIVE-INCLUSIVE DISJUNCTION, NON-EXHAUSTIVE-EXCLUSIVE DISJUNCTION, NON-EXHAUSTIVE-INCLUSIVE DISJUNCTION. The within-subject factor SENTENCE TYPE has the same 4 levels as in Experiment 1: simple/conjunction, simple disjunction and two types of complex disjunction, (1) and (2). Participants in the exhaustive groups were trained to exclude alternatives not present in the sentence but present in the picture (contextual alternatives), whereas participants in the non-exhaustive groups were trained to accept additional red shapes in the picture. Participants in the exclusive groups were trained to only consider strong readings of disjunction (exclusive), whereas inclusive groups were trained to accept weak readings of disjunctions (inclusive). As before, the dependent variable was the rate of 'inaccurate'-responses to critical SENTENCE TYPES in the testing phase, as that was indicative of participants' deriving relevant strong readings falsified by the picture.

model	npar	Chisq	Df	Pr(>Chisq)
resp $\sim$ (sentence type + lexical training + contextual training)^2 + (1 | subjectId)	17			
resp $\sim$ sentence type * lexical training * contextual training + (1 | subjectId)	21	48.961	12	$<$ 5.95e-10 ***

Table 9: Model comparison between a model with the three-way interaction between all three fixed factors and model without the three-way-interaction (but with two-way-interactions) including random intercepts for subjects (most complex converging model).

3.2.5 Predictions

Based on the results from Experiment 1, we predicted an effect of both LEXICAL TRAINING and CONTEXTUAL TRAINING on the rate of 'inaccurate'-responses to critical sentence types in the test phase. If they affect the critical SENTENCE TYPES (simple/conjunction, simple disjunction, complex disjunction (1), complex disjunction (2)) differently, we should find a three-way-interaction between LEXICAL TRAINING, CONTEXTUAL TRAINING and critical SENTENCE TYPE. This would be evidence in favor of H_3.

Additionally, we predict an effect of strong training for each training type on all critical sentence types. Specifically, strong contextual training should affect strong readings of all types of disjunctions, even if lexical training itself is weak (inclusive). If our conjecture regarding the results of Experiment 1 is right—that is, if the presence of symbols in the picture that are not mentioned in the sentence evokes a context search—we should also find an effect of lexical training on exhaustive readings. If this effect is present even when contextual training itself is weak (non-exhaustive), we have evidence that the activation of alternatives is sufficient to evoke an exclusion mechanism (a one-step model of pragmatic strengthening). All effects should, as before, go beyond the effect of either training on 'accurate'-controls, which served as baselines for the general asymmetric 'inaccurate'-bias created by the different training groups.

3.2.6 Results

We used the same statistical tools and classes of analyses as reported above for Experiment 1, so in this section we focus immediately on the details of direct relevance. The rate of 'inaccurate'-responses to critical sentence types per training groups is in Figure 4.

We looked for a three-way interaction between SENTENCE TYPE, CONTEXTUAL TRAINING and LEXICAL TRAINING. We compared a model with the three-way-interaction to a model without it. The models included random slopes for participants and no random effects for items. We found the interaction term to be justified by the data, as per Table 9.

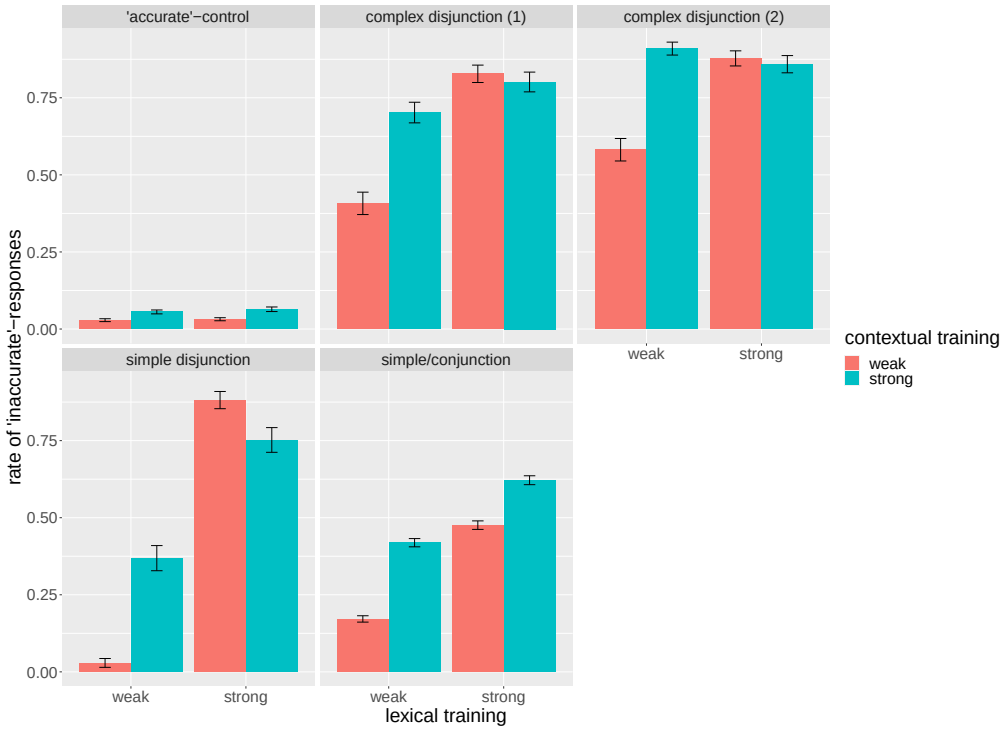


Figure 4: Rate of 'inaccurate'-responses in the testing phase by SENTENCE TYPE, LEXICAL TRAINING (weak/strong) and CONTEXTUAL TRAINING (weak/strong). Error bars indicate the standard error. Complex disjunction (1) refers to the intermediate reading of complex 'and-or' sentences, complex disjunction (2) to its reading.

