Probabilistic informativeness in implicature computation: the case of embedded disjunctions*†

Milica Denić

Abstract

Sentences are standardly assumed to trigger scalar and ignorance implicatures because there are more informative alternative utterances the speaker could have said. Which notion of informativeness is at play here, and what knowledge is integrated into its evaluation? In this paper, a family of novel empirical observations is discussed related to inferences of embedded disjunctions, and used to argue (i) in favor of probabilistic rather than entailment-based informativeness being at play in implicature computation; and (ii) that the role of contextual knowledge in informativeness evaluation of alternative utterances 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 many 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.). They however generally agree on the following: (1a) triggers as its implicature the negation of (1b) because (1b) is an alternative to (1a) which is more informative than (1a). In this paper, we focus on two dimensions of informativeness along which these approaches may differ. The first is what notion of informativeness is relevant for implicature computation. Major candidates here are entailment-based informativeness (if Sentence 1 asymmetrically entails Sentence 2, then Sentence 1 is more informative), and probabilistic informativeness (if Sentence 1 expresses a proposition which is less likely to be true than Sentence 2, then Sentence 1 is more informative). The second is whether

*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, 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 has received 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 was supported by ANR-17-EURE-0017.

†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'.

contextual knowledge is integrated into informativeness evaluation, or whether informativeness evaluation relies instead on notions of logical entailments or probabilities.

In this paper, a family of novel empirical observations will be discussed, and used to argue in favor of probabilistic rather than entailment-based informativeness being at heart of implicature computation. One such observation is that inferences of embedded disjunction are sometimes 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.

(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.

We will further argue that not all contextual knowledge is incorporated in informativeness evaluation. This argument relates to a second family of observations, one of which is that (5a) and other structurally similar sentences 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.

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

In the first part of the paper, we focus on the contrast between (2) and (3) and related observations, and derive empirical generalizations that are to be explained. We introduce the exhaustification approach to implicature derivation (Fox, 2007; Chierchia et al., 2008), and discuss why it cannot at present capture the empirical generalizations. We choose to focus on exhaustification approach for concreteness: importantly, however, *neither the failure to account for the data nor the proposal that will be put forward is restricted to the exhaustification approach*. Any approach (exhaustification approach included) according to which (2) and (3) activate comparable alternatives and which relies on the notion of entailment-based informativeness will predict that (2) and (3) trigger the same implicatures, contrary to fact. Finally, we show how the contrast between (2) and (3) can be captured if implicature computation relies on the notion of probabilistic informativeness. We will formalize the proposal as a modification of the exhaustification approach to implicatures, but the insights may be incorporated into other approaches to implicatures as well.

In the second part of the paper, we discuss the contrast in (5). We will propose that the degraded status of (5a) is due to the inferences it triggers. That inferences a sentence triggers may cause it to be degraded has already been argued for by Magri (2009a,b). If this is correct, (probabilistic) informativeness of a sentence and its alternatives is computed without full access to contextual knowledge.

2 Empirical generalizations: Part 1

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.

(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.

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.

We acknowledge however that absolute preference for distributive inferences with ALL-20-OR and for ignorance inferences with ALL-2-OR may not always hold, because of contextual relevance considerations or other factors. 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 restrictor of the universal quantifier has a large cardinality than when it has a small cardinality; the naturalness of ignorance inferences is higher when the restrictor of the universal quantifier has a small cardinality than when it has a large cardinality.

What happens when the cardinality of the restrictor takes values intermediate to two and 20? As far as linguistic intuitions go, the effect appears to be ***gradient***: (8a) is still reported to be quite naturally interpreted with ignorance inferences, (8b) less clearly so, and (8c) is already reported to be naturally interpreted with distributive inferences.

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

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

b. All four of Mary's friends are French or Spanish.
c. All five of Mary's friends are French or Spanish.
...
d. All twenty of Mary's friends are French or Spanish.

