Probabilities and logic in implicature computation: two puzzles with embedded disjunction*†

Milica Denić

Abstract

Sentences are standardly assumed to trigger scalar and ignorance implicatures because there are alternative utterances the speaker could have said. The central question in modeling these inferences is thus: *what counts as an alternative utterance for a given sentence in a given context?* In this paper, I will present two families of novel empirical observations related to inferences and deviance patterns of embedded disjunction, based on which I will argue that (i) probabilistic informativeness plays a role in selecting the set of alternatives; and (ii) the role of prior world knowledge in evaluating probabilistic informativeness of alternatives is limited.

1 Introduction

A sentence in (1a) typically triggers the inference (*scalar implicature*) that (1b) is false.

(1) a. John ate a cookie or a muffin.
    b. John ate a cookie and a muffin.

There are different approaches to how scalar implicatures of sentences such as (1a) are computed (Grice, 1975; Sauerland, 2004; van Rooij and Schulz, 2004; Schulz and van Rooij, 2006; Spector, 2006, 2007; Chierchia et al., 2008; Franke, 2011; Bergen et al., 2014, a.o.). All models of scalar implicatures however rely on a set of alternative utterances that the speaker could have said: (1a) triggers as its implicature the negation of (1b) because (1b) is an alternative to (1a). The central question in modeling these inferences is thus: *what counts as an alternative utterance for a given sentence in a given context?*

By now a standard answer to this question is that alternatives of a sentence S are other sentences which convey contextually relevant information and which are obtained by replacing constituents

*Many thanks to Emmanuel Chemla and Benjamin Spector for important input throughout the development of this work. Many thanks also to Gennaro Chierchia, with whom I first discussed this work when it was still in its early stages, and who encouraged me to pursue it further. Discussions with Amir Anvari and Moshe E. Bar-Lev also contributed greatly to my thinking about these issues. For helpful suggestions and comments, thanks also to David Barner, Marc Benhamou, Brian Buccola, Fausto Carcassi, Rachel Dudley, Émile Enguehard, Danny Fox, Roni Katzir, Angelika Kratzer, Manuel Križ, Jeremy Kuhn, Mora Maldonado, Dean McHugh, Jacopo Romoli, Philippe Schlenker, Andrés Soria Ruiz, as well as the audiences of SALT 2018 and SuB 2018. The research leading to these results was supported by funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 313610 and n. STG 716230 CoSaQ, and by a postdoctoral fellowship awarded to MD by the Azrieli Foundation.

†An early version of a part of this work was published as 'Denić, Milica. A new case of pragmatically deviant embedded disjunctions. In: *Proceedings of Semantics and Linguistic Theory*. 2018. p. 454-473'.

of S with expressions of the same syntactic category and of smaller or equal structural complexity (Katzir, 2007; Fox and Katzir, 2011).

In this paper, I will present two families of novel empirical observations, which will be argued to have important consequences for the question of what counts as an alternative sentence in implicature computation. We will refer to the first family of observations as the ***inference puzzle***. An example of this puzzle is that inferences of sentences with an embedded disjunction are sensitive to the domain size of the quantifier which embeds it. Consider (2) and (3). These two sentences are structurally very similar; they differ essentially in the domain size of the universal quantifier. Strikingly, (2) and (3) trigger very different inferences. (2) is interpreted preferably as suggesting (4). (3) however is preferably interpreted without inferences in (4), and as suggesting that the speaker is ignorant about whether (4a) and (4b) hold.

Inference puzzle:

(2) All 20 of Mary's friends are French or Spanish.

(3) Both of Mary's friends are French or Spanish.

(4) a. At least one of Mary's friends is French.
b. At least one of Mary's friends is Spanish.

Based on this and related observations, I will argue that alternatives need to satisfy an informativeness criterion to enter implicature computation. There are two candidates as to which notion of informativeness this may be: entailment-based informativeness (if Sentence 1 asymmetrically entails Sentence 2, then Sentence 1 is more informative as Sentence 2), and probabilistic informativeness (if Sentence 1 expresses a proposition which is more likely to be true than the proposition expressed by Sentence 2, then Sentence 1 is more informative as Sentence 2). We will see that the data can be accounted with probabilistic, but not with entailment-based, notion of informativeness.

I will however argue that not all prior world knowledge is incorporated in probabilistic informativeness evaluation. This argument relates to a second family of observations, which will be referred to as the ***deviance puzzle***. An example of this puzzle is that (5a) and other structurally similar sentences with an embedded disjunction are degraded. This is surprising: (5a) should be able to convey the same meaning as (5b), yet clearly it cannot be used to do so.

Deviance puzzle:

(5) a. #Each of these three girls is Mary, Susan, or Jane.
b. These three girls are Mary, Susan, and Jane.

I will propose that the degraded status of (5a) is due to the inferences it triggers which contradict prior world knowledge. That inferences a sentence triggers may cause it to be degraded has already been argued for other cases of implicatures by Magri (2009); Meyer (2013); Marty (2017); Marty and Romoli (2022). Importantly, if these proposals are on the right track, whether an alternative sentence satisfies an informativeness criterion is evaluated based on probabilities that do not align with those from prior world knowledge.

2 Inference puzzle

Suppose that what's being discussed is where Mary's friends are from. Consider the example

(6), which will be referred to as ALL-20-OR[1] henceforth. When the disjunction is in the scope of a universal quantifier as in ALL-20-OR, it typically triggers *distributive inferences* in (6a) (Chemla (2009); Chemla and Spector (2011); Crnič et al. (2015); Chierchia et al. (2008); Fox (2007); Klinedinst (2007); Spector (2006), a.o). Accordingly, *ignorance inferences* in (6b) are typically absent: even though the speaker can in principle both believe ALL-20-OR and be in the epistemic state is as in (6b), we do not typically infer (6b) upon hearing ALL-20-OR.[2]

(6) **All 20 of Mary's friends are French or Spanish.**
  a. $\rightsquigarrow$ At least one of them is French.
     $\rightsquigarrow$ At least one of them is Spanish.
  b. $\not\rightsquigarrow$ The speaker is ignorant about whether at least one of them is French.
     $\not\rightsquigarrow$ The speaker is ignorant about whether at least one of them is Spanish.

A novel observation is that, strikingly, this inference pattern is sensitive to the ***cardinality of the restrictor of the universal quantifier***. Consider the sentence (7), which will be referred to as ALL-2-OR henceforth. ALL-2-OR is minimally different from ALL-20-OR in that the cardinality of the restrictor of the universal quantifier is 20 in ALL-20-OR and two in ALL-2-OR. This change in cardinality reverses the inference pattern of ALL-2-OR as compared to ALL-20-OR: ALL-2-OR no longer seems to trigger the distributive inferences in (7a). Instead, it is naturally interpreted as suggesting (7b).

(7) **Both of Mary's friends are French or Spanish.**
  a. $\not\rightsquigarrow$ At least one is French.
     $\not\rightsquigarrow$ At least one is Spanish.
  b. $\rightsquigarrow$ The speaker is ignorant about whether at least one of them is Spanish.
     $\rightsquigarrow$ The speaker is ignorant about whether at least one of them is French.

There may be contexts where, due to relevance considerations, ALL-20-OR and ALL-2-OR don't trigger any inferences (distributive or ignorance). We will remain agnostic as to whether ALL-20-OR ever triggers ignorance inferences, and whether ALL-2-OR ever triggers distributive inferences. We thus re-state the empirical facts in weaker terms: distributive inferences are, all other things being equal, more naturally derived for ALL-20-OR than for ALL-2-OR; ignorance inferences are more naturally derived for ALL-2-OR than for ALL-20-OR.

In other words, there is an interaction between the cardinality of the restrictor of the universal quantifier and inferences triggered by the sentence: the naturalness of distributive inferences is higher when the cardinality of the restrictor is large; the naturalness of ignorance inferences is higher when the cardinality of the restrictor is small.

Another novel observation is that, in addition to the effect of the cardinality of the restrictor, the inference pattern is also sensitive to the ***number of disjuncts*** in the sentence. Consider (8), which will henceforth be referred to as SIMPLE-DISJ and (9), which will be henceforth referred to as COMPLEX-DISJ. The restrictor of the universal quantifier has the same cardinality in these two

[1]A small number of representative examples, such as (6), are given names because they are referred to frequently in the paper.

[2]Distributive inferences have been mainly studied theoretically and empirically for disjunction in the scope of *universally quantified noun phrases*, which is why we focus on that environment in the paper. More empirical work on inferences of disjunction embedded under other quantificational noun phrases would be welcome however.

examples (four), but the number of disjuncts is different: there are two disjuncts in SIMPLE-DISJ, and four disjuncts in COMPLEX-DISJ. SIMPLE-DISJ is reported to be more naturally interpreted with distributive inferences than COMPLEX-DISJ; COMPLEX-DISJ is reported to be more naturally interpreted with ignorance inferences than SIMPLE-DISJ.

(8) **All four of Mary's friends are French or Spanish.**

(9) **All four of Mary's friends are French, Spanish, German, or Portuguese.**

In other words, there is an interaction between the number of disjuncts in a universally quantified sentence and inferences triggered by the sentence: the naturalness of distributive inferences is higher with fewer disjuncts; the naturalness of ignorance inferences is higher with more disjuncts.

How does the interaction of the cardinality of the restrictor of the universal quantifier and the number of disjuncts influence the inference pattern? The contrasts between ALL-20-OR and ALL-2-OR and between SIMPLE-DISJ and COMPLEX-DISJ are compatible with two different empirical generalizations, in (10) and (11).

(10) ***Threshold generalization***: When cardinality of the restrictor is larger than the number of disjuncts, distributive inferences are preferably derived, otherwise ignorance inferences are preferably derived.

(11) ***Gradient generalization***: The larger the cardinality of the restrictor compared to the number of disjuncts, the greater the preference for distributive instead of ignorance inferences.

Which of the two generalizations is on the right track? There is some initial evidence in favor of the gradient generalization, coming from the contrast between (12) and (13). (12) is reported by some speakers to be more naturally interpreted with distributive and without ignorance inferences than (13). The contrast is however reported to be more subtle than the one between ALL-20-OR and ALL-2-OR, and the one between SIMPLE-DISJ and COMPLEX-DISJ, and calls for further investigation.

(12) All five of Mary's friends are French or German.

(13) All five of Mary's friends are French, Spanish, German or Portuguese.