lexical training	sentence type	contextual training (contrast)	est.	SE	p.value
weak	compl. disj. (1)	weak - strong	-1.4312	0.303	0.0001 ***
weak	compl. disj. (2)	weak - strong	-2.1309	0.368	<.0001 ***
weak	simple/conj.	weak - strong	-3.2218	0.648	<.0001 ***
weak	simple disj.	weak - strong	-3.3830	0.578	<.0001 ***
strong	simple/conj.	weak - strong	-1.4837	0.351	0.0014 **

Table 10: Pairwise comparison of weak and strong contextual training when lexical training was weak (above black line)/strong (below black line).

We looked at the effect of strong contextual training and strong lexical training by pairwise comparisons of weak and strong training feedback for critical SENTENCE TYPE when the other training was weak/strong, respectively. The comparisons are based on least square means and contrasts were run with the *contrast* function of the `emmeans` package in R. Effects are Bonferroni-corrected for multiple comparisons.

Concentrating on contextual training first, we now see an effect of contextual training on all critical sentence types, including all types of disjunction, even if lexical training itself was weak, see Table 10 (above the black line). When lexical training was strong, there was no additional effect of contextual training for any sentence type involving disjunction. Looking at the contrasts between weak and strong lexical training, we also see an effect of lexical training on all sentence types when contextual training was weak, see Table 11 (above the black line). We also see that strong lexical training had an additive effect when contextual training was strong for simple sentences/conjunctions and disjunctions, see 10 (below the black line). Strong lexical training had an additional effect on simple disjunctions and simple sentences/conjunctions when contextual training was strong, see 11 (below the black line). Comparing the effects of strong training on all critical sentence types to effects on 'accurate'-controls, we see that the former had significantly more impact for all of them, suggesting that the effects we see go beyond a general 'inaccurate'-bias (for the full contrast matrix see the Supplementary Materials).

To see how sentence types are affected differently by training type, we looked at the contrasts between sentence types for strong contextual and lexical training, respectively. We first considered cases where contextual training was strong while lexical training was weak. We see that complex disjunctions were more affected by strong contextual training than simple disjunctions or simple sentences/conjunctions, see Table 12 (above the black line). We find no evidence for exhaustive readings of simple sentences being more affected by contextual training than exclusive readings of disjunction. The two complex disjunctions, moreover, crucially differed from one another, with the strongest reading being impacted more by strong contextual training. We then looked at differences between sentence types when lexical training is strong while contextual training is weak. The effect of lexical training on the rate of exhaustive readings was weaker than the effect on the rates of exclusive readings of disjunction. It was also

contextual training	sentence type	lexical training (contrast)	est.	SE	p.value
weak	compl. disj (1)	weak - strong	-2.2875	0.332	<.0001 ***
weak	compl. disj. (2)	weak - strong	-1.8420	0.351	<.0001 ***
weak	simple/conj.	weak - strong	-3.0780	0.652	0.0001 ***
weak	simple disj.	weak - strong	-6.2728	0.621	<.0001 ***
strong	simple/conj.	weak - strong	-1.3399	0.341	0.0052 **
strong	simple disj.	weak - strong	-1.8610	0.353	<.0001 ***

Table 11: Pairwise comparisons between weak and strong lexical training when contextual training was weak (above black line)/strong (below black line).

training	sentence type (contrast)	est.	SE	p.value
cont. strong	compl. disj. (1) - - simple disjunction	1.5521	6.200	<.0001 ***
cont. strong	compl. disj. (2) – simple disjunction	3.1405	9.575	<.0001 **
cont. strong	compl. disj. (1) – simple/conj.	2.0151	7.739	<.0001 ***
cont. strong	compl. disj. (2) – simple/conj.	3.6035	10.720	0.0022 ***
cont. strong	compl. disj. (1) – compl. disj. (2)	-1.5884	-5.015	0.0026 ***
lex. strong	simple/conj. – simple disj.	-3.4966	-9.770	<.0001 ***
lex. strong	compl. disj. (1) – simple/conj.	3.0151	9.903	<.0001 ***
lex. strong	compl. disj. (2) – simple/conj.	3.4584	110.562	0.0022 ***

Table 12: Comparison between different sentence types regarding the effect of strong contextual and strong lexical training on them if the other training was weak.

stronger for both readings of complex disjunction compared to exhaustive readings of simple sentences/conjunctions, see Table 12 (below the black line).

