Linguistics and Philosophy manuscript No.
(will be inserted by the editor)

An Implicature account of Homogeneity and Non-maximality

Moshe E. Bar-Lev

Received: date / Accepted: date

Abstract I provide arguments in favor of an implicature approach to Homogeneity (Magri 2014) where the basic meaning of *the kids laughed* is *some of the kids laughed*, and its strengthened meaning is *all of the kids laughed*. The arguments come from asymmetries between positive and negatives sentences containing definite plurals with respect to (i) children's behavior (Tieu et al. 2015), (ii) the availability of non-maximal readings, and (iii) the robustness of neither-true-nor-false ('gappy') judgments (Križ and Chemla 2015). I propose to avoid some problems of Magri's analysis by modelling the Implicature account of Homogeneity after the Implicature account of Free Choice, based on a hitherto unnoticed analogy between the two phenomena. The approach that emerges has the advantages of Magri's Implicature account of Homogeneity (predicting asymmetries), while at the same time bears a close resemblance to recent approaches to Non-maximality (Malamud 2012; Križ and Spector 2017), which enables restating their account of Non-maximality as following from the context-sensitivity of implicature calculation.

Keywords Homogeneity · Non-maximality · Exhaustivity · Innocent Inclusion · Context dependency

Moshe E. Bar-Lev
The Hebrew University of Jerusalem
E-mail: mosheel.levin@mail.huji.ac.il

1 Introduction

1.1 Homogeneity

(1) exemplifies the issue of Homogeneity with definite plurals:[1] In out-of-the-blue contexts we infer from the sentence in (1a) that all the kids laughed and from the sentence in (1b) that none of the kids laughed (Löbner 1987, 2000; Schwarzschild 1994; Krifka 1996; Gajewski 2005; Magri 2014; Križ 2015, 2016; Križ and Spector 2017).

(1) a. The kids laughed.
   (i) ≈ *All* the kids laughed.
   (ii) ≉ *Some* of the kids laughed.
  b. The kids didn't laugh.
   (i) ≉ *Not all* the kids laughed.
   (ii) ≈ *None* of the kids laughed.

While it is entirely possible that the reading reported for (1b) can be derived by scoping negation below whatever it is that provides the universal quantification over the kids in (1a), it is still surprising that (at least with no special intonation) there is no reading for (1b) where it means *not all the kids laughed*, which is what we should expect if negation takes scope above that putative universal quantification. Moreover, even when the definite plural cannot possibly take scope above negation, for example when it contains a bound variable as in (2) which prevents the definite plural from taking wide scope, the only possible reading is one that can be paraphrased as negation taking scope above an existential quantifier (Križ 2015).[2]

(2) No boy found his presents.
  a. ≉ No boy found *all* of his presents.
  b. ≈ No boy found *any* of his presents.

Before we discuss why Homogeneity is puzzling and various ways to tackle this puzzle, we should pay attention to another property of sentences containing definite plurals called Non-maximality.

1.2 Non-maximality

The Homogeneity data reported in (1) are complicated by another property of sentences containing definite plurals, called Non-maximality, which is that they allow

[1] A note about the terminology: the term Homogeneity is sometimes used in the literature to refer to the phenomena described in (1), and sometimes to refer to the claim that the sentences in (1) are judged neither-true-nor-false in a situation where some but not all of the kids laughed. I will only use the term Homogeneity for referring to the former, i.e., the phenomena in (1); the latter phenomenon will be termed Gappiness. I will discuss the relation between the two phenomena in §6.

[2] Of course, this does not mean that (1b) is not structurally ambiguous between negation taking scope above and below the definite plural; it does however lead to the conclusion that both of those structures miraculously end up having the same meaning in out-of-the-blue contexts.

for exceptions. In contexts where it doesn't matter whether most or all of the kids laughed, as in (3a), we judge the sentence *the kids laughed* as true even if only most of the kids laughed (see Dowty 1987; Brisson 1998, 2003; Lasersohn 1999; Malamud 2012; Schwarz 2013; Križ 2015, 2016; Križ and Spector 2017).