3 Implicatures or domain-general reasoning about the world?

An important question to address is whether the distributive and ignorance inferences reported above are truly *implicatures*, by which we mean that they are a result of consideration of alternative sentences that could have been used. Alternatively, they could simply be a result of domain-general reasoning about how the world might be based on the literal meanings of sentences. Take ALL-20-OR and ALL-2-OR for instance — intuitively, if all we know about Mary's friends is that each of them is either French or Spanish, the more friends she has, the more likely it is that at least one of them is French and the more likely it is that at least one of them is Spanish. Domain-general reasoning about how the world might be can similarly account for the contrasts between SIMPLE-DISJ and COMPLEX-DISJ, and between (12) and (13).

Here is one argument in favor of these inferences not being (solely) a product of domain-general reasoning about how the world might be. [3]

[3] In Section 4, we will see more data suggesting that these inferences are truly implicatures.

Suppose that we are discussing where Mary's office-mates are from. Consider (14).

(14) Both of Mary's office-mates are American or British.

Imagine that Mary works in the US, and that this fact translates into high prior probability that her office-mates are from the US (and hence low prior probability that they are from elsewhere). On the domain-general reasoning approach, we may expect that (14) should suggest that at least one of Mary's office-mates is American, and that the speaker may be ignorant about whether at least one is British (in other words, a salient interpretation of the sentence should be that at least one and possibly both of Mary's office-mates are American). This interpretation does not seem to be readily available: even in such a context, (14) is reported to be preferably interpreted with ignorance inferences and without distributive inferences, similarly to (7).

We will thus consider that distributive and ignorance inferences patterns reported in Section 2 should be explained by a theory of implicatures.

4 Deviance puzzle

We will now describe the second puzzle, which, among others, further challenges the idea that distributive and ignorance inferences of sentences from Section 2 are a product of domain-general reasoning about how the world might be.

When a disjunction of definite noun phrases is embedded in the scope of a universal quantifier, the result is sometimes unexpectedly deviant. The deviance depends on the predicate that embeds the disjunction. To see this, consider (15), which will be referred to as DEVIANT-BE, (16), which will be referred to as NON-DEVIANT-CALLED, (17), which will be referred to as DEVIANT-WRITE, and (18), which will be referred to as NON-DEVIANT-READ. When the predicate in question is the identity copula as in DEVIANT-BE or the predicate *to write* in DEVIANT-WRITE, the result is deviant. When the predicate in question is minimally different, as the predicate *to be called* in NON-DEVIANT-CALLED or the predicate *to read* in NON-DEVIANT-READ, the result is acceptable.

(15) (Context: Peter invited three girls to the party.)
#Each of those three girls **is** Mary, Susan, or Jane.

(16) (Context: Peter invited three girls to the party.)
Each of those three girls **is called** Mary, Susan, or Jane.

(17) (Context: Tolstoy, Zola and Rowling are great writers.)
#Each of those three writers **wrote** Anna Karenina, Germinal, or Harry Potter.

(18) (Context: Ann, John, and Bob are great students.)
Each of those three students **read** Anna Karenina, Germinal, or Harry Potter.

To see why the deviance of DEVIANT-BE and of DEVIANT-WRITE is surprising, note that DEVIANT-BE is contextually equivalent to (19), assuming that it is common knowledge that Mary, Susan, and Jane have to be three different individuals (see (21) for definitions of common knowledge, context set and contextual equivalence). Likewise, DEVIANT-WRITE is contextually equivalent to (20), assuming that it is common knowledge that for any book there can be exactly one singular or plural individual who wrote it[4]. Yet, surprisingly, DEVIANT-BE cannot be naturally

[4] In the case of co-authorship there would be exactly one plural individual who wrote the book.

used to do convey the meaning of (19), and neither can DEVIANT-WRITE to convey the meaning of (20).

(19) One of those three girls is Mary, another one is Susan, and yet another one is Jane.

(20) One of those three writers wrote Anna Karenina, another one wrote Germinal, and yet another one wrote Harry Potter.

(21) **Common knowledge:** A proposition ϕ is commonly known to a group of individuals if and only if all individuals in the group know that ϕ, all know that all know it, all know that all know that all know it, etc.
Context set: Context set is the set of possible worlds in which all the propositions that are common knowledge between the interlocutors are true.
Contextual equivalence: Two sentences A and B are contextually equivalent iff they have the same truth value in all the worlds of the context set.
(Stalnaker, 1973, 1978, 2002, a.o.)

Note that the deviance observed in DEVIANT-BE and DEVIANT-WRITE is not specific to *each*: the pattern is the same with *every* and *all*.[5]

We have observed the deviance of an embedded disjunction with certain predicates, such as the identity copula or *to write*, but not with others, such as *to be called* or *to read*. Which property makes a predicate pattern with one group or the other? We will argue that the essential property that identity copula and *to write* have in common is that when their internal arguments are, respectively, a specific individual and a specific book, they can only be true of a unique (singular or plural) individual given common knowledge. More formally, given common knowledge, they have the property in (23), which we will call *left-uniqueness*.

(23) A predicate P is left-unique iff
$\forall y$ in the relevant domain $\forall x[P(x,y) \Rightarrow \forall z[P(z,y) \Rightarrow z = x]]$

To see that the identity copula (when its internal argument is from a domain of individuals) and *to write* (when its internal argument is from a domain of books) are left-unique but not the predicates *to be called* (when its internal argument is from a domain of names) and *to read* (when its internal argument is from a domain of books), observe that the continuations in (24a) and (24c) sound contradictory, but not in (24b) and in (24d).

(24) a. This girl is my sister Susan. #That other girl is my sister Susan too.
b. This girl is called Susan. That other girl is called Susan too.
c. John wrote this book. #Peter wrote this book too.
d. John read this book. Peter read this book too.

To see that the left-uniqueness is indeed relevant for the phenomenon in question, consider

[5]The deviance is also not limited to universally quantified sentences; see for instance (22). For the simplicity of exposition, we will focus on the disjunction in the scope of universally quantified noun phrases; the main ideas that will be presented can in principle be extended to cases in (22).

(22) a. #These three girls are Mary, Susan, or Jane.
b. #Three of those girls are Mary, Susan, or Jane.

what happens when the internal argument of *to write* is not from a relevant domain for it to be left-unique. For instance, when its internal argument is from a domain of letters of the alphabet, the predicate *to write* is not left-unique, and note that (25), which is structurally similar to DEVIANT-BE and DEVIANT-WRITE, is not deviant (it could perfectly be used in a situation in which, for instance, each of John's three students wrote a number of letters on the board):

(25) Each of John's three students wrote the letter A, the letter D, or the letter K on the board.

Now, why would left-uniqueness be relevant for the deviance pattern of universally quantified sentences with embedded disjunction?

We will suggest that this is connected to the inference pattern we have observed in Section 2. We have seen that quantified sentences with embedded disjunction trigger ignorance inferences under certain conditions. Consider now what happens if the deviant DEVIANT-BE and DEVIANT-WRITE trigger ignorance inferences, which are in fact expected given the ratio between the cardinality of the restrictor and the number of disjuncts in these sentences (cf. empirical generalizations in Section 2). These ignorance inferences for DEVIANT-BE and DEVIANT-WRITE are paraphrased in (26) and (27), respectively.

(26) The speaker is ignorant about whether some of these three girls is Mary (Susan, Jane).[6]

(27) The speaker is ignorant about whether some of these three writers wrote Anna Karenina (Germinal, Harry Potter).

One can immediately see the problem with these ignorance inferences.

If we take it to be common knowledge that the speaker believes their own utterances in DEVIANT-BE/DEVIANT-WRITE, due to the left-uniqueness of the identity copula and the predicate *to write*, ***the speaker cannot be ignorant about the sentences in (28)*** — they must know that these sentences are true.

(28) Some of these three girls is Mary (Susan, Jane)/ Some of these three writers wrote Anna Karenina (Germinal, Harry Potter)

The ignorance inferences that sentences DEVIANT-BE and DEVIANT-WRITE might trigger thus contradict common knowledge. We propose that this is the reason why these sentences are deviant (cf. (29)).

(29) **Deviance due to ignorance inferences:**
Sentences DEVIANT-BE and DEVIANT-WRITE are deviant because they trigger *ignorance inferences* which contradict common knowledge.

The core assumption of the proposal in (29) is that implicatures generally, and ignorance inferences specifically, are derived *blindly* from common knowledge. What is meant by this is that, once the set of alternatives is determined, implicatures are derived even if they contradict common knowledge; note however that there are ways for common knowledge to influence which alterna-

[6]'The speaker is ignorant about whether some of these three girls is Mary (Susan, Jane)' and the like should be understood henceforth as an abbreviation for 'The speaker is ignorant about whether some of these three girls is Mary, The speaker is ignorant about whether some of these three girls is Susan, The speaker is ignorant about whether some of these three girls is Jane'.

tives feed implicature computation due to relevance or salience consideration, for which there is ample empirical evidence (Matsumoto, 1995; Fox and Katzir, 2011; Degen and Tanenhaus, 2016, a.o.). The idea that the procedure which derives implicatures is blind to common knowledge has been in fact already defended by Magri 2009 for the case of scalar implicatures in order to account for deviance of (30) (cf. also Meyer, 2013; Marty, 2017; Marty and Romoli, 2022 for related data and ideas).

(30) #Some Italians come from a warm country.

Simplifying a lot, according to Magri's proposal, (30) is deviant because the conjunction of (30) and its scalar implicature in (31) contradicts common knowledge.

(31) Not all Italians come from a warm country.

Additionally, there is other data suggesting that ignorance inferences may result in the deviance of sentences which triggered them when they contradict common knowledge. One such data point relates to the ignorance inferences of the modified numeral *at least*. We provide in (32) an example from Buccola and Haida (2018); similar empirical observations have been first made by Nouwen (2010). Given the context in (32), (32a) and (32b) are contextually equivalent; yet (32a) is deviant and (32b) is not. A possible explanation for why (32a) is deviant is because it triggers the inference that the speaker is ignorant about whether Ann scored exactly 3 points, which contradicts common knowledge.

(32) Context: Ann played a card game in which, given the rules, the final score is always an even number of points. Bob knows this, and reports to Carl:
a. #Ann scored at least 3 points.
b. Ann scored at least 4 points.

If DEVIANT-BE and DEVIANT-WRITE are deviant because the inferences they trigger contradict common knowledge, this further strengthens the argument that these inferences are implicatures rather than the result of domain-general reasoning about how the world might be. In light of the data from this section, a proposal aiming to account for the inference pattern in Section 2 needs to allow for implicatures to be derived blindly to common knowledge (or alternatively to put forward a different account for the data presented in this section).

5 Plan for the paper

We will start by introducing an exhaustification approach to implicature computation, and discuss the challenges posed for it by the inference and deviance puzzles. Importantly, the challenges are not specific to the exhaustification approach: any approach in which implicatures are a function of (solely) entailment relations between a sentence and its alternatives — in addition to possibly considerations of contextual relevance — will face similar challenges.