The data in (8) thus suggests that the larger the cardinality of the restrictor, the more distributive inferences are derived (and conversely, the smaller the cardinality of the restrictor, the more ignorance inferences are derived). However, the empirical picture is more subtle. Consider (9), which will henceforth be referred to as SIMPLE-DISJ and (10), which will be henceforth referred to as COMPLEX-DISJ. The restrictor of the universal quantifier has the same cardinality in these two 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.

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

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

A more precise description of the data is thus that the inference pattern of universally quantified sentences with an embedded disjunction is sensitive to ***the ratio between the cardinality of the restrictor and the number of disjuncts***, rather than solely to the cardinality of the restrictor.

Finally, the phenomenon in question is not restricted to universally quantified sentences. For instance, consider (11), which will henceforth be referred to as 20-OR and (12), which will henceforth be referred to as 2-OR. 20-OR is reported to be more naturally interpreted with distributive inferences than 2-OR; 2-OR is reported to be more naturally interpreted with ignorance inferences than 20-OR.

(11) Twenty of Mary's friends are French or Spanish.

(12) Two of Mary's friends are French or Spanish.

Let us extract a generalization from the empirical observations so far: what is the difference between ALL-20-OR, SIMPLE-DISJ, 20-OR, and their counterparts ALL-2-OR, COMPLEX-DISJ, 2-OR, which might be responsible for the fact that the former are more naturally interpreted with distributive inferences than the latter, and the latter more naturally with ignorance inferences than the former? One salient difference between ALL-20-OR and ALL-2-OR is that, intuitively, ALL-20-OR carries more information than ALL-2-OR: ALL-20-OR assigns the property of being French or Spanish to 20 individuals, while ALL-2-OR does so to two. The same applies to 20-OR and 2-OR. Similarly, SIMPLE-DISJ carries more information than COMPLEX-DISJ, as SIMPLE-DISJ asymmetrically entails COMPLEX-DISJ. The empirical generalization that emerges from these considerations is thus in (13).

(13) ***Empirical generalization:*** The more informative the logical meaning of a quantified sentence with an embedded disjunction, the more likely distributive inferences are derived; the less informative it is, the more likely ignorance inferences are derived.

3 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 (14). (14) triggers ignorance inferences in (14a) (Gazdar 1979, a.o.). In addition to them, (14) also triggers the scalar implicature in (14b).

(14) John is French or Spanish.

a. The speaker is ignorant about whether John is French (Spanish)[2].
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 (15).

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

The semantics of *exh* is given in (16), 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 as many alternatives as possible from the set of alternatives activated by S, $ALT(S)$. There is however one 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. (16b)).

(16) 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) a sentence S activates are standardly assumed to be obtained by replacing the constituents of S with another expression of the same type 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. (17)).

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

Let us now see how the inferences of (14) are derived under this approach. We assume that $ALT((14))$ is in (18).

(18) Relevant formal alternatives of (14):

[2] 'The speaker is ignorant about whether John is French (Spanish)' and the like should be understood henceforth as abbreviation for 'The speaker is ignorant about whether John is French, The speaker is ignorant about whether John is Spanish'.

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

There are two maximal sets of alternatives of (14) which can be negated consistently with (14): (19a) and (19b).

(19) a. {John is French, John is French and Spanish}
b. {John is Spanish, John is French and Spanish}

The only IE alternative of (14) is thus 'John is French and Spanish', as it is the only alternative which appears in both (19a) and (19b). (14), parsed as (15), is thus interpreted as in (20).

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

How about the derivation of ignorance inferences? One approach[3] 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 he has. We will adopt the version of the maxim of quantity in (21), adapted from Fox (2007):

(21) *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 (14) follow from the maxim of quantity in (21). Assume that in a context in which the sentence (14) 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 (14) (both S' and S'' asymmetrically entail (14)), the maxim of quantity licenses the inferences in (23).