We again looked at the performance of participants during the training phase. The goal was to see whether our training regimes were equally effective for all groups. The performance at the beginning of training (first 5 trials) is illustrated in Figure 5. At the beginning of training, we see a relatively mixed picture, which is probably due to higher complexity of the sentence-picture pairings and feedback.[9] We observe that non-exhaustive sentences (two panels on the right) received overall more 'inaccurate'-responses than exhaustive sentences (two panels on the left). The same holds for inclusive (bottom two panels) versus exclusive conditions (top two panels). That is, we have no evidence for one having a higher baseline rejection rate than the other. However, it is important to note that participants in Experiment 2 are tested on different sentence-picture pairings than they received training on. A comparison to baseline rates of implicatures is therefore less straightforward than for Experiment 1. When looking at the last trials of the training phase, we see that training was overall effective, with the responses to training sentences being reflective of the training group. The

[9]The complexity is also reflected in the development of responses by trial number, which varies more than for Experiment 1, see Supplementary Materials for details.

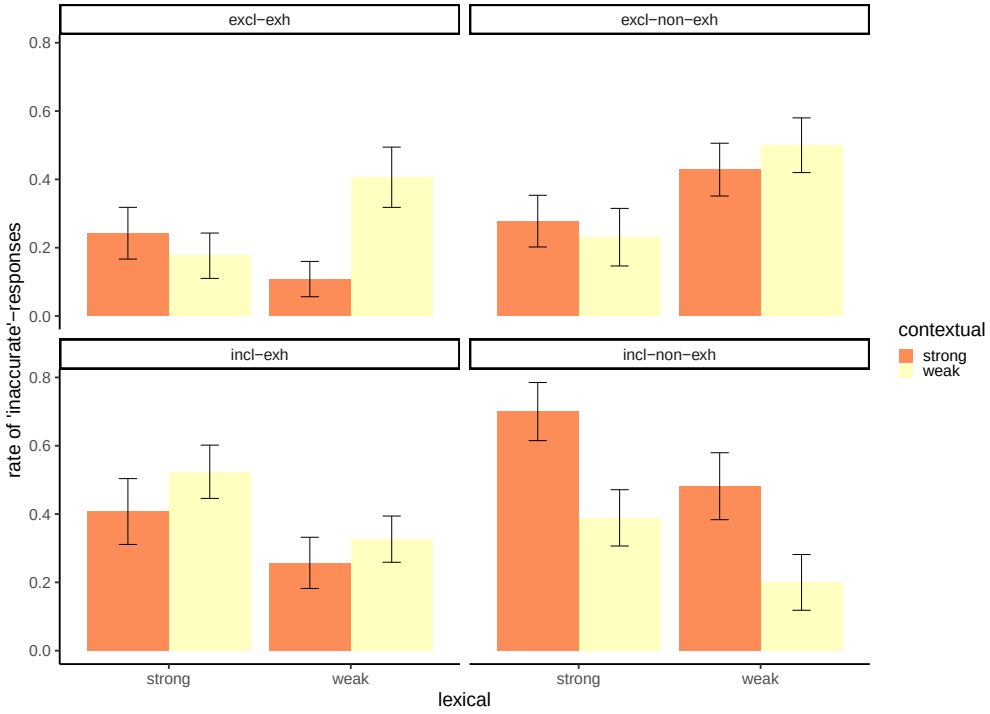


Figure 5: Rate of 'inaccurate'-responses in the training phase by training group (contextual strong-lexical weak, contextual strong -lexical strong, contextual weak-lexical weak, contextual weak-lexical strong) and sentence-picture pairings verifying one of the critical readings of simple disjunctions in the picture context (exhaustive-inclusive, exhaustive-exclusive, non-exhaustive-inclusive, non-exhaustive-exclusive) for the first 5 last trials.

performance of participants during the last 5 trials of training is depicted in Figure 5.

3.2.7 Discussion

We replicated the finding of Experiment 1 in that strong lexical training affects the rates of all types of implicatures involving lexical alternatives (simple and complex disjunctions (1) and (2)), and that strong contextual training increased the rate of exhaustive readings of simple sentences and conjunctions. That is, we again find evidence for training effects within the same alternative type. Moreover, we now observe that contextual training affects all (stronger) readings of simple and complex disjunctions. We also find an effect of lexical training on exhaustive readings of simple sentences and conjunctions. That is, we now also find training effects *across* alternative types, which is consistent with H_1 (*Uniformity*). We find that lexical training boosted exhaustive readings even when contextual training itself was weak (allowed for non-exhaustive