We will then put forward a proposal for how to incorporate probabilities into the exhaustication approach to implicatures so as to account for the inference puzzle, and discuss what additional assumptions are necessary to account for the deviance puzzle.

We will finish by discussing the remaining challenges and questions raised by our proposal. We will also discuss three alternative directions which could have been pursued to account for the inference and deviance puzzles, pointing out difficulties and open questions for each of these alternatives.

6 Implicatures of unembedded disjunction

We have introduced two types of implicatures so far: distributive and ignorance inferences. Distributive inferences are usually assumed to be a type of *scalar implicatures*. Let us see how scalar and ignorance inferences are derived according to the exhaustification approach to implicatures derivation (Fox, 2007; Chierchia et al., 2008).

Let us first consider a simple case such as (33). (33) triggers ignorance inferences in (33a) (Gazdar 1979, a.o.). In addition to them, (33) also triggers the scalar implicature in (33b).

(33) John is French or Spanish.
  a. The speaker is ignorant about whether John is French (Spanish).
  b. John isn't French and Spanish.

According to the exhaustification approach to implicatures, scalar implicatures are not the result of pragmatic reasoning. They are assumed to be a part of the logical meaning of the sentence as a result of the semantics of a silent exhaustivity operator *exh*. This operator is assumed to be present in the logical form of a sentence, as in (34).

(34) [*exh* [John is French or Spanish]]

The semantics of *exh* is given in (35), it is very similar to that of the focus operator *only* (Chierchia, 2006; Fox, 2007; Chierchia et al., 2008). In short, the semantic import of *exh* when it attaches to a sentence S is to negate alternatives activated by S, $ALT(S)$. There is however a restriction on the alternatives which can be negated: only those alternatives which are *innocently excludable (IE)* can be negated. IE alternatives of a sentence *S* are those which appear in every maximal set of alternatives of *S* which can be negated consistently with *S* (cf. (35b)).

(35)
  a. $Exh(S) = S \wedge \bigwedge_{q \in IE(S)} \neg q$
  b. $IE(S) = \bigcap\{A' \subseteq ALT(S)$: A' is a maximal set in $ALT(S)$ which can be negated consistently with $S\}$

What alternatives does a sentence activate? Simplifying somewhat, the *formal alternatives* (FA) of a sentence S are standardly assumed to be obtained by replacing the constituents of S with another expression of the same syntactic category and of smaller or equal structural complexity (Katzir, 2007; Fox and Katzir, 2011). The final set of alternatives a sentence S activates $ALT(S)$ in a given context are all those formal alternatives which are *relevant* in that context (cf. (36)) (Fox and Katzir, 2011).

(36) **Alternatives of a sentence *S***:
$ALT(S) = FA(S) \cap \{Y : Y$ expresses a contextually relevant proposition$\}$

Let us now see how the inferences of (33) are derived under this approach. We assume that $ALT((33))$ is in (37).

(37) Relevant formal alternatives of (33):

a. John is French (Spanish).
b. John is French and Spanish.

There are two maximal sets of alternatives of (33) which can be negated consistently with (33): (38a) and (38b).

(38) a. $\{$John is French, John is French and Spanish$\}$
b. $\{$John is Spanish, John is French and Spanish$\}$

The only IE alternative of (33) is thus 'John is French and Spanish', as it is the only alternative which appears in both (38a) and (38b). (33), parsed as (34), is thus interpreted as in (39).

(39) John is French or Spanish and he isn't French and Spanish.

How about the derivation of ignorance inferences? One approach[7] to ignorance inferences is pragmatic in nature: ignorance inferences of unembedded disjunction are a consequence of the maxim of quantity (Grice, 1975), according to which the speaker should convey all of the relevant information they have. We will adopt the version of the maxim of quantity in (40), adapted from Fox (2007):

(40) *Basic maxim of quantity:* If two sentences S and S' are both relevant to the topic of conversation, and S' is more informative than S, if the speaker believes both S and S' to be true, the speaker should say S' rather than S.

Let us see how ignorance inferences of (33) follow from the maxim of quantity in (40). Assume that in a context in which the sentence (33) is uttered and relevant, the sentences $S' =$ *John is French* and $S'' =$ *John is Spanish* are also relevant. As S' and S'' are more informative than (33) (both S' and S'' asymmetrically entail (33)), the maxim of quantity licenses the inferences in (41).

(41) The speaker doesn't believe that John is French (Spanish).

Assuming that the speaker believes their own utterance (33) (maxim of quality), the inferences in (41) amount to ignorance inferences in (42):

(42) The speaker is ignorant about whether John is French (Spanish).

More generally, assuming together with Fox (2007) that relevance is closed under conjunction and negation, ignorance inferences are predicted to be derived about any relevant sentence S' whose truth is not settled by the utterance S. The reason is that if S is relevant and S' is relevant, so is $S \wedge S'$, as well as $S \wedge \neg S'$. As both of these are more informative than S, the maxim of quantity licenses inferences that the speaker doesn't believe $S \wedge S'$ or $S \wedge \neg S'$: together with the maxim of quality this amounts to the ignorance inference about S'.

[7] Another approach to ignorance inferences within the exhaustification framework derives ignorance inferences as semantic inferences, that is, within grammar (Meyer, 2013, 2014; Buccola and Haida, 2018), cf. also discussion in Section 10.

7 Embedded disjunction: a problem

Let us now see what implicatures are predicted under the exhaustification approach for ALL-20-OR and for ALL-2-OR.

The predictions of any theory of implicatures for a given sentence depend on the alternatives that the sentence is assumed to activate. ALL-20-OR and ALL-2-OR have two scalar items, both of which can activate alternatives: the universal quantifier (*all*, *both*), and the disjunction *or*. If both of these scalar items activate their alternatives, the set of formal alternatives consists of all sentences in which the universal quantifier, the disjunction, or both, are replaced by alternative expressions they activate. For presentation purposes, we will focus only on alternatives without connectives, as the other alternatives don't play a role in distributive and ignorance inference derivation[8]. Concretely, for ALL-20-OR and ALL-2-OR, this means that the alternatives are as in (43) and (44) respectively: we will henceforth refer to the set of alternatives in (43) and (44) as *ALT-all-or*.

(43) *ALT-all-or*(ALL-20-OR):
  a. All 20 are French
  b. All 20 are Spanish
  c. Some are French
  d. Some are Spanish

(44) *ALT-all-or*(ALL-2-OR):
  a. Both are French
  b. Both are Spanish
  c. Some are French
  d. Some are Spanish

Another possibility is that only one of the two scalar items activates alternatives. If only the disjunction activates its alternatives, the formal alternatives of ALL-20-OR and ALL-2-OR are in (45) and (46) respectively.[9][10] We will henceforth refer to the set of alternatives in (45) and (46) as *ALT-or*.[11]

(45) *ALT-or*(ALL-20-OR):
  a. All 20 are French
  b. All 20 are Spanish

(46) *ALT-or*(ALL-2-OR):
  a. Both are French
  b. Both are Spanish

Which implicatures are predicted for ALL-20-OR and for ALL-2-OR by the exhaustification approach, under each of the two sets of alternatives *ALT-all-or* and *ALT-or*? To give a preview of what follows, ignorance inferences are predicted for both ALL-20-OR and ALL-2-OR under the set of alternatives *ALT-all-or*, while distributive inferences are predicted for both ALL-20-OR and ALL-2-OR under the set of alternatives *ALT-or*. Let us see why.

If ALL-20-OR activates the alternatives *ALT-all-or*, three different maximal sets of alternatives as in (47) can be negated consistently with ALL-20-OR:

(47) a. {All 20 are French, All 20 are Spanish}

[8]The actual set of alternatives *ALT-all-or* in (43) would contain *All 20 are French and Spanish*, *Some are French and Spanish*, and *Some are French or Spanish*. The first two can be shown to be innocently excludable with no consequences for distributive or ignorance inferences, and the third one is entailed by the assertion if the domain of individuals is non-empty. Similarly for *ALT-all-or* in (44).

[9]Note that if only the universal quantifier activates its alternatives, the only alternative of ALL-20-OR and ALL-2-OR would be 'Some is French or Spanish': if the restrictor of the universal quantifier is non-empty, this alternative is entailed by the original sentence, so no implicatures are derived.

[10]The actual set of alternatives *ALT-or* in (45) would contain *All 20 are French and Spanish*, which can be shown to be IE with no consequences for distributive or ignorance inferences. Similarly for *ALT-or* in (46).

[11]Fox (2007) and Magri (2009) assume that the *ALT-or* alternatives are the only alternatives that sentences such as ALL-20-OR and ALL-2-OR activate; see also the discussion in Bar-Lev and Fox (2017), fn. 7.

b. {All 20 are French, Some are French}
c. {All 20 are Spanish, Some are Spanish}

No alternative appears in all three sets in (47), hence no alternative is IE. Assuming that the alternatives 'All 20 are French (Spanish)', 'Some are French (Spanish)', are relevant, ignorance inferences about them are derived as a consequence of the maxim of quantity in (40). The same applies to ALL-2-OR.

Let us now see what the predictions are if ALL-20-OR activates the alternatives *ALT-or*. All alternatives in *ALT-or* can be negated consistently with (43), i.e. they are all IE. This accounts for the distributive inferences: the negation of (45a) together with ALL-20-OR entails that some of Mary's friends are Spanish, and the negation of (45b) together with ALL-20-OR entails that some of Mary's friends are French. The same applies to ALL-2-OR.

How do these predictions match the actual inferences people get with ALL-20-OR and with ALL-2-OR? We have seen that ALL-2-OR preferably triggers ignorance inferences while ALL-20-OR preferably triggers distributive inferences. Assuming that sentences ALL-20-OR and ALL-2-OR activate the alternatives *ALT-all-or*, correct inferences are predicted for ALL-2-OR but not for ALL-20-OR. Assuming alternatively that sentences ALL-20-OR and ALL-2-OR activate the alternatives *ALT-or*, correct inferences are predicted for ALL-20-OR but not for ALL-2-OR.[12]

To summarize, ALL-20-OR triggers distributive inferences more naturally than ALL-2-OR; ALL-2-OR triggers ignorance inferences more naturally than ALL-20-OR. The exhaustifiction approach to implicature derivation, as it stands, cannot capture this difference. The reason for this is fully general. According to the exhaustification approach to implicatures, similarly to many other approaches, implicatures are a function of the entailment relations between a sentence and its alternatives (in the exhaustification approach, this follows from the semantic entry of *exh*). Crucially, to the extent that ALL-20-OR and ALL-2-OR activate comparable sets of alternatives, they stand in the same entailment relations to them, and will thus necessarily be predicted to have the same implicatures.