(3) a. *Context: There was a clown at my kid's birthday party. Someone asks me if they gave a funny performance. I reply:*
b. The kids laughed.

As has been emphasized in the literature, whether a non-maximal interpretation is available is highly dependent on the context. Given the context in (4a), *the kids laughed* only receives a maximal interpretation.

(4) a. *Context: Someone asks me who among the five adults and the ten kids at the birthday party laughed. I reply:*
b. The kids laughed.

Križ (2015) convincingly argues that there is a non-accidental connection between Homogeneity and Non-maximality: Non-maximal readings are only possible for expressions that give rise to Homogeneity. When Homogeneity disappears, so does Non-maximality. Following Križ, I take it that a complete account of Homogeneity should offer an explanation for this connection. For expository reasons I will however dedicate the remainder of this introduction to the various approaches to Homogeneity, pretending that only maximal readings exist.

1.3 Approaches to Homogeneity and asymmetry

Here is one way to illustrate the problem of Homogeneity with definite plurals. Suppose Kelly is one of the kids. Then, based on the judgment in (1a), one might suggest that (5a) holds. And based on the judgment in (1b), one might suggest that (5b) holds. Given the rule of contraposition and these two assumptions, we conclude (5c), which is obviously nonsensical if taken at face value: *Kelly laughed* does not intuitively entail that *the kids laughed*.

(5) **The puzzle of Homogeneity**:
a. The kids laughed $\Rightarrow$ Kelly laughed
b. $\neg$ The kids laughed $\Rightarrow$ $\neg$ Kelly laughed
c. $\therefore$ The kids laughed $\Leftrightarrow$ Kelly laughed

The main solutions that have been proposed to this problem can be divided into two groups: the Trivalence approach and the Ambiguity approach. The Trivalence approach holds that *the kids laughed* is only defined if none or all of the kids laughed (Schwarzschild 1994; Löbner 2000; Gajewski 2005; Križ 2015, 2016, a.o.). This view solves the problem by accepting the inference in (5), under the assumption that entailment is understood as *Strawson*-entailment (von Fintel 1999), i.e., replacing $\Rightarrow$ with $\Rightarrow_{st}$.[3] Indeed, if *Kelly laughed* is true and *the kids laughed* is defined, which

[3] $A \Rightarrow_{st} B$ *iff* B is true whenever A is true and B is defined.

within the Trivalence approach means that none or all of the kids laughed, then all the kids laughed. So *Kelly laughed* Strawson-entails *the kids laughed*. For the Ambiguity approach (Krifka 1996; Spector 2013; Magri 2014; Križ and Spector 2017; Spector 2018), *the kids laughed* is ambiguous between an existential and a universal reading, and possibly many intermediate readings as well; this assumption is supplemented with a mechanism that decides which of the readings is chosen for which sentence, e.g., Dalrymple et al.'s (1994) Strongest Meaning Hypothesis. This solves the problem by denying the identity of the two instances of *the kids laughed* in (5a) and (5b), hence invalidating the inference in (5).

In this paper I pursue a third approach, the Implicature approach proposed by Magri (2014), in which the burden of explanation shifts to the theory of implicature calculation.[4, 5] Within the Implicature approach, the basic meaning of *the kids laughed* is the same as *some of the kids laughed*, and this meaning is strengthened in positive sentences into *all the kids laughed*, which I will refer to as the Maximality implicature. The solution proposed by this view amounts to denying (5a): the basic meaning of *the kids laughed* does not entail that *Kelly laughed*; this inference is rather the result of calculating an implicature.

I will argue in this paper in favor of the Implicature account, capitalizing on a distinguishing feature of this account which I will call Asymmetry. In both the Ambiguity and Trivalence approaches the reasoning behind the inference pattern of the positive sentence in (1a) and that of the negative sentence in (1b) is the same. In other words, they are symmetric in nature. This is in contrast with the Implicature account, in which the Maximality inference in the positive sentence in (1a) does not follow from the basic semantics (i.e., the meaning before implicature computation), (6a), but from implicature calculation, whereas the Maximality inference in the negative sentence in (1b) follows from the basic semantics, (6b). Taken together with the independently established generalization that implicatures are not derived under negation (except under special circumstances, see Fox and Spector 2018), the Implicature approach is asymmetric in nature.[6]

[4] Since Magri's Implicature account (as well as what I will propose) is couched within a view where implicatures are a special case of ambiguity (i.e., an ambiguity between the basic meaning and the strengthened meaning), the Implicature account is in fact a special case of the Ambiguity account. I choose to distinguish between them since as we will shortly see the Implicature account makes substantially different predictions than the other Ambiguity accounts mentioned above.