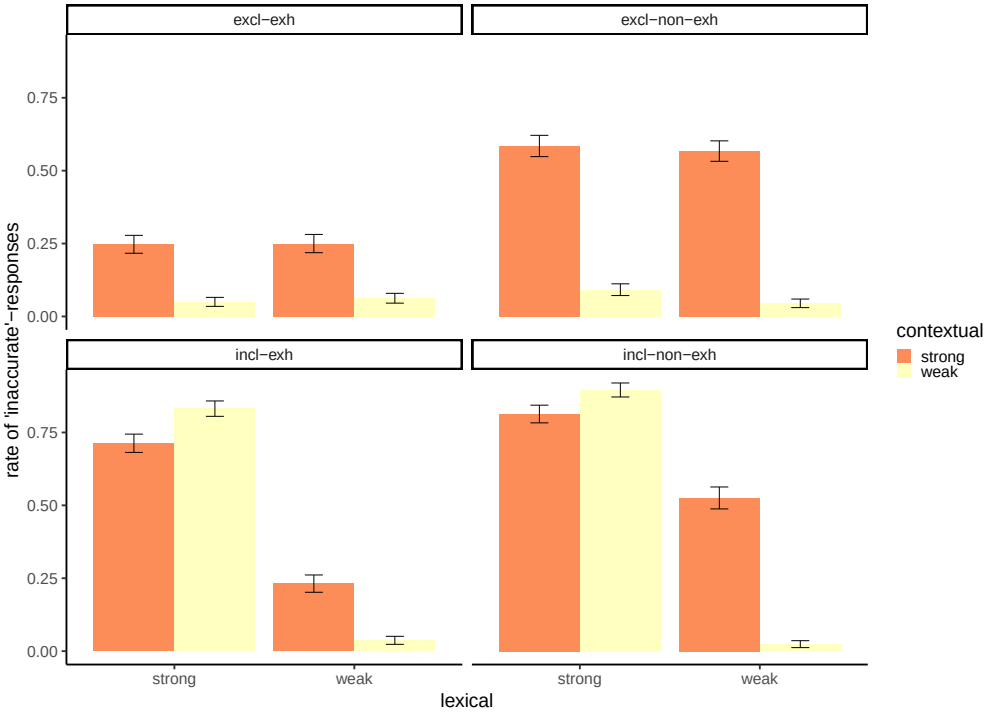


Figure 6: Rate of 'inaccurate'-responses in the training phase by training group (contextual strong-lexical weak, contextual strong -lexical strong, contextual weak-lexical weak, contextual weak-lexical strong) and sentence-picture pairings verifying one of the critical readings of simple disjunctions in the picture context (exhaustive-inclusive, exhaustive-exclusive, non-exhaustive-inclusive, non-exhaustive-exclusive) for last 5 trials.

readings). This suggests that the activation of the alternative was sufficient to trigger an exclusion mechanism for contextual alternatives, even though training did not include the exclusion of contextual alternatives. The same holds for the training effect in the other direction: strong contextual training led to an increased rate of exclusive readings of simple disjunctions, even when lexical training itself was weak (allowed for inclusive readings). Note that this effect cannot be due to participants being exposed to disjunction and presence of additional alternatives in the picture alone, as this was also true for the group receiving weak contextual training. Together with the fact that complex disjunction was affected by contextual training, this suggests that training participants to exclude contextual alternatives when faced with lexical cases opened up another path to getting to the strong readings of simple and complex disjunctions.

We do find asymmetries in how successful training is across alternative types compared to within the same alternative type. Contextual training had more impact on both readings of complex disjunctions than other critical sentence types. We see no difference between the effect of contextual training on exclusive versus exhaustive readings, however. By contrast, we do find that lexical training is more efficient for all disjunctions involving lexical alternatives, and less efficient in boosting exhaustive readings. We also see that contextual training has an additional effect on exhaustive readings when lexical training is strong, and lexical training has an additional effect on exclusive readings of disjunction when contextual training is strong. Taken together with the results of Experiment 1, we have evidence for H_2, *partial uniformity*. Specifically, we see effects of strong contextual and lexical training in both directions if a search of the visual context is required. We see that this search of the visual context was enough to trigger an exclusion mechanism, even if training itself did not necessitate exclusion of the relevant alternatives. However, we also see that the effect of training was even stronger when it did involve the exclusion of the exact alternative needed for the implicature (additive effect of one type of training when the other was already strong). This suggests that activation of alternatives and their exclusion should be considered *distinct mechanisms*. Activation might be sufficient in triggering exclusion, but training for both is more efficient than training for the activation of alternatives alone.

4 General Discussion

Using a new paradigm of training-with-feedback, we investigated whether deriving different types of implicatures—involving contextual versus lexical alternatives—share the same underlying mechanisms. We find different results for Experiment 1 and 2.

We only find training effects of simple disjunctions on other (simple and complex) disjunctions in Experiment 1. Participants were trained to derive exclusive or inclusive readings of simple disjunctions for lexical training, and exhaustive or non-exhaustive readings of simple sentences/conjunctions for contextual training. Lexical training affected the rate of intermediate and strong readings of complex disjunctions. That is, the effect of lexical training went beyond pushing participants for an interpretation of 'or' as 'Xor', thereby validating our novel methodology.

In Experiment 2 we replicated the positive results of Experiment 1, but we find additionally that training was effective *across alternative types* (lexical/contextual).

Training always involved simple disjunctions ('*a* or *b* is red') but differed with regard to (i) whether it allowed for the presence of additional red shapes in the picture (non-exhaustive readings) and (ii) whether it allowed for both *a* and *b* to be red (inclusive readings). Crucially, strong contextual training (disallowing for additional red shapes) affected the rate of exclusive readings of disjunction, even when participants were trained to interpret disjunction itself as inclusive. Furthermore, lexical training with exclusive disjunction made people derive more exhaustive inferences with contextual alternatives, even when training did not require the exclusion of these alternatives (allowing for additional red shapes).

Our results speak against a completely uniform view of quantity implicatures involving lexical versus contextual alternatives (H_1 from page 9). If the exclusion mechanism and the derivation of alternatives were uniform across these cases, we would have seen training effects across alternative types in Experiment 1. We do see training effects across alternatives types in Experiment 2 however, so our findings are also problematic for theories that propose entirely independent mechanisms and alternatives for the two types of implicatures (H_2).