8 How about relevance?

In the previous section, we have explained why the contrast in the inference pattern of ALL-20-OR and ALL-2-OR is problematic for the exhaustification-based approach: if ALL-20-OR and ALL-2-OR activate comparable sets of alternatives, they stand in the same entailment relations to them, and will thus necessarily be predicted to have the same implicatures.

It is however standardly assumed that contextual relevance plays a role in which alternatives enter implicature computation. In other words, some of the formal alternatives can sometimes be 'ignored' when implicatures are computed because they don't convey contextually relevant information — the term that is usually employed for this is *alternative pruning* (Horn, 1972; Fox and Katzir, 2011; Katzir, 2014; Crnič et al., 2015; Bar-Lev, 2018). This assumption is incorporated in the definition of the set of alternatives of a sentence S, $ALT(S)$, as the set of formal alternatives

[12] This is a simplification — it has been argued by Crnič et al. (2015) that distributive inferences of sentences such as ALL-20-OR should be derived not by negating alternatives 'All 20 are French' and 'All 20 are Spanish', but by negating a modified version of these alternatives, namely 'All 20 are only French' and 'All 20 are only Spanish'. For simplicity of exposition, we will consider that *ALT-or* alternatives predict correct inferences for sentences such as ALL-20-OR in the main body of the paper, and come back to Crnič et al. (2015) in Appendix. We will discuss in Appendix how the proposal we will put forward can be incorporated within the proposal of Crnič et al. (2015).

which express contextually relevant propositions (cf. (36)), repeated below: essentially, formal alternatives expressing contextually irrelevant propositions are pruned from $ALT(S)$, i.e. they are 'ignored' in the process of implicature computation.

(36) **Alternatives of a sentence *S***:
$ALT(S) = \{X : X \in FA(S)\} \cap \{Y : Y \text{ expresses a contextually relevant proposition}\}$

Could it then be that ALL-20-OR and ALL-2-OR do not activate comparable sets of alternatives because of contextual relevance, which would eliminate the problem for the exhaustification approach? In particular, could it be that ALL-2-OR activates the alternatives *ALT-all-or*, while ALL-20-OR activates the alternatives *ALT-or*?

Recall the discussion from Section 7: if ALL-20-OR and ALL-2-OR activate alternatives *ALT-all-or*, ignorance inferences are derived, and if they activate alternatives *ALT-or*, distributive inferences are derived. Importantly, note that *ALT-all-or* is a superset of *ALT-or*. This allows for the following theoretical possibility. ALL-20-OR and ALL-2-OR have as their *formal* alternatives *ALT-all-or*. When no alternatives are pruned from this set of formal alternatives, ignorance inferences are derived. When alternatives obtained by replacing the universal quantifier with the existential (we will refer to these as *existential alternatives* henceforth) are pruned, distributive inferences are derived.

Crucially, if existential alternatives could be preferably pruned from the alternative set of ALL-20-OR but not from the alternative set of ALL-2-OR (i.e., if the alternatives of ALL-20-OR and ALL-2-OR were preferably as in Table 1 and Table 2 respectively), this would resolve the tension between the exhaustification approach to implicatures and the contrast between ALL-20-OR and ALL-2-OR.

Alternatives of ALL-20-OR	Inferences
All 20 are French	**Some are Spanish**
All 20 are Spanish	**Some are French**
~~**Some are French**~~	—
~~**Some are Spanish**~~	—

Table 1: Left: Alternatives of ALL-20-OR with the existential alternatives pruned (in strikethrough text). Right: Inferences of ALL-20-OR which result from the alternatives on the left-hand side of the table.

Alternatives of ALL-2-OR	Inferences
Both are French	**ignorance**
Both are Spanish	**ignorance**
Some are French	**ignorance**
Some are Spanish	**ignorance**

Table 2: Left: Alternatives of ALL-2-OR (no alternatives are pruned). Right: Inferences of ALL-2-OR which result from the alternatives on the left-hand side of the table.

However, we will see below that such a contrast in pruning preferences is not predicted by existing approaches to pruning.

Namely, it has been recognized that we cannot prune just any alternative: that would create a massive overgeneration problem (Fox and Katzir, 2011). Take for instance unembedded disjunction, as in (33):

(33) John is French or Spanish.

Recall that its formal alternatives are {John is French, John is Spanish, John is French and Spanish}. If we could simply prune the alternative 'John is French', (33) would have as implicature that John is not Spanish. This implicature arguably never arises. This example, among many others, motivated developing explicit proposals about what kind of alternatives can be pruned.

A proposal that is grounded in contextual relevance considerations has been put forward by Fox and Katzir (2011). They propose two ways in which pruning is constrained.

The first is a constraint on the set of contextually relevant sentences. Following von Fintel and Heim (1997); Fox (2007), Fox and Katzir (2011) propose that (i) if a sentence S is relevant, so is $\neg S$, and (ii) if sentences S_1 and S_2 are relevant, so is $S_1 \wedge S_2$.

The second is a constraint on how the set of contextually relevant sentences can restrict the set of formal alternatives. They propose that the set of formal alternatives $FA(S)$ can be restricted via pruning to the set $ALT(S)$ if and only if the following conditions are met:

a. $S \in ALT(S)$
b. No member of $FA(S) \setminus ALT(S)$ is *exhaustively relevant* given $ALT(S)$
(where p is exhaustively relevant given $ALT(S)$ is exhaustifying p with respect to $ALT(S)$ is in the Boolean closure of $ALT(S)$)

Do these two constraints allow for restricting *ALT-all-or* to *ALT-or* by pruning? It can be shown that their second constraint in principle allows for restricting *ALT-all-or* to *ALT-or*.[13] Importantly, however, there is nothing in the proposal which would predict that the restriction from *ALT-all-or* to *ALT-or* should be more often done with ALL-20-OR than with ALL-2-OR.

There are other approaches to constraints on pruning, which are not rooted in contextual relevance considerations. Simplifying somewhat, Crnič et al. (2015) propose that one can only prune alternatives of S if the exhaustification of S with respect to the set of alternatives after pruning results in a weaker interpretation than the exhaustification of S with respect to the set of alternatives before pruning. This approach wouldn't help for the contrast between ALL-2-OR and ALL-20-OR: pruning existential alternatives results in distributive inferences, not pruning them results in ignorance inferences. The two interpretations are logically independent: the constraint on pruning by Crnič et al. (2015) is thus incompatible with restricting *ALT-all-or* to *ALT-or* via pruning. Bar-Lev (2018) argues for a stronger version of Crnič et al. (2015) (he argues that additional criteria need to be satisfied for pruning to be possible); his proposal is thus incompatible with restricting *ALT-all-or* to *ALT-or* via pruning for the same reason as that of Crnič et al. (2015).

[13]To see this, consider what happens if *Some are French* is exhaustified with respect to {*All are French or Spanish, All are French, All are Spanish, All are French and Spanish*}. All of these alternatives are innocently excludable; the result is thus *some are French and not all are French or Spanish* (which entails that not all are French, not all are Spanish, and not all are French and Spanish). This is not in the Boolean closure of {*All are French or Spanish, All are French, All are Spanish, All are French and Spanish*} — in other words, *Some are French* is not exhaustively relevant with respect to this set of alternatives. It can similarly be shown that *Some are Spanish*, as well as *Some are French and Spanish* and *Some are French or Spanish*, are not exhaustively relevant with respect to {*All are French or Spanish, All are French, All are Spanish, All are French and Spanish*}.

Let's take stock. Exhaustification approach — and any approach in which implicatures are a function of entailment relations between a sentence and its alternatives — coupled with existing approaches to pruning, cannot account for the contrast between ALL-2-OR and ALL-20-OR or other observations pertaining to the inference puzzle.

9 Proposal and Inference puzzle

In this section, I will propose that evaluation of informativeness of sentences is incorporated into alternative pruning, which can account for the inference puzzle. Importantly, however, that we can derive implicatures contradicting world knowledge (cf. the deviance puzzle) suggests that such informativeness evaluation is computed independently of (at least some aspects of) prior world knowledge — this will be discussed in Section 10.

Recall that the contrast between ALL-20-OR and ALL-2-OR could be derived if for the latter but not for the former it were possible to preferably derive the set of alternatives *ALT-or* from *ALT-all-or* by pruning existential alternatives.

The proposal we will put forward that achieves this has two components. The first component is that *alternative pruning is, in addition to contextual relevance, also sensitive to how informative alternatives are*. The set of alternatives of a sentence S, $ALT(S)$, is thus defined in (48).

(48) **Alternatives of a sentence *S*: proposal**
$ALT(S) = \{X : X \in FA(S)\} \cap \{Y : Y$ expresses a contextually relevant proposition$\} \cap$ $\{Z : Z$ expresses an informative proposition$\}$

The second component of the proposal states that informativeness employed for pruning is probabilistic: the more unlikely the proposition expressed by an alternative sentence is, the more informative the alternative sentence is (cf. Shannon (1948)). For presentation purposes, we start with a very simple version of the proposal which only cares about the informativeness of the alternatives, and not about the informativeness of the original utterance.

(49) **Informative propositions and pruning: proposal *(to be revised)***
Let A be a formal alternative of S, and $P(A)$ the probability that A is true. The probability to prune A from $ALT(S)$ increases with $P(A)$ (and thus decreases with the informativeness of A!).

An intuitive reason for why (49) might hold of pruning is that the more likely an alternative A is to be true, the less pressure there is for the speaker to utter A, and thus the less pressure there is to consider it as an alternative utterance the speaker could have said instead of their original utterance S.

Let us first see how this proposal accounts for the contrast between ALL-20-OR and ALL-2-OR. In particular, for some domain size n, for a sentence of the form (50), let us consider what happens with its alternatives of the form (50a,b).

(50) All of n people are A or B.
 a. Some of n people are A (B)
 b. All of n people are A (B)

Let us make intuitively plausible assumptions that (i) ***for any domain of individuals D, for any***

predicate A, the larger the cardinality of D, the more likely it is that someone in D is in A, and that (ii) ***if $|D| > 1$, then it is (strictly) more likely that someone in D is in A than that everyone in D is in A***. These assumptions plausibly hold when the interlocutors possess little prior world knowledge about individuals in the domain D (apart from how many of them there are) and about the property A. Obviously, these assumptions are most often not met in the actual world. We will however adopt them for the time being; we will see in Section 10 arguments that implicature computation proceeds as if these assumptions are met (i.e., as if access to prior world knowledge is limited).