[5] Malamud's (2012) account of Non-maximality (in the footsteps of which we'll follow in §5) is based on the idea that the interpretation of definite plurals should be treated in the same way implicatures are. Since accounting for Homogeneity isn't her main goal, and since as Križ and Spector (2017) observed her account fails to derive the correct Homogeneity facts, I do not consider her account to be an Implicature account of Homogeneity.

[6] Of course, one may add auxiliary assumptions to the Ambiguity and Trivalence accounts in a way that would make them asymmetric. For example, within the Ambiguity account one could suggest that for a sentence containing a definite plural to be true the truth of both the universal and the existential readings is required in principle (as in Spector 2013; Križ and Spector 2017), but they are not on equal footing: we might be forgiving (i.e., allow for Non-maximality) in cases where the universal reading is false but not in cases where the existential reading is. However, such a move would be stipulative; If asymmetry is borne out, as I will argue, it is obviously more parsimonious and thus preferable to have an account in which it follows from the core assumptions. For this reason I will maintain the division between symmetric and asymmetric accounts. As Benjamin Spector pointed out to me, the Rational Speech Act implementation of Križ and Spector (2017) found in Spector (2018) predicts a slight asymmetry between positive and

(6) **Asymmetry within the Implicature account** (given basic meaning):

a. The kids laughed $\not\Rightarrow$ Kelly laughed

b. $\neg$ The kids laughed $\Rightarrow$ $\neg$ Kelly laughed

I will first present in §2 evidence supporting the asymmetric architecture of Magri's account, but argue that Magri's own implementation does not fully account for the facts. As a background motivation to the alternative implementation I propose, in §3 I present an analogy between the problem of Free Choice and the problem of Homogeneity, motivating a unified account for the two phenomena. The account I will propose subsequently in §4 will rely on the Implicature account of Free Choice (Kratzer and Shimoyama 2002; Fox 2007, a.o.) and extend it to Homogeneity (along similar lines to what has been proposed in Bassi and Bar-Lev 2016 for Homogeneity with bare conditionals). The gist of the proposal is the claim that what Homogeneity and Free Choice phenomena share is a set of alternatives which is not closed under conjunction. I will then move on in §5 to claim that non-maximal readings are predicted in this view to arise as a result of ignoring some alternatives ('pruning'). Additional assumptions will however be needed in order to capture how the context affects the availability of such readings; I will propose a novel view of the context-sensitivity of implicature calculation to this end. In §6 I discuss the issue of 'gappy' (neither-true-nor-false) judgments often associated with Homogeneity, and propose that they are rather a side effect of Non-maximality. In §7 I return to the connection between Free Choice and Homogeneity and argue that differences between them can be reasonably explained despite the underlying similarity between the two phenomena within the view proposed here. Finally, in §8 I discuss Homogeneity with non-distributive predicates (observed by Križ 2015) and Homogeneity removal by plural quantification.

2 Asymmetries between positive and negative sentences

In this section I discuss two pieces of evidence for substantiating the Implicature account's prediction of asymmetry between positive and negative sentences (a third piece of evidence comes from Gappiness and will be discussed in §6):

(7) **Evidence for asymmetry**:

a. Experimental evidence by Tieu et al. (2015) that while some children accept sentences like *the hearts are red* as true in a situation where some but not all of the hearts are red (Karmiloff-Smith 1981; Caponigro et al. 2012), no children accept sentences like *the hearts aren't red* as true in such a situation (§2.1).

b. Novel data showing that non-maximal readings are more pervasive in positive sentences than in negative ones (§2.2).

negative sentences due to differences in the cost assigned to their alternatives. I will not pursue here a detailed comparison with this implementation, but merely point out that such differences are unlikely to provide a satisfactory explanation for the asymmetry in children's behavior to be discussed in §2.1, since children have been argued to lack access to the alternatives Spector's analysis relies on (see §7).