Overall, our results are consistent with a *partial uniformity* view (H_3) where the same exclusion mechanism is involved in both types of implicatures, but the mechanism fails in the absence of the right alternatives. In Experiment 1, consulting the picture context for which of the shapes were red was not required to verify the exclusive reading of disjunction. It only required checking whether one or two of the shapes were red. We hypothesized that, as a result, lexical training was unsuccessful for increasing the rate of exhaustive readings. We checked this hypothesis with Experiment 2. Participants needed to consult the visual context when judging disjunctions, as they needed to check *which* of the three shapes in the picture were red. With this modified design, we found training effects across alternative types in both directions. This supports our hypothesis that some amount of attention to context is required for training effects across alternative types. We also found that training participants to *exclude* these alternatives further increased the rate of exhaustive readings with simple sentences and conjunctions. A higher rate of implicatures for simple and complex disjunctions was obtained through strong contextual training requiring the exclusion of contextual alternatives.

To what extent is the support we found for H_3 consistent with different theories of alternatives and their role in quantity implicatures? In terms of a structural view of alternatives, an explanation in line with our data is that requiring lexical substitution with simple disjunctions facilitated lexical substitution of scalar words for more complex cases of disjunction. However, as training across alternative types was not successful in Experiment 1, it did not automatically facilitate replacement of content words with *contextual* alternatives. This kind of distinction between alternative types speaks against the exact derivation mechanism proposed by Fox and Katzir (2011). Crucially, the advantage of that proposal is that structural and contextual alternatives are considered at the same time for both, and are subject to the same constraints. In Experiment 2, where contextual alternatives were made salient for disjunction due to the presence of another shape in the picture, training with exclusive disjunction did increase the rate of exhaustive inferences in the testing phase. This suggests that an additional threshold has to be met to consider contextual alternatives, and that they are not

automatically activated along with lexical alternatives. The fact that strong contextual training in Experiment 2 also boosted the rate of implicatures associated with simple and complex disjunctions further suggests that lexical substitution did not need to be activated to derive these implicatures. The result indicates that this training raised the possibility of local exclusion of contextual alternatives. That is, each of the disjuncts was exhaustified with regard to the contextual alternatives, irrespective of the strength of disjunction itself. Specifically, to derive the contextually strong and lexically weak meaning of disjunction, '*a* and *b*' might have been interpreted as '[*a* and not *c*] or [*b* and not *c*]', for *c* the additional shape in the picture. However, given that contextual training did not affect any type of disjunction in Experiment 1, our findings suggest that this strategy can only be activated when the sentence type in training was a disjunction, and the exclusion of contextual alternatives was trained. The mere presence of contextual alternatives with the use of disjunction cannot be considered a sufficient incentive for this strategy, as weak contextual training did not affect disjunctions in Experiment 2.

In the terms of a minimal model view, the results of Experiment 1 suggest that activating minimal alternatives from within the simple disjunction (*a, b*) might have prompted the activation of sentence internal alternatives for more complex cases (*a, b, c* for complex disjunction). However, since a context search was not required to build these models for disjunction, lexical training did not affect building models from contextual alternatives in Experiment 1. A context search was invoked for both types of training in Experiment 2 through the presence of a third, contextual alternative in the picture. As a result, they were now considered for building models for disjunction and simple sentences/conjunctions alike in the testing phase. Thus, the training affected the rate of strengthened meanings across alternative types.

Regarding a model for implicature computation, our results are at least partly compatible with a simple *salience model* as suggested by Rees and Bott (2018), in which the activation of alternatives and mechanism proceeds in one step. Our data offer a more fine-grained view on the *salience model*, however. Our findings suggest that it is not the activation of the exact alternative to be excluded that matters, but rather the activation of a specific *mechanism* for deriving alternatives. If the (higher) activation threshold for deriving contextual alternatives is met, these alternatives are considered for the exclusive interpretation of disjunction as well as exhaustive readings of simple sentences/conjunctions. Activating these alternatives was sufficient to trigger the mechanism of exclusion, as a training effect was there even when the activated alternative itself did not need to be excluded. However, our data show that if training also involved the exclusion of the exact alternative needed for the implicature, its rate increased even more. That is, we do find evidence that the two operations (activation and exclusion of the alternative) should be considered distinct steps, which can coincide or not. Further research is needed to see under which conditions they do fall together.

Further support for a model that distinguishes between the activation of contextual versus lexical alternatives (for disjunction) comes from a recent acquisition study. Gotzner, Barner, and Crain (2020) show that 4–5 year-old children calculate quantity implicatures with contextual alternatives to a higher degree than those involving disjunction. They suggest that contextual cases do not necessarily require access to lexical scales, which is in line with our findings in the case of adults. They also argue

that disjunction allows for the construction of sub-domains more easily than conjunction, based on the fact that children calculate more contextual implicatures based on disjunction than conjunction. In line with what Gotzner, Barner, and Crain (2020) find for children, disjunction did facilitate the search for contextual alternatives for adults, but only if these contextual alternatives were made salient.

Our results, especially Experiment 2, shed new light on what factors influence the activation of contextual alternatives. In both experiments, all participants were exposed to alternative ways of describing the picture scenario involving conjunction. Our findings suggest that exposure to the lexical alternative 'and' is not sufficient to meet the activation threshold for contextual cases, against the findings by Rees and Bott (2018). The global question '(Guess) what is red?,' which was stable across groups in both experiments, was not enough to meet the threshold either. Since it was not necessary to consider sentence-external alternatives for disjunction at all in Experiment 1, we included an additional element in the picture (not contained in the sentence) in Experiment 2. This was sufficient to make contextual alternatives salient or perhaps even relevant, highlighting the importance of the visual input. Our results overall suggest that the threshold for activating contextual alternatives is higher. If met, we see that there is overlap between implicatures based on lexical versus contextual alternatives, in line with previous findings (Bott and Chemla 2016; Degen and Tanenhaus 2015; Rees and Bott 2018).