Under these assumptions, the alternative such as 'Some of the n individuals are A' is more likely to be true for larger ns, and therefore it is more likely to be pruned from some set of alternatives for larger ns. In addition, a sentence of the form 'All of the n individuals are A' is less likely to be pruned from some set of alternatives than the sentence of the form 'Some of the n individuals are A' as soon as the cardinality of the domain of individuals is larger than 1. It thus follows that we are more likely to end up with the restricted set of alternatives which yields distributive inferences for larger ns than for smaller ns, and that we are more likely to end up with the full set of alternatives which yields ignorance inferences for smaller ns than for larger ns.

Concretely, this means that ALL-20-OR is more likely to have the set of alternatives as in Table 1 than ALL-2-OR is to have a parallel set of alternatives: with such a set of alternatives, distributive inferences are derived. Furthermore, ALL-2-OR is more likely to have the set of alternatives as in Table 2 than ALL-20-OR is to have a parallel set of alternatives: with such a set of alternatives, ignorance inferences are derived (cf. Section 7). Distributive and not ignorance inferences are thus more likely to be derived with ALL-20-OR than with ALL-2-OR, and ignorance and not distributive inferences are more likely to be derived with ALL-2-OR than with ALL-20-OR.

The proposal in (49) can thus capture the contrast between ALL-20-OR and ALL-2-OR.

However, the proposal does not yet capture that the inference pattern is sensitive to the number of disjuncts. As a reminder, consider (8) (i.e. SIMPLE-DISJ) and (9) (i.e. COMPLEX-DISJ), repeated below: SIMPLE-DISJ is more naturally interpreted with distributive inferences than COMPLEX-DISJ, and COMPLEX-DISJ more naturally with ignorance inferences than SIMPLE-DISJ.

(8) All four of Mary's friends are French or Spanish.

(9) All four of Mary's friends are French, Spanish, German, or Portuguese.

Let us see why the proposal in (49) cannot capture this by focusing on inferences predicted for COMPLEX-DISJ. COMPLEX-DISJ is predicted to trigger ignorance inferences when no alternatives are pruned, as in Table 3, and distributive inferences when the existential alternatives of the form 'Some are French', 'Some are French or Spanish' etc. are pruned, as in Table 4.[14]

[14]In table cells of Table 3 and 4 '...' in the 'Alternatives' column is intended to convey 'and other alternatives with the same number of disjuncts or conjuncts'; '...' in the 'Inferences' column is intended to convey 'the inferences derived from the alternatives with the same number of disjuncts/conjuncts'.

Alternatives of COMPLEX-DISJ	Inferences
All 4 are French...	**ignorance**
All 4 are French or Spanish...	**ignorance**
All 4 are French or Spanish or German...	**ignorance**
All 4 are French and Spanish...	Not all 4 are French and Spanish...
Some are French...	**ignorance**
Some are French or Spanish...	**ignorance**
Some are French or Spanish or German...	**ignorance**
Some are French and Spanish...	No one is French and Spanish

Table 3: Left: Alternatives of COMPLEX-DISJ (no alternatives are pruned). Right: Inferences of COMPLEX-DISJ which result from the alternatives on the left-hand side of the table.

Alternatives of COMPLEX-DISJ	Inferences
All 4 are French...	**Someone is Spanish, German, or Portuguese**...
All 4 are French or Spanish...	**Someone is German or Portuguese**...
All 4 are French or Spanish or German...	**Someone is Portuguese**...
All 4 are French and Spanish...	Not all 4 are French and Spanish
~~**Some are French**...~~	—
~~**Some are French or Spanish**...~~	—
~~**Some are French or Spanish or German**...~~	—
Some are French and Spanish...	No one is French and Spanish

Table 4: Left: Alternatives of COMPLEX-DISJ with the existential alternatives pruned (in strike-through text). Right: Inferences of COMPLEX-DISJ which result from the alternatives on the left-hand side of the table.

The problem of the proposal in (49) is the following: we have assumed that *P(Some of Mary's friends are French)* depends on the total number of Mary's friends, but nothing we have said so far relates it to the number of disjuncts in the original utterance. In other words, whether alternatives such as 'Some of Mary's friends are French' are pruned and hence whether distributive or ignorance inferences are derived is expected to vary as a function of the total number of Mary's friends, rather than as a function of the number of disjuncts in the sentence.

A very minor refinement of the proposal, in (51), solves these two problems.

(51) **Informative propositions and pruning: proposal *(final)***
Let A be a formal alternative of S, and $P(A \mid S)$ the conditional probability that A is true given that S is true. The probability to prune A from $ALT(S)$ increases with $P(A \mid S)$ (and thus decreases with the informativeness of A given S!).

An intuitive reason for why a constraint on pruning such as (51) might hold is related to what has been said to conceptually motivate (49): the more likely an alternative A is to be true given the utterance S (the closer it is to being entailed by S), the less pressure there is for the speaker to utter both A and S instead of only S if they believe both A and S to be true.

How does this modification capture the contrast between ALL-20-OR and ALL-2-OR, and between SIMPLE-DISJ and COMPLEX-DISJ?

Let us start with the contrast between ALL-20-OR and ALL-2-OR. We have already seen that, under the intuitive assumptions discussed above, as n increases, so does the P(Someone is A | n people are A or B). Alternatives such as 'Some of Mary's friends are French (Spanish)' are more likely to be pruned from *ALT(*ALL-20-OR*)* than from *ALT(*ALL-2-OR*)*. This will result in distributive inferences being more likely to be derived in the case of ALL-20-OR than in the case of ALL-2-OR, and ignorance inferences being more likely to be derived in the case of ALL-2-OR than in the case of ALL-20-OR.

Let us now see how the revised proposal also captures the difference between SIMPLE-DISJ and COMPLEX-DISJ.

Let us make another intuitively plausible assumption[15] that ***in a domain D of n people it is less likely that there is someone in D who is A_1 when it is known that everyone in D is A_1, A_2, A_3,... or A_n, than when it is known that everyone in D is A_1, A_2, A_3,... or A_m, with $m < n$.***

Because of this we may conclude that the likelihood of pruning alternatives of the form 'Some of the n individuals are A' decreases with the number of disjuncts of the original sentence (again, this is true for any A). It thus follows that the set of alternatives that we will end up with is more likely to be the set without the existential alternatives (i.e. without the alternatives such as 'Some of the n individuals are A') for sentences with smaller numbers of disjuncts. Concretely, this means that we will be more likely to derive distributive inferences for SIMPLE-DISJ than for COMPLEX-DISJ, and more likely to derive ignorance inferences with COMPLEX-DISJ than with SIMPLE-DISJ.

There is however an important piece in the present proposal that is left underspecified. How exactly does $P(A|S)$ (conditional probability that the alternative A is true given that the sentence S is true) map to the probability of pruning A from the set of alternatives of S? The formulation of the proposal in (51) states that the function from the first set of probabilities to the second set of probabilities is some increasing function, but we haven't specified what kind of increasing function it is, let alone said anything about its parameters. To be able to specify this part of the proposal, experimental and computational work is necessary. There are many options to be explored. We will mention just two. One interesting possibility is that the probability of pruning A from the set of alternatives of S increases linearly with $P(A|S)$, i.e., $P($pruning A from $ALT(S)) = a \cdot P(A|S) + b$. Another interesting possibility is that there is a threshold θ, such that if $P(A|S) \geq \theta$, A is pruned from the set of alternatives of S. That would entail, for instance that, $P($'Someone is French'|'Both are French or Spanish'$)$ is often, if not always[16], lower than θ. There are a lot of outstanding questions to be pursued for each of these possibilities. How are parameters of the function (i.e., θ, a, b) computed? Can they be affected by certain aspects of the context, and if so, which? Can they be affected by aspects of literal meaning of the sentence, and if so, which? Do they vary across different people?

There is an important empirical question that is related to this. We discussed in Section 2 whether the judgments about distributive or ignorance inferences of universally quantified sentences with embedded disjunction are better described by *threshold* or by *gradient* generalization.

[15] Again, this assumption may not always be met in the actual world; see discussion above.

[16] One could argue that it is always lower than θ if it was shown that sentences such as 'Both are French or Spanish' never trigger distributive inferences. As we pointed out in Section 2, we are not sure about the data. Investigating this point is thus left for future empirical work.

According to the threshold generalization, when cardinality of the restrictor is larger than the number of disjuncts, distributive inferences are preferably derived, otherwise ignorance inferences are preferably derived. According to the gradient generalization, the larger the cardinality of the restrictor compared to the number of disjuncts, the greater the preference for distributive instead of ignorance inferences. The proposal we put forward is in principle compatible with both of these generalizations. If threshold generalization turns out to be empirically correct, the function mapping $P(A|S)$ to the probability of pruning A from the set of alternatives of S may be a threshold function. If gradient generalization turns out to be empirically correct, the function mapping $P(A|S)$ to the probability of pruning A from the set of alternatives of S may be a linear increasing function. Initial data suggests that gradient generalization is on the right track (e.g., contrast between (12) and (13)). However, investigating empirically how exactly the judgments vary with the cardinality of the restrictor - number of disjuncts ratio is crucial for inferring properties of the function mapping $P(A|S)$ to the probability of pruning A from the set of alternatives of S.

Specifying this function, and answering interesting conceptual and empirical questions such a function would raise, will remain open for future work.

To summarize, the proposal in (51) accounts for the two aspects of the inference puzzle: the influence of cardinality of the restrictor and the number of disjuncts on distributive and ignorance inference derivation in quantified sentences with embedded disjunction. As a reminder, the proposal relies on the assumption that alternatives can be pruned under certain considerations (Horn, 1972; Fox and Katzir, 2011; Katzir, 2014; Crnič et al., 2015; Bar-Lev, 2018). The core of the proposal is that alternative pruning is sensitive to the informativeness of an alternative conditioned on the original utterance. The proposal is relatively independent of the specifics of the mechanism which derives implicatures: we have demonstrated how it can be implemented with the exhaustification-based framework for implicature derivation, but it is in principle also compatible with other approaches to implicature derivation.

10 Proposal and Deviance puzzle

Can the proposal put forward account for the deviance puzzle? In this section, we will discuss additional assumptions needed for this.

10.1 Obligatory (grammatical) implicatures

If problematic inferences are behind the deviance of sentences such as DEVIANT-BE and DEVIANT-WRITE, it must be the case that these inferences are obligatory. Within the exhaustification approach to implicatures, this would entail that sentences are obligatorily parsed with the exhaustifying operator *exh*. This assumption is arguably needed for any account aiming to explain deviance of certain sentences as a consequence of their problematic inferences. For instance, this assumption is already present in Magri (2009), who argued that sentences such as (30), repeated here, are deviant due to the problematic scalar implicatures (*Not all Italians come from a warm country*).

(30) Some Italians come from a warm country.