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

Assuming that the speaker believes his own utterance (14) (maxim of quality), the inferences in (22) amount to ignorance inferences in (23):

(23) 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'.

[3] Another approach to ignorance inferences within the exhaustification framework derives ignorance inferences as semantic inferences (Meyer, 2013, 2014; Buccola and Haida, 2018). In other words, ignorance inferences, just like scalar implicatures, end up being part of the logical meaning of the sentence.

4 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. Concretely, for ALL-20-OR and ALL-2-OR, this means that the alternatives are as in (24) and (25) respectively: we will henceforth refer to the set of alternatives in (24) and (25) as *ALT-all-or*.

(24) *ALT-all-or*(ALL-20-OR):
- a. All 20 are French
- b. All 20 are Spanish
- c. All 20 are French and Spanish
- d. Some are French
- e. Some are Spanish
- f. Some are French or Spanish
- g. Some are French and Spanish

(25) *ALT-all-or*(ALL-2-OR):
- a. Both are French
- b. Both are Spanish
- c. Both are French and Spanish
- d. Some are French
- e. Some are Spanish
- f. Some are French or Spanish
- g. Some are French and 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 (26) and (27) respectively.[4] We will henceforth refer to the set of alternatives in (26) and (27) as *ALT-or*.[5]

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

(27) *ALT-or*(ALL-2-OR):
- a. Both are French
- b. Both are Spanish
- c. Both are French and 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? 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 (28) can be negated consistently with ALL-20-OR:

(28) a. {All 20 are French, All 20 are Spanish, Some are French and Spanish, All 20 are French and Spanish}

[4]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.

[5]Fox (2007) and Magri (2009b) 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, Some are French and Spanish, All 20 are French and Spanish}
c. {All 20 are Spanish, Some are Spanish, Some are French and Spanish, All 20 are French and Spanish}

The only IE alternatives will thus be (24c) and (24g), as they are the only alternatives which appear in all three sets in (28). The predicted scalar implicatures will thus be their negations. 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 (21). 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 (24), i.e. they are all IE. This accounts for the distributive inferences: the negation of (26a) together with ALL-20-OR entails that some of Mary's friends are Spanish, and the negation of (26b) 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 for ALL-2-OR? We have seen that ALL-2-OR triggers ignorance inferences while ALL-20-OR does not (or at least to a very different degree). 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.

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. It is important to note again that the reason for this is fully general: what is at heart of the exhaustification approach, as well as many other approaches to implicature derivation, are the entailment relations between a sentence and its alternatives. 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.

5 Proposal: constraint on alternative pruning

In this section, we put forward a proposal for how to account for the data presented so far. The proposal crucially relies on changing the notion of informativeness involved in implicature computation from entailment-based to probabilistic. We will couch the proposal within the exhaustification approach to implicature derivation, but the central ideas are completely general, and could thus be incorporated into other approaches as well.

5.1 Informativeness and alternative pruning

Two assumptions will be central to our proposal. The first assumption is that some of the formal alternatives can sometimes be 'ignored' when implicatures are computed — the term that is usually employed for this is *alternative pruning* (Fox and Katzir, 2011; Katzir, 2014; Crnič et al., 2015). This assumption is in fact already tacitly incorporated in the definition of the set of alternatives of a sentence S, $ALT(S)$, as the set of formal alternatives which express contextually relevant propositions (cf. (17)): essentially, formal alternatives expressing contextually irrelevant propositions are thus pruned for $ALT(S)$, i.e. they are 'ignored' in the process of implicature computation.

The second assumption is that the set of formal alternatives of sentences such as ALL-20-OR and ALL-2-OR is the *ALT-all-or* alternative set, but that the alternative set *ALT-or* is derivable from *ALT-all-or* by pruning some of the alternatives from it.

The proposal we will put forward to account for the data presented so far 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 (29).

(29) **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.

(30) **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 larger *P(A)* (and thus the less informative A!), the more likely *A* is to be pruned from *ALT(S)*.