Our findings further suggest that this overlap is due to the mechanism being shared, and automatically being successful when combining different methods for deriving lexical and contextual alternatives. It is important to note, however, that previous studies looked at other lexical cases (involving 'some'). These lexical cases might differ from disjunction as the alternatives of the latter can be found sentence internally. Furthermore, there is only one way of deriving the strong interpretation of 'some', which relies on the presence of other quantified expressions (Degen and Tanenhaus 2015). Another difference between previous studies and ours is the method: the training-with-feedback paradigm we used taught participants to employ a general global strategy for interpretation. The fact that this strategy was used across different sentence types suggests that it operates on a high level and is based on pragmatic interpretation, rather than learning an *ad hoc* ruled based on visual or lexical input. It speaks to the success of the methodology that our results are consistent with findings resulting from priming paradigms.

Our training paradigm can be modified in interesting ways to further study pragmatic strengthening. First, our guessing task allows for felicitous uses of disjunctions, and thus the paradigm is suitable for further testing their properties compared to other scalar expressions. Moreover, the fact that our training leads to long-term pragmatic strategies means that the paradigm can be used to test the effect of pragmatic behavior on other task types, for example reasoning behavior. In fact, the prediction would be that these strategies play a role in general purpose reasoning (Mascarenhas 2014). At least in principle then, the consequences of our findings go beyond exploring different alternative and scale types.

5 Conclusion

Results from two experiments employing a training-with-feedback paradigm offer new insight into the roles of different types of alternatives (contextual versus lexical) involved in implicature computation. We find that different activation thresholds for contextual access and lexical access to alternatives need to be included into models of implicature computation. Furthermore, our findings suggest that the presence of the relevant alternative expression in the experiment is not sufficient or necessary to meet a threshold for activation. Our findings are in line with a view where the difference between *particularized* and *generalized conversational implicatures* is rooted in this difference in activation thresholds, not in the involvement of two distinct mechanisms for excluding these alternatives. However, we also see evidence that excluding alternatives is a distinct step in implicature computation, and, if activated, leads to even higher rates of implicatures. Lastly, our results speak to the success of a new training-with-feedback method to investigate implicatures.

Funding acknowledgments. The research reported here was supported by *Agence Nationale de la Recherche* grant ANR-18-CE28-0008 (LANG-REASON; PI: Mascarenhas)

Data availability. The full set of materials is provided in the appendices below. Links to the experiments, analysis scripts and data are available via the following OSF link.

References

Bott, Lewis and Emmanuel Chemla (2016). "Shared and distinct mechanisms in deriving linguistic enrichment." In: *Journal of Memory and Language* 91, pp. 117–140.

Breheny, Richard, Heather J Ferguson, and Napoleon Katsos (2013). "Taking the epistemic step: Toward a model of on-line access to conversational implicatures." In: *Cognition* 126.3, pp. 423–440.

Breheny, Richard, Napoleon Katsos, and John Williams (2006). "Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences." In: *Cognition* 100.3, pp. 434–463.

Chemla, Emmanuel and Lewis Bott (2014). "Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice." In: *Cognition* 130.3, pp. 380–396.

Chierchia, Gennaro (2013). *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector (2012). "The grammatical view of scalar implicatures and the relationship between semantics and pragmatics." In: *Semantics: An International Handbook of Natural Language Meaning*. Ed. by Paul Portner, Claudia Maienborn, and Klaus von Heusinger. Berlin: Mouton de Gruyter.

Degen, Judith and Michael K Tanenhaus (2015). "Processing scalar implicature: A constraint-based approach." In: *Cognitive science* 39.4, pp. 667–710.

Fox, Danny (2007). "Free choice disjunction and the theory of scalar implicature." In: *Presupposition and Implicature in Compositional Semantics*. Ed. by Uli Sauerland and Penka Stateva. Pelgrave McMillan, pp. 71–120.

Fox, Danny and Roni Katzir (2011). "On the characterization of alternatives." In: *Natural Language Semantics* 19, pp. 87–107.

Franke, Michael (2011). "Quantity implicatures, exhaustive interpretation, and rational conversation." In: *Semantics and Pragmatics* 4, pp. 1–1.

Geurts, Bart (2010). *Quantity Implicatures*. Cambridge University Press.

Gotzner, Nicole, David Barner, and Stephen Crain (Apr. 2020). "Disjunction Triggers Exhaustivity Implicatures in 4- to 5-Year-Olds: Investigating the Role of Access to Alternatives." In: *Journal of Semantics* 37.2, pp. 219–245. ISSN: 0167-5133. DOI: 10.1093/jos/ffz021. eprint: https://academic.oup.com/jos/article-pdf/37/2/219/33211246/ffz021.pdf. URL: https://doi.org/10.1093/jos/ffz021.

Gotzner, Nicole, Isabell Wartenburger, and Katharina Spalek (2016). "The impact of focus particles on the recognition and rejection of contrastive alternatives." In: *Language and Cognition* 8, pp. 59–95.

Grice, Paul (1975). "Logic and conversation." In: *Syntax and Semantics: Speech Acts*. Ed. by P. Cole and J. Morgan. Vol. 3. New York: Academic Press.