There is a related challenge for these accounts: even if the implicature computing mechanism is triggered whenever we interpret a sentence, why can't pruning the alternatives which would lead to problematic inferences save the sentence? In other words, why can't pruning the alternative *All Italians come from a warm country* from the set of altenatives of (30), or pruning the existential alternatives from the set of alternatives of DEVIANT-BE or DEVIANT-WRITE, save these sentences from deviance?

This is an important challenge to which we don't have a complete answer. There are two directions one could pursue.

One option would be to try to propose additional constraints on what type of alternatives can or cannot be pruned: for instance, Magri (2009) proposes one such constraint which prohibits pruning the alternative *All Italians come from a warm country* when we compute implicatures of (30).

Another possibility is that there is something about the architecture of implicature computation that disallows potential deviance of the sentence to influence pruning. In other words, pruning may be guided solely by relevance and informativeness considerations, and the information about whether the implicatures of a sentence contradict common knowledge might not be accessible for guiding the decision about which alternatives to prune.

Finally, throughout the paper, we have been working with the pragmatic account of ignorance inferences as in Fox (2007). However, the blind derivation of ignorance inferences suggests that they too need to be derived in grammar (cf. Meyer (2013, 2014); Buccola and Haida (2018)). According to these approaches, ignorance inferences, just like scalar implicatures, end up being part of the logical meaning of the sentence. We will remain agnostic as to what grammatical machinery is needed to ensure this (in principle, one could even stipulate that the derivation of ignorance inferences about all alternatives whose truth is not settled is encoded in the meaning of *exh*).

10.2 Probabilistic informativeness and blindness

If sentences such as DEVIANT-BE, DEVIANT-WRITE, or Magri's cases such as (30) are indeed deviant due to implicatures they trigger, there is another important consequence for the proposal that probabilistic informativeness considerations guide pruning.

This is that not only does the derivation of the ignorance inferences have to proceed in a blind manner, *but the mechanism which calculates the informativeness of alternatives must be blind to common knowledge too*. The reason is simply that, given common knowledge, P(All Italians come from a warm country | Some Italians come from a warm country) $= 1$: this means that the alternative 'All Italians come from a warm country' should be pruned due to its lack of informativeness from ALT(Some Italians come from a warm country), and that the problematic implicature should not arise.

Likewise, informativeness-based pruning that is blind to common knowledge is necessary to account for the deviance of DEVIANT-BE within our approach. To see why this is the case, recall that in order to derive ignorance inferences of DEVIANT-BE, the alternatives in (52) need to **not** be pruned from ALT(DEVIANT-BE).

(52) Someone is Mary,... (Susan, Jane), Someone is Mary or Susan,... (Susan or Jane, Mary or Jane)

However, given common knowledge P(Some of the girls is Mary| Each of the girls is Mary, Susan, or Jane) = 1, and similarly for all of the alternatives from (52).

This means that, if our proposal is on the right track and alternatives in (52) aren't pruned from the alternative set of DEVIANT-BE, the computation of informativeness according to the proposal in (51) has to be blind to (most of) common knowledge: the only things that seem to matter are domain size and logical words (quantifiers, disjunctions etc.) in the sentences. In other words, this means that $P(\text{Someone is } A \mid \text{Everyone is } A \text{ or } B \text{ or } C)$ is not influenced by the common knowledge about predicates A, B, C.

11 Empirical challenges

In this section, we discuss two types of challenges to the proposal presented in this paper.

The first challenge is that the proposal according to which probabilistic informativeness plays a role in pruning is in tension with current approaches to the so-called *symmetry problem*.

The second challenge consists in three additional observations of deviant sentences with embedded disjunction, whose deviance is not straightforwardly accounted for by problematic ignorance inferences. We discuss how the deviance of those three cases may be accounted for, but we acknowledge that further work on each of those three cases is needed.

11.1 The symmetry problem

According to our proposal, there are two sources of alternative pruning in implicature computation. The first is pruning due to contextual relevance considerations, whose existence has been argued for in much previous work on implicatures. The second is pruning due to informativeness considerations, which we have argued for in the present paper.

Recall from the discussion in Section 8 that previous work has established that pruning needs to be constrained — in other words, not all alternatives are prunable. A representative example is (33), repeated here, which can never be interpreted as 'John is French' or 'John is Spanish' (these interpretations would be available if it were possible to prune one of the alternatives 'John is French', 'John is Spanish' without the other).

(33) John is French or Spanish.

This data point belongs to a larger data pattern according to which, when a sentence has two alternatives which are *symmetric* (noting that different definitions of symmetry have been proposed, see Katzir, 2014 for details), it is not possible to prune one without pruning the other. In the case of (33), the alternatives 'John is French' and 'John is Spanish' are symmetric, and it is thus not possible to prune one without the other (i.e., to 'break' symmetry). This data pattern is one aspect of the so-called *symmetry problem*; see Fox and Katzir, 2011 and Breheny et al. (2018) for more comprehensive discussions of the symmetry problem.

Why can't pruning due to contextual relevance considerations break symmetry? To our knowledge, this question hasn't yet received a complete answer, although various proposals exist. For instance, Fox and Katzir, 2011 propose that this follows from a set of (natural) assumptions about relevance. That proposal has some empirical shortcomings however — see Katzir, 2014 for discussion of Fox and Katzir, 2011, and Breheny et al. (2018) for a recent review of empirical and conceptual challenges of all existing accounts of the symmetry problem.

Importantly, it appears that pruning based on informativeness considerations cannot break symmetry either. To see this, consider (53).

(53) All of Mary's 5 cousins or all of her 20 friends are French.

Formal alternatives of (53) are {All of Mary's 5 cousins are French, All of Mary's 20 friends are French, All of Mary's 5 cousins and all of her 20 friends are French}. As before, the disjunct alternatives 'All of Mary's 5 cousins are French' and 'All of Mary's 20 friends are French' are symmetric alternatives. If prior knowledge (dis)connecting Mary's family or friends to France doesn't enter into informativeness evaluation but the cardinality of individuals (domain size) does, P(All of Mary's 5 cousins are French$|$(53)) $>$ P(All of Mary's 20 friends are French$|$(53)). Our proposal may thus (in principle[17]) predict that we may be able to prune 'All of Mary's 5 cousins are French' due to informativeness considerations, and that (53) could thus have as implicature that not all of Mary's 20 friends are French. Again, this reading is intuitively unavailable. This suggests that we cannot prune one of the symmetric alternatives without pruning the other due to informativeness considerations.

How to account for this within the present proposal? A possible route would be to propose a set of constraints on informativeness-based pruning that would prevent it from breaking symmetry. This would however miss the generalization that symmetry cannot be broken by either of the two types of pruning (informativeness-based and contextual relevance considerations-based). We leave explaining this generalization as an important direction for future work.

11.2 Three additional observations

11.2.1 Modal contrast

Sentences such as DEVIANT-BE can be saved if the possibility modal is inserted below the universal quantifier and above disjunction, but not if the necessity modal is: (54a) is reported deviant, while (54b) is perfectly felicitous.

(54) a. #Each of these three girls must be Mary, Susan, or Jane.
b. Each of these three girls might be Mary, Susan, or Jane.

Can this contrast be explained by inferences triggered by (54a) which contradict common ground and which are not triggered by (54b)? Under the assumption that in (54a,b) the universal quantifier, the modal and the disjunction all activate alternatives, alternatives such as (55) will be IE for (54b), but not for (54a) (assuming (54a)/(54b) are parsed with *exh* taking the widest scope).

(55) Some of these three girls must be Mary (Susan, Jane...)

If one was to derive ignorance inferences about alternatives such as (55) when they are not IE, one could explain the contrast between (54a) and (54b) along similar lines as the contrast between DEVIANT-BE and NON-DEVIANT-CALLED. It is not clear however that postulating the existence of such ignorance inferences is desirable on independent grounds. In particular, alternatives such as (55) contain an epistemic modal *must*, and it is not clear what ignorance inferences about alternatives that contain epistemic modals should amount to.

11.2.2 Larger domain size

The intuitions about (56) appear to be more subtle than those for DEVIANT-BE or DEVIANT-WRITE, but at least some speakers find the sentence deviant.

(56) ?Each of the twenty girls in this photo is Lisa or one of our neighbors.

If the sentence in (56) triggered ignorance inferences, we could explain its deviance in the same way as we did for the sentences DEVIANT-BE and DEVIANT-WRITE. However, we have established

[17] Note however that the exact prediction depends on the function mapping informativeness to pruning, cf. discussion in Section 9.

that universally quantified sentences with the cardinality of restrictor and the number of disjuncts as in (56) are naturally interpreted with distributive and not ignorance inferences (cf. ALL-20-OR).

We would thus like to point to an alternative approach for the deviance of (56), which is nonetheless in the same spirit as the current proposal. Spector (2018) observes that sentences such as ALL-20-OR trigger not only distributive inferences according to which at least one of the twenty girls is French, and at least one is Spanish, but also an inference about how many of the twenty girls (approximately) are French, and how many are Spanish (we will refer to this in the continuation as the *distribution estimate inference*). The content of this inference for a sentence such as ALL-20-OR seems to be that there is approximately as many of the twenty girls who are French as those who are Spanish.

Such a distribution estimate inference in the case of (56) would amount to (57), which clearly contradicts common knowledge and could thus explain the deviance of (56).[18]

(59) Approximately the same number of the girls in the photo are Lisa as the number of girls in the photo who are our neighbors.

11.2.3 Downward-entailing contexts

Finally, we discuss sentences such as (60), in which the universally quantified sentence with a disjunction in its scope is embedded under a downward-entailing operator such as negation. Like for (56), the intuitions about (60) appear to be more subtle than those for DEVIANT-BE or DEVIANT-WRITE, but at least some speakers find the sentence deviant.

(60) ?It's not the case that both of these girls are Susan or Jane.

This empirical pattern is entirely parallel to deviance cases discussed by Magri (2009) which motivated the proposal that scalar implicatures are derived blindly to common knowledge. Consider (61), which is deviant just like (30), repeated here in (62).

(61) #It's not the case that some Italians come from a cold country.

(62) #Some Italians come from a warm country.

To explain the deviance of (61), Magri (2009) proposes that implicatures are in cases such as

[18] Extending the exhaustification approach to capture the *distribution estimate inference* is straightforward. The only necessary components are (i) to assume that sentences such as ALL-20-OR activate not the alternatives in which the universal quantifier is substituted with an existential, but the alternatives in which the universal quantifier is substituted with the full range of numeric expressions between (at least) 1 and the (at least) $n-1$, with n being the cardinality of the restrictor, and (ii) to assume that there is a threshold numeral such that all and only alternatives headed by numerals lower than the threshold numeral are not informative enough and are thus pruned. Taking as an example the numeral *at least 12* as the threshold numeral for ALL-20-OR, alternatives in (57) are sufficiently informative not to be pruned. They can all be negated consistently with ALL-20-OR, giving rise to inference in (58).