An intuitive reason for why (30) 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 his original utterance S.

Let us first see how this proposal accounts for the observation that, when keeping the number of disjuncts constant, the larger the cardinality of the restrictor of the universal quantifier scoping over a disjunction, the more distributive and the less ignorance inferences are derived (cf. ALL-20-OR and ALL-2-OR). In particular, for some domain size n, for a sentence of the form (31), let us consider what happens with its alternatives of the form (31a,b).

(31) 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 can be shown to follow from more general simplifying assumptions in (32). The simplifying assumptions in (32) plausibly hold when the speaker possesses little contextual knowledge apart from that relating to cardinalities of various sets: the speaker knows the number of individuals in the domain but not which predicate(s) hold of them; in addition to this, the speaker knows that all predicates hold of some but not all individuals, and they may know of how many individuals each predicate holds, but they do not know whether there are relations between indi-

viduals and predicates. Obviously, these assumptions are most often not met in the actual world. We will however adopt them for the time being, and see evidence at a later point in the paper that some aspects of implicature computation proceed as if assumptions such as these are met (i.e. as if little contextual knowledge is involved).

(32) **Simplifying assumptions:**
Let there be some set of predicates *Pred*. For any domain D, for any predicate A from *Pred*, for any individuals x, y from D, $P(x \in A) = P(y \in A)$ (*equality assumption*), with $P(x \in A) \in (0, 1)$ (*contingency assumption*). For any domain D, for any two predicates A_k and A_l from *Pred* and any two individuals x_i and x_j from D, such that $A_k \neq A_l$ and $x_i \neq x_j$, $P(x_i \in A_k)$, $P(x_j \in A_k)$, $P(x_i \in A_l)$, $P(x_j \in A_l)$ are all independent (*independence assumption*).

Under assumptions in (32), the alternative such as 'Some of the n individuals are A' is more likely to be true for larger n, 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 (cf. Section 4). 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 4). 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.

Alternatives of ALL-20-OR	Inferences
All 20 are French	**Some are Spanish**
All 20 are Spanish	**Some are French**
All 20 are French and Spanish	Not all are French and Spanish
~~**Some are French**~~	—
~~**Some are Spanish**~~	—
~~Some are French or Spanish~~	(entailed by ALL-2-OR)
Some are French and Spanish	No one is French and Spanish
All 20 are French and Spanish	Not all are French and Spanish

Table 1: Left: Alternatives of ALL-20-OR with the least informative alternatives pruned (in strike-through 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**
Both are French and Spanish	Not both are French and Spanish
Some are French	**ignorance**
Some are Spanish	**ignorance**
Some are French or Spanish	(entailed by ALL-2-OR)
Some are French and Spanish	No one is French and Spanish
Both are French and Spanish	Not both are French and Spanish

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.

The proposal in (30) (under simplifying assumptions as in (32)) can thus capture that the smaller the cardinality of the restrictor of the universally quantified sentence with an embedded disjunction, the more ignorance inferences and the less distributive inferences are derived.

As it stands, however, the proposal in (30) does not capture the entire empirical pattern from Section 2. Firstly, the proposal does not yet capture that this effect is not restricted to universally quantified sentences. As a reminder, consider (11) (i.e. 20-OR) and (12) (i.e. 2-OR), repeated here: 20-OR is more naturally interpreted with distributive inferences than 2-OR, and 2-OR more naturally with ignorance inferences than 20-OR.

(11) Twenty of Mary's friends are French or Spanish.

(12) Two of Mary's friends are French or Spanish.

Assuming that, prior to any pruning, 20-OR and 2-OR activate the alternatives in (33) and (34), the proposal in (30) does not capture the contrast between 20-OR and 2-OR. The reason is that we have so far granted that *P(Some of Mary's friends are French)* depends on the total number of Mary's friends, but we made no connections to the information conveyed by 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 which numeral heads the sentence.