At the same time, I will argue (following and elaborating on Križ 2015) that Magri's own Implicature account suffers from several problems due to the way in which the implicature is derived (§2.3). I will conclude that while the general architecture proposed by Magri is promising, a different implementation is needed.

2.1 Asymmetry in the acquisition of Homogeneity

There has been evidence for a while (Karmiloff-Smith 1981; Caponigro et al. 2012) that, in contrast with adults, some children have a weak interpretation for positive sentences containing definite plurals. A sentence like *the hearts are red* is interpreted by these children as *some of the hearts are red* (an existential interpretation). Tieu et al. (2015) further investigated children's behavior with definite plurals in negative sentences, examining whether they have a parallel weak interpretation where a sentence like *the hearts aren't red* is interpreted as *not all the hearts are red* (a negated universal interpretation).

Given the symmetric approaches to Homogeneity, we expect children to behave in a symmetric way in positive and negative sentences. Namely, if they are more permissive than adults are with respect to positive sentences, they should be similarly permissive with respect to negative sentences, namely have a weak (negated universal) reading for such sentences. Given the Implicature account, on the other hand, being permissive in positive sentences might just amount to not computing the Maximality implicature. If this is the case, then it is expected that they would not be permissive in negative sentences and only get the basic strong (negated existential) interpretation.[7]

Tieu et al. (2015) have examined these predictions with 2 experiments. In the first experiment French-speaking children were asked to indicate their judgment of a puppet's description of various pictures by pointing at a sad face ('no' response) or a happy one ('yes' response). On the target conditions the puppet would describe for instance a picture where there are 2 red hearts and 2 yellow ones using the sentence *the hearts are(n't) red* (originally in French). Strikingly, while 8 out of the 24 children approved of the puppet's description of such a situation with a positive sentence like *the hearts are red* (namely, they responded 'yes' on at least two out of three cases of this sort presented to them), no child approved of the puppet's description of such a situation with a negative sentence like *the hearts aren't red*.[8] Compared to adults' responses for the same task, it turns out that only in positive sentences are children more permissive than adults are in their interpretation of sentences containing definite plurals.[9]

[7] This will only be the expectation with the aid of an auxiliary assumption ensuring that the permissive kids do not have access to an LF where negation takes scope below the definite plural.

[8] Among all 78 test trials of this condition (3 such test trials for each child, including 2 children who were excluded from analysis) there was only one 'yes' response; the two other responses by the same child on the same condition were 'no' responses.

[9] The opposite is true for negative sentences: adults turned out to be more permissive than children with such sentences. I set this difference aside for the sake of the discussion here, though one might entertain the idea that the behavior of the minority of adults who are permissive in negative sentences results from allowing to compute an implicature under negation (perhaps reanalyzing the sentence *the hearts aren't red* as involving 'meta-linguistic' negation).

	Adults	Children
Homogeneous	23	12
Existential	0	10
Universal	2	0

Table 1 Tieu et al.'s characterization of their experiment 2 results (based on their table 6). 'Homogeneous' behavior means universal interpretation in positive contexts and existential interpretation in negative contexts. 'Existential' behavior means existential interpretations in both and 'Universal' behavior means universal interpretations in both.

They further tested children's behavior using a ternary judgment task where instead of binary 'yes'/'no' responses the children were asked to reward the puppet according to the accuracy of its descriptions with one, two or three strawberries. Here too, children turned out to be more permissive than adults in positive sentences but not in negative ones: for instance, while in positive sentences the rate to which children gave the minimal reward to the puppet in the above situation was significantly lower than the adults' rate, it was adult-like in negative ones. Tieu et al.'s description of the findings of their second experiment are in table 1 (for the details and the analysis I refer the interested reader to their paper).

Crucially, while some children can be categorized as interpreting definite plurals existentially and some as interpreting them universally in positive sentences, no child can be categorized as having a weak negated universal interpretation in negative sentences. These results are puzzling for symmetric accounts, but corroborate the predictions of the asymmetric Implicature account.