— (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grodner, Daniel J et al. (2010). ""Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment." In: *Cognition* 116.1, pp. 42–55.

Groenendijk, Jeroen and Martin Stokhof (1984). "Studies in the Semantics of Questions and the Pragmatics of Answers." PhD thesis. University of Amsterdam.

Horn, Laurence (1972). "On the semantic properties of the logical operators in English." PhD thesis. UCLA.

Huang, Yi Ting and Jesse Snedeker (2009). "Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface." In: *Cognitive psychology* 58.3, pp. 376–415.

Katzir, Roni (2007). "Structurally-defined alternatives." In: *Linguistics and Philosophy* 30, pp. 669–690.

Koralus, Philipp and Salvador Mascarenhas (2013). "The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference." In: *Philosophical Perspectives* 27, pp. 312–365. DOI: 10.1111/phpe.12029.

Lenth, Russell V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. published online. https://CRAN.R-project.org/package=emmeans.

Levinson, Stephen C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Magri, Giorgio (2009). "A theory of individual-level predicates based on blind mandatory scalar implicatures." In: *Natural language semantics* 17.3, pp. 245–297.

Mascarenhas, Salvador (2014). "Formal Semantics and the Psychology of Reasoning: Building new bridges and investigating interactions." PhD thesis. New York University. URL: https://ling.auf.net/lingbuzz/002213.

Picat, Léo (2019). "Inferences with disjunction, interpretation or reasoning?" MA thesis. Ecole Normale Supérieure. URL: http://web-risc.ens.fr/~lpicat/website/picat-m2-thesis-pre-print.pdf.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rees, Alice and Lewis Bott (2018). "The role of alternative salience in the derivation of scalar implicatures." In: *Cognition* 176, pp. 1–14.

Sablé-Meyer, Mathias and Salvador Mascarenhas (2021). "Indirect illusory inferences from disjunction: a new bridge between deductive inference and representativeness." In: *Review of Philosophy and Psychology* 12.2. DOI: 10.1007/s13164-021-00543-8. URL: https://psyarxiv.com/pwkzm/.

Sauerland, Uli (2004). "Scalar implicatures in complex sentences." In: *Linguistics and Philosophy* 27, pp. 367–391.

Schulz, Katrin and Robert Van Rooij (2006). "Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation." In: *Linguistics and philosophy* 29.2, pp. 205–250.

Spector, Benjamin (2007). "Scalar implicatures: exhaustivity and Gricean reasoning." In: *Questions in Dynamic Semantics*. Ed. by Maria Aloni, Paul Dekker, and Alastair Butler. Elsevier.

— (2016). "Comparing exhaustivity operators." In: *Semantics & Pragmatics* 9.11, pp. 1–33.

Van Rooij, Robert and Katrin Schulz (2004). "Exhaustive interpretation of complex sentences." In: *Journal of logic, language and information* 13.4, pp. 491–519.

Van Tiel, Bob and Walter Schaeken (2017). "Processing conversational implicatures: alternatives and counterfactual reasoning." In: *Cognitive science* 41, pp. 1119–1154.

Waldon, Brandon and Judith Degen (2020). "Symmetric alternatives and semantic uncertainty modulate scalar inference." In: *Proceedings of the Annual Meeting of the Cognitive Science Society*.

A Theoretical Background

This section goes into details regarding the exact alternatives considered for the cases under discussion for different theories. The relevant sentence we look at are simple disjunction, complex disjunction, and simple sentences with contextual alternatives, the schematic depictions are repeated in (13) to (15) below.

(13) CONTEXT: $\{a, b\}$

a. $a \vee b$
'a or b' SIMPLE DISJUNCTION

b. S-M of (13a) = $(a \vee b) \wedge \neg(a \wedge b)$
'a or b but not both a and b'

(14) CONTEXT: {a, b}

a. a SIMPLE
'a'

b. S-M of (14a) = a $\wedge\neg$ b
'a and not b'

(15) CONTEXT: {a, b, c}

a. (a $\wedge$ b) $\vee$ c COMPLEX DISJUNCTION (2)
'a and b, or else c'

b. S-M of (15a) = (a $\wedge$ b $\wedge\neg$ c) $\vee$ (c $\wedge$ $\neg$ a $\wedge\neg$ b)
'a or b and not c, or else c and not a and not b'

A.1 Alternatives for exh$_{mm}$

For an operator using minimal models the same type of minimal alternatives can be assumed for all three cases above, {a,b} for (13) and (14), {a,b,c} for (15). The exact models construed differ for the first two cases, however. Disjunction has the models in Table 13, focus has the ones in Table 14.

w_1	w_2
a	$\neg$a
$\neg$b	b

Table 13: Models for disjunction

w_1	w_2
a	a
$\neg$b	b

Table 14: Models for simple sentences

Since neither of the two models is more minimal than the other in Table 13, the results of strengthening is the disjunction of both, i.e. *(a* $\wedge\neg$*b)* $\vee$ *(b* $\wedge\neg$*a)*. For the second table, it holds that $w_1 < w_2$.[10] We thus derive the S-M *a* $\wedge\neg$*b*. There are five different models compatible with the case *a and b, or c*, see Table 15.

w_1	w_2	w_3	w_4	w_5
$\neg$a	a	a	a	$\neg$a
$\neg$b	b	$\neg$b	b	b
c	c	c	$\neg$c	c

Table 15: Models for complex disjunction

w_1, w_3, w_4 and w_5 are all more minimal than w_2. Furthermore, $w_1 < w_3$ and $w_1 < w_5$. However, there are no models more minimal than w_1 and w_4. The strongest reading is the disjunction of both, see (16).