(33) Alternatives of 20-OR

a. Twenty are French
b. Twenty are Spanish
c. Some are French
d. Some are Spanish

(34) Alternatives of 2-OR

a. Two are French
b. Two are Spanish
c. Some are French
d. Some are Spanish

Secondly, the proposal does not yet capture that the inference pattern is sensitive to the number of disjuncts. As a reminder, consider (9) (i.e. SIMPLE-DISJ) and (10) (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.

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

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

Let us see why the proposal in (30) 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 alternatives of the form 'Some are French', 'Some are French or Spanish' etc. are pruned, as in Table 4.[6]

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 least informative 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 (30) 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 (35), solves these two problems.

(35) **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

[6]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'.

given that *S* is true. The larger $P(A \mid S)$ (and thus the less informative *A* given *S*!), the more likely we are to prune *A* from *ALT(S)*.

An intuitive reason for why a constraint on pruning such as (35) might hold is related to what has been said to conceptually motivate (30): 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 he believes both A and S to be true.

We have already seen that, under the simplifying assumptions in (32), as n increases, so does the $P(\text{Someone is A} \mid n \text{ people are A or B})$. The proposal thus now captures straightforwardly both the original contrast between ALL-20-OR and ALL-2-OR, as well as the contrast between 20-OR and 2-OR. 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*)*, and from *ALT(*20-OR*)* than from *ALT(*2-OR*)*. This will result in distributive inferences being more likely to be derived in the case of ALL-20-OR and 20-OR than in the case of ALL-2-OR and 2-OR, and ignorance inferences being more likely to be derived in the case of ALL-2-OR and 2-OR than in the case of ALL-20-OR and 20-OR.

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

We submit that it is reasonable to assume 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 $m < n$. This can be shown to follow from the simplifying assumptions as in (32).

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 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 restricted set (i.e. without the alternatives such as 'Some of the n individuals are A') for 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.

6 Interim discussion

To summarize, the proposal in (35) accounts for the two aspects influencing the derivation of ignorance inferences of the disjunction: the cardinality of the restrictor and the number of disjuncts in quantified sentences with embedded disjunction. The proposal also extends straightforwardly to sentences headed with quantifiers other than universal (cf. 20-OR and 2-OR). 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). 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 neo-Gricean approaches (e.g. Sauerland (2004)).

We finish the interim discussion by drawing comparisons to related work. First, the current proposal 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.

Second, our proposal is conceptually related to rational speech act (RSA) models (Bergen et al., 2016; Goodman and Stuhlmüller, 2013, a.o.). Simplifying somewhat, in RSA models, the amount of information carried by alternative utterances contributes positively to their utility (modulo the cost of the utterance) in a situation in which they can be used truthfully; the higher the utility of an utterance, the more likely the speaker is to use the utterance, and thus the more likely the pragmatic listener is to infer that the utterance cannot be uttered truthfully if the speaker doesn't use it. It should thus be possible to develop an RSA model which can capture the observations made thus far, although we will not make an attempt at that in the continuation. There is an important caveat, however: an underlying assumption of RSA models is that speakers and listeners are rational agents, and that inferences listeners draw depend heavily on contextual knowledge. As we will see in the continuation however, not all contextual knowledge seems to be integrated in implicature computation.

7 Informativeness and contextual knowledge

In previous section, we have demonstrated that under certain simplifying assumptions about probability assignments (cf. (32)), informativeness-based pruning can capture the empirical pattern discussed in Section 2. As discussed previously, these assumptions hold when the speaker possesses little contextual knowledge about individuals in the domain and predicates, apart from their cardinalities. What consequences does contextual knowledge, which clearly often doesn't validate these simplifying assumptions, have on probability assignments to alternatives, and thus on inferences?

Strikingly, the contextual probabilities do not have the influence one might expect them to. To see this, consider (36).

(36) Both of Mary's office-mates are from the US or the UK.