2.2 Asymmetry with Non-maximality

Another reason to believe there is an asymmetry between positive and negative sentences which I would like to put forward comes from Non-maximality. Similarly to the explanation it provides for children's behavior which we have just discussed, the Implicature approach provides a natural source for Non-maximality as a case where the Maximality implicature is not computed. I will develop this perspective in detail in §5; for now, let us only have the following hypothesis in place:

(8) **Hypothesis**: Non-maximal readings may result from not computing the Maximality implicature.

This hypothesis predicts a major difference between positive and negative sentences: Only in positive sentences, where the maximal reading results from calculating an implicature, would this source of Non-maximality be available. When the definite plural is under negation, namely when the maximal reading follows from the basic semantics, it wouldn't. The goal of this section is to examine this prediction of the Implicature account combined with the hypothesis in (8) that there should be an asymmetry between positive and negative sentences with respect to Non-maximality. Whether this prediction is borne out will however turn out to be an elusive issue on which I will have to spend some ink. I will argue that once we control for the scope of the definite plural relative to negation in negative sentences, we indeed find asymmetries between positive and negative sentences which are difficult to explain assuming a non-implicature account.

Let us begin first with what may seem like a reason to believe there is no asymmetry. It has been claimed (by, e.g., Malamud 2012; Križ 2015) that Non-maximality applies to positive and negative sentences alike. Many speakers indeed judge (9b) as true as long as most of the kids don't laugh (assuming that this suffices for the clown's performance to be considered not funny):[10]

(9) a. *Context: There was a clown at my kid's birthday party. Someone asks me if they gave a funny performance. I reply:*
b. The kids didn't laugh.

However, concluding that there is no asymmetry based on (9b) would be too hasty for several reasons. One reason is that the facts are not as clean as they are sometimes reported. Some speakers do not in fact accept non-maximal readings for such sentences, while readily accepting parallel non-maximal readings for positive sentences, e.g., the judgment reported for *the kids laughed* in the same context, (3a). For these speakers, (9b) can only be true if no kid laughed, namely they only have a maximal reading.[11] In other words, it might seem like we find an asymmetry for some but not all speakers. Another reason is that nothing guarantees that the definite plural *the kids* in (9b) takes scope below negation. An LF where it takes scope above negation will be no different in terms of the predictions regarding Non-maximality than an LF with no negation at all since in both Homogeneity will be the result of implicature calculation. Namely, if we want to check for asymmetries examples like (9b) are not the right examples to look at. Finally, and perhaps most importantly, it is worth emphasizing that even if the definite plural could not take scope above negation, finding non-maximal readings for (9b) will not by itself falsify the predictions of our hypothesis. This is since it is conceivable that Non-maximality does not have just one source; there may well be routes to Non-maximality other than the one paved by the Implicature account, ones that can apply when the definite plural is under negation. I will leave a more detailed discussion of such routes to §5.3, and at this point only point out that nothing in our hypothesis leads to denying the existence of alternative routes.

What all of this goes to show is that we should not expect an asymmetry with Non-maximality to be easy to detect, even if there is indeed such an asymmetry. I will however argue that we can after all reveal an asymmetry by blocking one route to Non-maximality. Specifically, my argument will rely on asymmetry in cases where we control for the scope of the definite plural relative to negation. As we discussed in connection with (2), this can be done by having a variable within the definite plural

[10] The Non-maximality data reported in this section and in §5.3 are based on judgments by several native speakers of English as well as several native speakers of Hebrew on parallel Hebrew sentences. Informants were told what the context of utterance is (trying to establish a question under discussion) and were then asked whether the sentence seems true in a certain situation (e.g., if 5 out of the 30 kids laughed). Informants were also asked about what they think would be a natural inference from the sentence given only a context of utterance (e.g., for (10b), whether they infer that no kid took any vitamin or not).