(57) At least 12 of Mary's friends are French, At least 12 of Mary's friends are Spanish,
...
At least 19 of Mary's friends are French, At least 19 of Mary's friends are Spanish,
All 20 of Mary's friends are French, All 20 of Mary's friends are Spanish

(58) At least 8 of Mary's friends are Spanish and at least 8 are French.

(61) derived *locally* instead of *globally*, that is to say, that implicatures are derived at the embedded level, below negation, rather than at the matrix level.

Furthermore, it is possible to construct deviant cases with the modified numeral *at least n* in downward-entailing contexts (recall that this modified numeral triggers ignorance inferences in upward-entailing contexts which may cause the sentence to be deviant, cf. (32)). Consider the slightly adapted scenario from Buccola and Haida (2018) in (63).

(63) Context: Ann played a card game in which, given the rules, the final score is always an even number of points. According to the rules, if a person scores 2 or 4 points, they get a small prize, and if they score 6 or more, they get a big prize. Carl believes that Ann scored at least 6 points; Bob however sees that Ann is awarded a small prize, and reports to Carl:

a. ?Ann got a small prize, so it can't be the case that she scored at least 5 points.
b. Ann got a small prize, so it can't be the case that she scored at least 6 points.

It thus seems to be the case that, quite generally, sentences which are deviant (arguably because they trigger certain problematic inferences) remain deviant when embedded under a downward-entailing operator. Whether this is because of a local derivation of problematic inferences or the deviance in such cases has a different source remains to be understood.

12 Three alternative directions for inference and deviance puzzles

We will now introduce three alternative directions one may attempt to pursue as competing accounts for the inference and deviance puzzles. We will point out the difficulties and open questions for each of these alternatives.

12.1 Alternative 1: Iterated rationality models

We demonstrated that the inference puzzle poses fundamental challenges to the exhaustification approach to implicatures, and more generally, to any approach in which implicatures are a function of (solely) entailment relations between a sentence and its alternatives — in addition to possibly considerations of contextual relevance. We proposed a solution to the inference puzzle according to which probabilistic informativeness plays a role in pruning.

There are however existing probabilistic approaches to implicatures (Franke et al., 2009; Franke, 2011; Goodman and Stuhlmüller, 2013; Franke and Jäger, 2014; Bergen et al., 2016). We will follow Fox and Katzir (2021) in referring to these models as iterated rationality models (IRMs).

Shortcomings of IRMs for certain types of implicatures, such as scalar implicatures and exhaustivity implicatures, have already been discussed (Franke and Bergen, 2020; Fox and Katzir, 2021; Cremers et al., 2022). IRMs may however still be appropriate models of various other inferences we draw when we interpret languages. Could they be used to model inference patterns of embedded disjunction?

An underlying assumption of IRMs models is that speakers and listeners are rational agents: the speaker reasons about how the listener will interpret the utterance, and the listener in turn reasons about how the speaker selects utterances, which results in inferences enriching the literal meaning of the speaker's sentences. In *most* IRMs, the inferences listeners draw depend heavily on prior world knowledge (common knowledge). The data pertaining to the deviance puzzle suggests however that inference patterns of embedded disjunction are largely independent of prior world knowledge. This in turn suggests that most existing IRMs wouldn't be appropriate to account for the deviance puzzle, even if an IRM account for the inference puzzle were to be developed.

There is however a version of IRMs developed in Degen et al. (2015), in which listeners reason about the prior world knowledge based on the speaker's utterance: if the utterance has certain properties, the listeners can suspend some of their prior beliefs in the process of implicature computation. It may be possible to develop a version of such a model which would account for the inference and the deviance puzzles: this would require working out a proposal for why prior world knowledge is systematically suspended in sentences with embedded disjunction.

12.2 Alternative 2: Disambiguation

There is yet another way to derive distributive inferences with recursive exhaustification[19], proposed by Bar-Lev and Fox (2016). It consists in applying recursively the exhaustifying operator at the matrix position, assuming that sentences with the disjunction embedded in the scope of a universal quantifier activate the alternatives as in *ALT-all-or*, i.e. the set of alternatives which includes the existential alternatives.

We have already established in Section 7 that if a sentence such as (64) is parsed as in (65) and it activates the alternatives as in *ALT-all-or*, when no alternatives are pruned, ignorance inferences are derived and not distributive inferences.

(64) Everyone is A or B.

(65) [*exh* [Everyone is A or B]]

However, Bar-Lev and Fox (2016) demonstrate that, if the sentence in (64) is parsed as (66), distributive inferences are derived.

(66) [*exh* [*exh* [Everyone is A or B]]]

Let us see why. The alternatives of the topmost *exh* are [20]:

(67) a. *exh* [Everyone is A] = Everyone is A and no one is B
b. *exh* [Everyone is B] = Everyone is B and no one is A
c. *exh* [Someone is A] = Someone is A and not everyone is A and no one is B
d. *exh* [Someone is B] = Someone is B and not everyone is B and no one is A

All of these alternatives are IE: negating them derives the distributive inferences in (68).

(68) Someone is A and someone is B.

In other words, the situation is the following: considering that sentences with the disjunction embedded in the scope of a universal quantifier activate the alternatives as in *ALT-all-or*, on the assumption that recursive exhaustification at the matrix level is possible, exhaustifying a sentence like (64) once derives ignorance inferences via the basic maxim of quantity, and exhaustifying it twice derives distributive inferences.

If this is indeed the way distributive and ignorance inferences of sentences such as (64) are

[19] Thanks to Moysh Bar-Lev for pointing this out to me.

[20] There are also the alternatives *exh*(Someone is A or B), *exh*(Someone is A and B), *exh*(Everyone is A and B); we are ignoring them for simplicity because they don't play a role in the derivation of distributive and ignorance inferences.

derived, one can put forward a proposal alternative to pruning to capture the empirical pattern of distributive and ignorance inferences, one that would possibly relate the informativeness of a sentence to a propensity to parse it with recursive matrix exhaustification. The idea in brief would be that, given that (69a) is more informative than (69b), and that (70a) is more informative than (70b), we are more likely to parse (69a) with recursive matrix exhaustification as compared to (69b), and likewise (70a) as compared to (70b).[21] In other words, the disambiguation process (between a parse with a single *exh* and parse with double *exh*) would have to be guided by the informativeness of a sentence.

(69) a. All of the 20 individuals are A or B.
b. Both individuals are A or B.

(70) a. All four individuals are A or B.
b. All four individuals are A, B, C, or D.

A problem with this approach however is the deviance puzzle. If the deviance pattern reported in Section 4 is indeed due to problematic inferences contradicting, one would need to propose that sentence disambiguation is not sensitive to whether one of the meanings is unlikely — or, in extreme case, contradictory — given prior world knowledge. This seems to be wrong: how we interpret the sentence 'I like banks' will likely differ when the sentence is uttered in a bank (financial institution) and when it's uttered at a bank (riverside).

12.3 Alternative 3: Implicature suspension

Our proposal according to which probabilistic informativeness plays a role in pruning is close in spirit to the proposal in Chemla and Romoli (2015), which was developed for other purposes. In their framework, implicatures of a sentence are eliminated if the informativeness of the implicature is too high. According to our proposal, alternatives are eliminated if the informativeness of the alternative given the original utterance is too low. The two ideas 'co-vary' in most cases, since in most cases the implicature is a consequence of the negation of an alternative.

Importantly, however, pruning an alternative from the whole process of implicature derivation (as in the current proposal) may have radically different effects than eliminating an implicature coming out of the presence of this alternative. To give a concrete example from the empirical domain explored in this paper, under the exhaustification approach to implicature derivation, pruning certain alternatives of sentences ALL-20-OR derives distributive inferences, and not pruning them derives ignorance inferences. Crucially, however, eliminating ignorance inferences (because they are too informative) would not immediately lead to the derivation of distributive inferences, or vice versa. This fact allows to differentiate our proposal from that of Chemla and Romoli (2015) on empirical grounds.

This is not to say however that the proposal in Chemla and Romoli (2015) cannot be extended to capture for the data discussed here. In particular, there are free parameters in Chemla and Romoli (2015) to be set up in terms of the set of alternatives assumed and the approach to implicature derivation taken in order to be able to fully compare it to the current proposal. We leave this comparison for future work.

[21] This is true under the same assumptions as elsewhere in the paper.

13 Conclusion

In this paper, two novel empirical puzzles with embedded disjunction have been explored: the inference and the deviance puzzle.

The inference puzzle taught us that quantified sentences with embedded disjunction trigger inferences which are sensitive in some way to the informativeness of the utterance and its alternatives (as evidenced by the effect of the domain size and the number of disjuncts on whether the ignorance or the distributive inferences are derived). The account we have put forward to capture this effect is that pruning of alternatives is sensitive to how much information the alternative carries over and above the original utterance: the more informative the alternative is, the more likely it is to be kept in the alternative set in the computation of implicatures. Importantly, even if the specifics of the pruning account turn out to be incorrect, the data pattern that the account aims to capture strongly suggests that informativeness other than logical or contextual entailment plays a role in some way in implicature computation.

The deviance puzzle is about a novel case of deviance of sentences with embedded disjunction, which we have argued to be caused by ignorance inferences. Importantly, if the proposed account is on the right track, ignorance inferences need to be derived blindly to common knowledge (much like scalar implicatures have been argued to be derived blindly to common knowledge by Magri, 2009), and crucially, the computation of informativeness of alternatives needs to be blind to (at least) some aspects of common knowledge (i.e., there has to be some level of modularity when informativeness is calculated). We consider these two conclusions — that probabilistic informativeness plays a role in implicature derivation, and that it is computed in a modular way — to be our main contributions.

Appendix: Recursive exhaustification and distributive inferences

There is an alternative way to derive distributive inferences in the exhaustification-based framework if one assumes that the exhaustifying operator can apply recursively, argued for in Crnič et al. (2015). They provide experimental results showing that sentences such as (71a) trigger the distributive inferences in (71b) *without necessarily triggering the inference in* (71c). This suggests that negating the disjunct alternatives, which is how distributive inferences are standardly derived, as discussed in Section 7, may not be the (only) way to derive distributive inferences.

(71) a. Every box contains an A or a B.
b. Some box contains an A and some box contains a B.
c. Not every box contains an A and not every box contains a B.

In order to derive the inferences in (71b) without deriving the inferences in (71c), Crnič et al. (2015) propose that the exhaustification operator applies twice in a sentence such as (71a). More specifically, they propose that the logical form of (71a) is (72).