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). One thus expects that the alternative 'Some of Mary's office-mates are from the US' will be less informative, and thus more likely to be pruned that the alternative 'Some of Mary's office-mates are from the UK'. If the alternative 'Some of Mary's office-mates are from the US' is pruned but not the alternative 'Some of Mary's office-mates are from the UK', (36) should have the inference that not both of them are from the UK (and so at least one is from the US), and the ignorance inference about whether both of Mary's office-mates are from the US.

This interpretation does not seem to be readily available: it thus seems that asymmetries in informativeness of alternatives that are due to contextual knowledge do not translate straightforwardly into an effect on the inferences triggered by sentences such as (36).

Building on (36) and a series of observations that will be discussed in the continuation of the paper, we will propose that informativeness computation does not have complete access to contextual knowledge. What is then the source of probabilities assigned to alternatives? We would like to propose that, in addition to being able to assign probabilities to various propositions based on contextual knowledge, humans also possess a cognitive mechanism for the computation of logical probabilities, very much in the spirit of Carnap (1950), and that this second computation has important repercussions for implicature computation. Essentially, such a computation would be guided by formal properties of the sentence, and not by the meaning of various predicates or contextual knowledge about them. An interesting direction for future research in the domain of psychology of reasoning could be to examine the relevance of logical probabilities in human probabilistic inference outside of implicature computatiton.

In the remainder of the paper, we turn to the observation (5) from Section 1. We will propose that the deviance of this and related cases is due to ignorance inferences such sentences give rise to. If that explanation is on the right track, it will be absolutely crucial that probabilities assigned to alternatives do not derive from contextual knowledge, as announced above.

8 Empirical generalizations: Part 2

When disjunctions of definite noun phrases are 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 (37), which will be referred to as DEVIANT-BE, (38), which will be referred to as NON-DEVIANT-CALLED, (39), which will be referred to as DEVIANT-WRITE, and (40), 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.

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

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

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

(40) (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 (41), assuming that it is common knowledge that Mary, Susan, and Jane have to be three different individuals (see (43) for definitions of common knowledge, context set and contextual equivalence). Likewise, DEVIANT-WRITE is contextually equivalent to (42), assuming that it is common knowledge that for any book there can be exactly one singular or plural individual who wrote it[7]. Yet, surprisingly, DEVIANT-BE cannot be naturally used to do convey the meaning of (41), and neither can DEVIANT-WRITE to convey the meaning of (42).

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

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

(43) **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*. The deviance is also not limited to universally quantified sentences; see for instance (44). For the simplicity of exposition, we will focus on the disjunction embedded in the scope of *each*; the main ideas that will be presented carry over to the two other universal quantifiers and to cases in (44).

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

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 world knowledge. More formally, given world knowledge, they have the property in (45), which we will call *left-uniqueness*.

(45) 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

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

argument is from a domain of books), observe that the continuations in (46a) and (46c) sound contradictory, but not in (46b) and in (46d).

(46) 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 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 (47), 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):

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

9 Proposal: blind ignorance inferences

Our proposal for the deviance of DEVIANT-BE and DEVIANT-WRITE will build on the conclusions reached in the first part of the paper. We have seen in the first part of the paper 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 (48) and (49), respectively.

(48) The speaker is ignorant about whether each of these three girls is Mary... (Susan, Jane), and he is ignorant about whether some of these three girls is Mary... (Susan, Jane).

(49) The speaker is ignorant about whether each of these three writers wrote Anna Karenina, ... (Germinal, Harry Potter), and he 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. Because of the left-uniqueness of the identity copula and the predicate *to write*, the sentences like those in (50) contradict general world knowledge. ***The speaker thus cannot possibly be ignorant about them*** — he must know that they are false.