[11] The idea that there might be a difference between positive and negative sentences with respect to Non-maximality was first pointed out to me by Elitzur Bar-Asher Siegal (p.c.), and was also brought to my attention by Manuel Križ (p.c.).

which is bound by a negative quantifier, as in (10b). And indeed, we find that there is a clear asymmetry when we compare (10b) with its positive counterpart in (10a). The maximal reading according to which no kid took any vitamin is clearly preferred for (10b); this is in contrast with (10a), in which for the sentence to be considered true it is enough that all the kids fulfill their requirement, i.e., that all the kids take at least two vitamins.

(10) *Context: the kids are required to take at least 2 of the 4 vitamins I gave each of them.*

a. All of the kids took their vitamins.
≈ Every kid took at least 2 vitamins. (non-maximal)

b. None of the kids took their vitamins.
≈ No kid took any vitamin (i.e., not even one vitamin). (only maximal)

If Non-maximality is a property of positive and negative sentences alike as predicted by non-implicature accounts of Non-maximality, this contrast is surprising. For why should we disprefer a reading for (10b) paraphrasable as *no kid took at least two vitamins*, which would be true as long as no kid fulfills the requirement?[12] Admittedly, the asymmetry here is relatively subtle: while the preferred reading is the maximal one, we would not obviously judge (10b) as false if we find out that one of the kids did take one of their vitamins. Note however that we did not control for all possible sources of Non-maximality: the subtlety here might be due to some other routes to Non-maximality being available, in which case a subtle asymmetry is arguably all we can hope for (for more on this see §5.3).

The nature of the argument for the Implicature account from (10) is hence the following: Once we control for scope, having non-maximal readings for negative sentences is difficult relative to positive ones. This is expected if there is one route to Non-maximality which is only available in positive sentences; since only the Implicature account provides such a route (namely the one based on not deriving the Maximality implicature), this constitutes evidence in its favor. To further illustrate the point and show that the asymmetry in (10) is not accidental, consider the following minimal pair.

(11) a. All of the kids came with their parents to the meeting.
b. None of the kids came with their parents to the meeting.

Consider now two possible contexts for uttering these sentences:

(12) a. *Context A: The kids were told to come with both of their parents to the meeting.*
b. *Context B: The kids were told to come with at least one parent to the meeting.*

[12] An alternative analysis of (10a) with no Non-maximality involved could be one where *their vitamins* is interpreted as *the vitamins they were supposed to take*, where this in turn is analyzed as an existential quantifier over pluralities of two vitamins. First, I would like to point out that it is not obvious to explain why given this analysis we should not have a parallel reading for (10b) where it means *no kid took at least two vitamins*. Second, note that other examples I use here to argue for asymmetry are not amenable to this kind of analysis, especially (14)-(15). I thank Kai von Fintel and Irene Heim for discussions of this issue.

The interpretation of the positive example in (11a) is easily affected by the difference between the two contexts, namely a non-maximal reading is possible in Context B but not in Context A, and in both the meaning can be paraphrased as *all the kids fulfilled their requirement*. The interpretation of the negative sentence (11b), on the other hand, is immune to contextual manipulations: in both contexts the maximal interpretation is preferred; while in Context B this meaning coincides with the paraphrase *no kid fulfilled their requirement*, in Context A it is stronger and its truth requires that every kid come with no parent to the meeting.

(13) a. (i) Interpretation of (11a) in Context A:
Every kid came with both parents to the meeting. (maximal)
(ii) Interpretation of (11a) in Context B:
Every kid came with at least one parent to the meeting. (non-maximal)
b. Interpretation of (11b) in both Context A and Context B:
No kid came with any parent to the meeting. (only maximal)

(14) and (15) are two more examples showing the same kind of asymmetry. In both, the definite plural contains a relative clause the interpretation of which varies with different assignments by a positive or negative quantifier over situations in order to control for scope. Once again, it is surprising that for instance while (15a) can be easily understood as conveying that the professor is attentive since she makes eye contact with most of the students in her classes, (15b) can't be easily understood as conveying that she is not so attentive since most of the students in her classes are such that she doesn't make eye contact with them. In contrast with (15a), the second sentence in (15b) provides a much stronger statement than needed to give justification for the first sentence; rather than merely conveying that she is not very attentive, it conveys that she is very inattentive.[13]

(14) a. In every performance by this clown the kids who sit on the first 3 rows laugh.
(Well, of course, not all of those kids do.)