(16) (($\neg$a and $\neg$b) and c) or (a and b and $\neg$ c)

[10] Where '<' stands for 'is more minimal than' and is defined as fewer alternatives being true.

A.2 Alternatives for exh_{ie}

An operator using innocent exclusion requires alternatives derived by deletion or substitution. In the contextual case, it is the lexical content *a* that is substituted (with *b* or *c*, for example), depending on the context. In the scalar case it is the scalar term 'or', whose alternative is specified by the lexicon. One consequence of assuming the operations of deriving alternatives to be restricted in this was is that they can at most be as complex as the sentence they are derived from. For the lexical case, the alternative is derived by substituting the scalar item 'or' with 'and'. However, for the contextual case the lexical item *a* is replaced with the contextual alternative *b*. Conjunction is not part of the alternatives, as *a and b* is more complex. For the complex 'and-or' case, the strongest excludable alternative is *a or b, and c*, see the derivation in (17).[11]

(17) $\text{EXH-IE}_{(a \vee b) c}((a \wedge b) \vee c) = ((a \wedge b) \vee c) \wedge \text{not}((a \vee b) \wedge c)$
$((a \wedge b) \vee c) \wedge (\neg (a \vee b) \vee \neg c)$
$(\neg a \wedge \neg b \wedge c) \vee (a \wedge b \wedge \neg c)$

B Experiment 1

Table 16 on page 42 gives the full feedback matrix for contextual training. Table 17 on page 43 does the same for lexical training.

C Experiment 2

Tables 18 to 21 (pp. 43–45) give full matrices for all four combinations of INCLUSIVE/-EXCLUSIVE × EXHAUSTIVE/NON-EXHAUSTIVE.

[11] The set of original alternatives considered for this case is much larger, and their derivation raises two issues: one with plausibility given the computational resources necessary to get the right alternatives, the other with intermediate alternative being weaker than the one excluded, see Mascarenhas (2014) for more discussion and details.

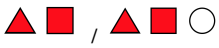
Sentence-picture	Group	Response	Feedback
The triangle is red. 	exhaustive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match'
The triangle is red. 	non-exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'
The triangle and the square are red. 	exhaustive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match!'
The triangle and the square are red. 	non-exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'
The triangle is red. 	exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle is red. 	non-exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle is red. 	exhaustive	yes	'Wait, that was actually a bad match!'
		yes	'Wait, that was actually a bad match.'
The triangle is red. 	non-exhaustive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match.'
The triangle and the square are red. 	exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle and the square are red. 	non-exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle and the square are red. 	exhaustive	yes	'Wait, that was actually a bad match.'
		no	'Great, that was indeed a good match!'
The triangle and the square are red. 	non-exhaustive	yes	'Wait, that was actually a bad match.'
		no	'Great, that was indeed a good match!'

Table 16: Appendix B: Experiment 1 feedback for contextual training according to training groups

Sentence-picture	Group	Response	Feedback
The triangle or the square is red. 	exclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match'
The triangle or the square is red. 	inclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'
The triangle or the square is red. 	exclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'
The triangle or the square is red. 	inclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'
The triangle or the square is red. 	exclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match'
The triangle or the square is red. 	inclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match'

Table 17: Appendix B: Experiment 1 feedback for lexical training according to training groups

Sentence-picture	Group	Response	Feedback
The triangle or the square are red. 	exhaustive, inclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match!'
The triangle or the square are red. 	exhaustive, exclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match'
The triangle or the square are red. 	non-exhaustive, inclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match!'
The triangle or the square are red. 	non-exhaustive, exclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'

Table 18: Appendix C: Experiment 2 feedback for exclusive, non-exhaustive sentence-picture pairings according to priming group.

Sentence-picture	Group	Response	Feedback
The triangle or the square are red.	strong focus, weak disjunction	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match!'
The triangle or the square are red.	exhaustive, exclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match'
The triangle or the square are red.	inclusive, non-exhaustive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match!'
The triangle or the square are red.	non-exhaustive, exclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match'

Table 19: Appendix C: Experiment 2 feedback for inclusive, exhaustive sentence-picture pairings according to training group.

Sentence-picture	Group	Response	Feedback
The triangle or the square are red.	exhaustive, inclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match!'
The triangle or the square are red.	exhaustive, exclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match!'
The triangle or the square are red.	non-exhaustive, inclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match!'
The triangle or the square are red.	non-exhaustive, exclusive	yes	'Wait, that was actually a bad match!'
		no	'Great, that was indeed a bad match!'

Table 20: Appendix C: Experiment 2 feedback for non-exhaustive, inclusive sentence-picture pairings according to training group.

Sentence-picture	Group	Response	Feedback
The triangle or the square are red. 	exhaustive, inclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle and the square are red. 	exhaustive, exclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle or the square are red. 	non-exhaustive, inclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'
The triangle and the square are red. 	non-exhaustive, exclusive	yes	'Great, that was indeed a good match!'
		no	'Wait, that was actually a good match.'

Table 21: Appendix C: Experiment 2 feedback for sentences that made the exclusive and exhaustive reading true for different training groups.