(50) Each of these three girls is Mary, Each of these three writers wrote Anna Karenina

Furthermore, if we take it to be common knowledge that the speaker believes their own utterances in DEVIANT-BE/DEVIANT-WRITE, again due to the left-uniqueness of the identity copula and the predicate *to write*, ***the speaker cannot be ignorant about the sentences in (51)*** — he must know that they are true.

(51) Some of these three girls is Mary, Some of these three writers wrote Anna Karenina

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. (52)).

(52) **Deviance due to ignorance inferences:**
If a sentence S triggers *ignorance inferences* which contradict common knoweldge, S is deviant.

There is in fact already some empirical data suggesting that ignorance inferences may result in the deviance of sentences which triggered them when they contradict common knowledge, in line with (52). These data relate to the ignorance inferences of the modified numeral *at least*. We provide in (53) an example from Buccola and Haida (2018); similar empirical observations have been first made by Nouwen (2010).

(53) 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.

While we will not go into the the question of how ignorance inferences are derived for a sentence such as (53a) (see for instance Spector (2019) for a recent review article), a possible reason why (53a) is deviant is because it triggers the ignorance inference that the speaker is ignorant about whether Ann scored exactly 3 points, which contradicts common knowledge.

The core assumption of the proposal in (52) is that implicatures generally, and ignorance inferences specifically, are derived *blindly* from common knowledge. The idea that the procedure which derives implicatures is blind to contextual knowledge has been in fact already defended by Magri 2009a for the case of scalar implicatures (see also Degen et al. (2015) for experimental data on unexpected patterns of implicature derivation given contextual knowledge). The pragmatically deviant (54) is an example of the cases that motivated this proposal.

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

Simplifying a lot, according to Magri's proposal, (54) is deviant because the conjunction of (54) and its scalar implicature in (55) contradicts common knowledge.

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

There is another important point to be made here. If Magri's proposal according to which (54) is deviant because of its scalar implicature is correct, and our informativeness-based pruning proposal is on the right track, 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 contextual knowledge too**. The reason is simply that, given contextual 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. This fact reiterates the observations made in Section 7, according to which contextual probabilities do not seem to influence the inferential pattern of embedded disjunctions in the ways in which one might expect them to.

Likewise, informativeness-based pruning that is blind to contextual 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 (56) need to **not** be pruned from ALT(DEVIANT-BE).

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

However, given common world 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 (56).

This means that, if our proposal is on the right track and alternatives in (56) aren't pruned from the alternative set of DEVIANT-BE, the computation of informativeness according to the proposal in (35) 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(Someone is *A* | Everyone is *A* or *B* or *C*) is not influenced by the common knowledge about predicates A, B, C.

10 Empirical challenges

In this section, we discuss three additional data points which are not as of yet straightforwardly captured by the current proposal, despite being evidently related to the observations presented in Section 8. We discuss in what way they are challenging for the proposal and how they may be in some cases accommodated under it, but we acknowledge that further work on each of these three interesting cases would be welcome.

10.1 Modal contrast

There is a similar contrast between the necessity and possibility modal as between predicates *to write* and *to read* (cf. Section 8): (57a) is reported deviant, while (57b) is perfectly felicitous.

(57) 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 (57a) which contradict common ground and which are not triggered by (57b)? Under the assumption that in (57a,b) the universal quantifier, the modal and the disjunction all activate alternatives, alternatives such as (58) will be IE for (57b), but not for (57a) (assuming (57a)/(57b) are parsed with *exh* taking the widest scope).

(58) Each of these three girls must be Mary (Susan, Jane...)

If one was to derive ignorance inferences about alternatives such as (58) when they are not IE, one could explain the contrast between (57a) and (57b) along similar lines as the contrast between, for instance DEVIANT-BE and NON-DEVIANT-CALLED. It is not clear however that postulating the existence of such ignorance inferences is desirable on independent grounds. First, alternatives such as (58) contain an epistemic modal *must*, and it is not clear what ignorance inferences about alternatives that contain epistemic modals should amount to. Second, more empirical work is needed to establish whether (59a) indeed triggers different ignorance inferences from (59b). We leave this exploration for future work.