(72) *exh* [Every box$_x$ *exh* [x contains an A or a B]]

Importantly, they assume that the conjunctive alternative ('x contains A and B'), which would have been the only IE alternative in the domain of the embedded *exh*, is pruned. Because of this, the embedded *exh* doesn't affect the meaning of (73a) — in other words, the meaning of (73a) is (the literal meaning of) (73b).

(73) a. [Every box$_x$ *exh* [x contains an A or a B]]
b. Every box contains an A or a B.

According to the parse (72), the alternatives on which the matrix *exh* operates are in (74):

(74) a. [Every box$_x$ *exh* [x contains an A]] = Every box contains an A and not a B
b. [Every box$_x$ *exh* [x contains an B]] = Every box contains a B and not an A
c. [Every box$_x$ *exh* [x contains an A and a B]] = Every box contains an A and a B

All of these alternatives can be negated consistently with (73a), i.e., with the original proposition which is an argument to the matrix *exh*. Negating (74a) obtains the inference in (75a), and negating (74b) the inference in (75b).

(75) a. It's not the case that every box contains an A and not a B.
$\rightsquigarrow$ Some box contains a B.
b. It's not the case that every box contains a B and not an A.
$\rightsquigarrow$ Some box contains an A.

Crucially, then, under the parse in (72), distributive inferences in (71b) are derived without the inferences in (71c).

Let us see what inferences are predicted by recursive exhaustification once we count in the inferences obtained by replacing the universal quantifier with an existential.

The alternatives activated by (71a), parsed with two exhaustification operators as in (72), are in (76):

(76) a. [Every box$_x$ *exh* [x contains an A]] = Every box contains an A and not a B
b. [Every box$_x$ *exh* [x contains an B]] = Every box contains a B and not an A
c. [Every box$_x$ *exh* [x contains an A and a B]] = Every box contains an A and a B
d. [Some box$_x$ *exh* [x contains an A]] = Some box contains an A and not a B
e. [Some box$_x$ *exh* [x contains an B]] = Some box contains a B and not an A
f. [Some box$_x$ *exh* [x contains an A or a B]] = Some box contains an A or a B
g. [Some box$_x$ *exh* [x contains an A and a B]] = Some box contains an A and a B

Now, which are the maximal sets of alternatives from (76) which can be negated consistently with (73a), the original proposition which is an argument to the matrix *exh*?

The following sets of alternatives are the maximal sets which can be negated consistently with (73a):

(77) a. (76a), (76b), (76d), (76e)
b. (76a), (76b), (76c), (76e)
c. (76a), (76b), (76c), (76d)
d. (76a), (76b), (76c), (76g)
e. (76b), (76c), (76e), (76g)
f. (76a), (76c), (76d), (76g)

These sets have an empty intersection. In other words, no alternative in (76) is IE. Under the same assumption as before about ignorance inferences (i.e. that they are derived about all of the alternatives whose truth is not settled by the utterance), we derive the ignorance inferences about all alternatives in (76) except (76f); the alternative in (76f) is entailed by (73a) under the assumption that the domain of boxes is non-empty.

Crucially, then, we have a very similar situation with recursive approach to distributive inferences as it was the case with the standard approach: having the existential alternatives derives ignorance inferences and doesn't derive distributive inferences; not having them derives distributive inferences and not ignorance inferences.

This means that the proposal in (51) can capture the cardinality of the restrictor effect under the recursive exhaustification approach too: intuitively, the reason is that the larger the domain size, the higher P((78a)|(73a)) (and P((78a) | (73a)) $>$ P((78b) | (73a)) whenever the domain size size is larger than 1). In other words, the larger the domain size, the more likely we will be to prune the alternatives such as (78a) and end up with the set of alternatives as in Table 6 rather than as in Table 5, deriving distributive rather than ignorance inferences.

(78) a. Some of the n boxes contains an A and not a B
b. All of the n boxes contain an A and not a B

Alternatives of (73a)	Inferences
[Every box$_x$ *exh* [x contains an A]]	**ignorance**
[Every box$_x$ *exh* [x contains an B]]	**ignorance**
[Every box$_x$ *exh* [x contains an A and a B]]	**ignorance**
[Some box$_x$ *exh* [x contains an A]]	**ignorance**
[Some box$_x$ *exh* [x contains an B]]	**ignorance**
[Some box$_x$ *exh* [x contains an A or a B]]	(entailed by (73a))
[Some box$_x$ *exh* [x contains an A and a B]]	**ignorance**

Table 5: Inferences of (7) — no alternatives pruned

Alternatives of (73a)	Inferences
[Every box$_x$ *exh* [x contains an A]]	Some box contains a B
[Every box$_x$ *exh* [x contains an B]]	Some box contains an A
[Every box$_x$ *exh* [x contains an A and a B]]	Not every box contains an A and a B
~~[Some box$_x$ *exh* [x contains an A]]~~	—
~~[Some box$_x$ *exh* [x contains an B]]~~	—
~~[Some box$_x$ *exh* [x contains an A or a B]]~~	(entailed by (73a))
[Some box$_x$ *exh* [x contains an A and a B]]	No box contains an A and a B

Table 6: Inferences of (6) — the least informative alternatives pruned

There is however one aspect of the proposal in Crnič et al. (2015) which is at odds with the account for the inference puzzle proposed here. In order to motivate the possibility of pruning the conjunctive alternative from the domain of embedded exhaustifying operator in (72), while avoiding optional pruning of conjunctive alternative in any sentence with disjunction (which would lead to an overgeneration problem), they propose the constraint on pruning discussed in Section 8: simplifying somewhat, they propose that one can only prune alternatives if that results in a weaker interpretation than not pruning them. This constraint is not met for existential alternatives of (72) as explained in Section 8: their constraint is thus not compatible with the account of distributive and ignorance inferences developed above. Resolving this tension is left for future work.

Word count: 15200

References

Bar-Lev, M. E. (2018). Free choice, homogeneity, and innocent inclusion. *Hebrew University of Jerusalem dissertation*.

Bar-Lev, M. E. and D. Fox (2016). On the global calculation of embedded implicatures. In *Poster presented at MIT workshop on exhaustivity*.

Bar-Lev, M. E. and D. Fox (2017). Universal free choice and innocent inclusion. SALT 2017.

Bergen, L., R. Levy, and N. Goodman (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics 9*.

Bergen, L., R. Levy, and N. D. Goodman (2014). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*.

Breheny, R., N. Klinedinst, J. Romoli, and Y. Sudo (2018). The symmetry problem: current theories and prospects. *Natural Language Semantics 26*(2), 85–110.

Buccola, B. and A. Haida (2018). Obligatory irrelevance and the computation of ignorance inferences. *Manuscript*.

Chemla, E. (2009). Similarity: Towards a unified account of scalar implicatures, free choice permission and presupposition projection. *Under revision for Semantics and Pragmatics*.

Chemla, E. and J. Romoli (2015). The role of probability in the accessibility of scalar inferences.

Chemla, E. and B. Spector (2011). Experimental evidence for embedded scalar implicatures. *Journal of semantics*, 359–400.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic inquiry 37*(4), 535–590.

Chierchia, G., D. Fox, and B. Spector (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. *Semantics: An International Handbook of Natural Language Meaning*.

Cremers, A., E. G. Wilcox, and B. Spector (2022). Exhaustivity and anti-exhaustivity in the rsa framework: Testing the effect of prior beliefs. *arXiv preprint arXiv:2202.07023*.

Crnič, L., E. Chemla, and D. Fox (2015). Scalar implicatures of embedded disjunction. *Natural Language Semantics 23*(4), 271–305.

Degen, J. and M. K. Tanenhaus (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive science 40*(1), 172–201.

Degen, J., M. H. Tessler, and N. D. Goodman (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. In *CogSci*.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, pp. 71–120. Springer.

Fox, D. and R. Katzir (2011). On the characterization of alternatives. *Natural language semantics 19*(1), 87–107.

Fox, D. and R. Katzir (2021). Notes on iterated rationality models of scalar implicatures. *Journal of Semantics 38*(4), 571–600.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics 4*, 1–82.

Franke, M. et al. (2009). *Signal to act: Game theory in pragmatics*. Institute for Logic, Language and Computation Amsterdam.

Franke, M. and L. Bergen (2020). Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language 96*(2), e77–e96.

Franke, M. and G. Jäger (2014). Pragmatic back-and-forth reasoning. In *Pragmatics, semantics and the case of scalar implicatures*, pp. 170–200. Springer.

Gazdar, G. (1979). Pragmatics, implicature, presuposition and logical form. *Academic Press*.

Goodman, N. D. and A. Stuhlmüller (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science 5*(1), 173–184.

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

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Ph. D. thesis, University of California, Los Angeles.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy 30*(6), 669–690.

Katzir, R. (2014). On the roles of markedness and contradiction in the use of alternatives. In *Pragmatics, semantics and the case of scalar implicatures*, pp. 40–71. Springer.

Klinedinst, N. W. (2007). Plurals, possibilities, and conjunctive disjunction. *UCL Working Papers in Linguistics 19*, 261–284.

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural language semantics 17*(3), 245–297.

Marty, P. and J. Romoli (2022). Presupposed free choice and the theory of scalar implicatures. *Linguistics and Philosophy 45*(1), 91–152.

Marty, P. P. (2017). *Implicatures in the DP domain*. Ph. D. thesis, Massachusetts Institute of Technology.

Matsumoto, Y. (1995). The conversational condition on horn scales. *Linguistics and philosophy*, 21–60.

Meyer, M.-C. (2013). *Ignorance and grammar*. Ph. D. thesis, Massachusetts Institute of Technology.

Meyer, M.-C. (2014). Deriving hurford's constraint. In *Semantics and Linguistic Theory*, Volume 24, pp. 577–596.

Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics 3*, 3–1.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and philosophy 27*(3), 367–391.

Schulz, K. and R. van Rooij (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy 29*(2), 205.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal 27*(3), 379–423.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques*. Ph. D. thesis, Paris 7.

Spector, B. (2007). Scalar implicatures: Exhaustivity and gricean reasoning. In A. B. M. Aloni and P. Dekker (Eds.), *Questions in dynamic semanticsy*, pp. 225–249. Elsevier.

Spector, B. (2018). What do we learn from game-theoretic pragmatics? SALT28 presentation.

Stalnaker, R. (1973). Presuppositions. *Journal of philosophical logic 2*, 447–457.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy 25*(5-6), 701–721.

Stalnaker, R. C. (1978). Assertion. In *Pragmatics*, pp. 315–332. Brill.

van Rooij, R. and K. Schulz (2004). Exhaustive interpretation of complex sentences. *Journal of logic, language and information 13*(4), 491–519.

von Fintel, K. and I. Heim (1997). Classnotes on pragmatics. *Massachusetts Institute of Technology (MIT)*.