(59) a. Each of these three girls must be from France, Germany, or Spain.
b. Each of these three girls might be from France, Germany, or Spain.

10.2 Larger domain examples

Consider the deviant sentence in (60).

(60) #Each of the twenty girls in this photo is Mary's daughter Lisa or one of her classmates.

If the sentence in (60) 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 that universally quantified sentences with the cardinality of restrictor and the number of disjuncts as in (60) 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 (60), 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 (60) would amount to (61), which clearly contradicts common knowledge and could thus explain the deviance of (60).[8]

(63) Approximately the same number of the girls in the photo are Mary's daughter Lisa as the number of girls in the photo who are her classmates.

10.3 Downward-entailing contexts

Finally, we discuss sentences such as (64), in which the universally quantified sentence with a disjunction in its scope is embedded under a downward-entailing operator such as negation. The intuitions about (64) appear to be less stable than those for DEVIANT-BE or DEVIANT-WRITE, but at least some speakers find the sentence deviant.

[8]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 (61) are sufficiently informative not to be pruned. They can all be negated consistently with ALL-20-OR, giving rise to inference in (62).

(61) 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

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

(64) ?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 (2009a) which motivated the proposal that scalar implicatures are derived blindly to common knowledge. Consider (65), which is deviant just like (54), repeated here in (66).

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

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

To explain the deviance of (65), Magri (2009a) proposes that implicatures are in the cases such as (65) derived *locally* instead of *globally*, that is to say, that the 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. (53)). Consider the slightly adapted scenario from Buccola and Haida (2018) in (67).

(67) 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.

11 Concluding remarks

In this paper, two novel empirical puzzles have been explored.

The first puzzle taught us that quantified sentences with embedded disjunction trigger ignorance 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 second puzzle explores a novel case of deviance, 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 contextual knowledge (much like scalar implicatures have been argued to be derived blindly to contextual knowledge by Magri, 2009b), and crucially, the computation of informativeness of alternatives needs to be blind to (at least) some aspects of contextual

knowledge (i.e. there has to be some level of modularity when informativeness is calculated). We consider these two conclusions — that probabilistic informativeness might be at heart of implicature derivation, and that it is computed in a modular way — to be our main contributions.

Zooming into the features of the pruning proposal, we have proposed to distinguish between relevance-based pruning and informativeness-based pruning, assuming that they play complementing roles in restricting the formal alternative set for implicature derivation: the set of formal alternatives that is used in the implicature derivation is the set of formal alternatives which are both contextually relevant and sufficiently informative. Importantly, however, relevance-based pruning is essentially connected to contextual knowledge, while we have argued that the informativeness-based pruning is blind to at least some aspects of it. How to accommodate the co-existence of these two types of pruning? One possibility is that there are in fact two routes to implicatures: one domain-general (with contextual knowledge integration), and the other modular (without contextual knowledge integration). Relevance-based pruning might thus be a component of the first route only, and informativeness-based pruning a component of both.

A straightforward objection to the two routes idea is why the cases of deviance due to blind implicatures ever occur to begin with, instead of such sentences being 'saved' by the domain-general route. A possible reason why is that modular calculation of implicatures might be less costly in terms of cognitive resources needed to integrate contextual knowledge, and is thus a default one: there may be certain contextual requirements to be met for someone to engage in the domain-general calculation of implicatures.

Clearly, arguing in favor or against the two routes idea requires much further research.

Word count: 12049

References

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*.

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

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press.

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*.

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

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.

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

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

Magri, G. (2009b). *A theory of individual-level predicates based on blind mandatory scalar implicatures: constraint promotion for Optimality Theory*. Ph. D. thesis, Ph. D. dissertation, MIT.

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.

Spector, B. (2019). Modified numerals. In R. Z. Matthewson, Meier (Ed.), *Wiley Companion to Semantics*.

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.