[13] Note furthermore the effect of fixing the scope of the definite plural using a relative clause: In contrast with (14b), a non-maximal reading is accessible when the definite plural contains no relative clause as in (i). Unlike (14b), in this case the definite plural may refer to the plurality of all kids who are in any performance by this clown and take scope above negation.

(i) In no performance by this clown the kids laugh. (Well, of course, some of those kids do.)

This contrast may speak against responses which attribute the asymmetries discussed here to independent differences between positive and negative sentences. One such response has been proposed to me by Sophia Malamud (p.c.), suggesting that the typical Question Under Discussion is different with and without negation. Another response of this kind has been proposed to me by Manuel Križ (p.c.), suggesting that there is a difference between the conditions under which we conceive of the difference between *no kid laughed* and *almost no kid laughed* as irrelevant and those under which we conceive of the difference between *every kid laughed* and *almost every kid laughed* as irrelevant.

b. In no performance by this clown the kids who sit on the first 3 rows laugh.
(# Well, of course, some of those kids do.)

(15) a. She is a very attentive professor. Whenever she enters a class, she always makes eye contact with the students who sit in front of her.
(Well, of course, she doesn't always make eye contact with all of them.)

b. She is not a very attentive professor. Whenever she enters a class, she never makes eye contact with the students who sit in front of her.
(# Well, of course, she does occasionally make eye contact with some of them.)

The discussion in this section aimed to establish that there is an asymmetry between positive and negative sentences with respect to Non-maximality which supports the Implicature account: While positive sentences easily allow for non-maximal readings, such readings are at least dispreferred with negative sentences. This evidence for asymmetry is far from constituting a decisive argument in favor of the Implicature account. It does however point in the direction of such an account: it's difficult to explain the asymmetries we have found without assuming that some route to Non-maximality is available for positive sentences but not for negative ones. The Implicature account is unique in providing such a route, hence I take the data discussed here as an argument in its favor.

2.3 Shortcomings of Magri's Implicature account

We have seen two reasons to think that there are asymmetries between positive and negative sentences, based on children's behavior and Non-maximality. An Implicature account is thus promising since it holds the prospects for predicting such asymmetries, while no other account does. This is since it is unique in having the following asymmetry, repeated from (6), at its core:

(16) **Asymmetry within the Implicature account** (given basic meaning):

a. The kids laughed $\nRightarrow$ Kelly laughed

b. $\neg$ The kids laughed $\Rightarrow$ $\neg$ Kelly laughed

Unfortunately, Magri's own Implicature account suffers from several problems which make it an inadequate analysis of the phenomena at hand. This section is devoted to a brief discussion of Magri's account and some of the problems it faces (for a detailed criticism of his approach see Križ 2015), which will serve as a motivation for providing a different implementation of the Implicature account.

Magri's proposal can be summarized as follows (glossing over several important details): *The kids laughed* has the same basic meaning as *some of the kids laughed* (**some**). The Maximality implicature is derived by negating an alternative the meaning of which is *some but not all of the kids laughed* (**some** $\wedge\neg$ **all**), which together with the basic meaning entails that all the kids laughed (since **some** $\wedge\neg$ (**some** $\wedge\neg$ **all**) = **all**). That alternative is itself the result of computing a *some but not all* implicature for the sentence *some of the kids laughed*.

As has been discussed in Križ (2015), Magri crucially relies on the non-standard assumption that alternativehood is intransitive in order to avoid deriving *all of the kids laughed* as an alternative of *the kids laughed*, which would lead to blocking the computation of the Maximality implicature. While this assumption is not conceptually appealing, Magri's account faces more serious empirical problems, two of which are particularly relevant. First, while it may nicely explain why children are permissive in positive but not in negative sentences as we have discussed in §2.1, Tieu et al. (2015) point out that it does not account for the full trajectory of the acquisition of Homogeneity. Specifically, it does not account for children who have a 'homogeneous' behavior but do not compute *some-but-not-all* implicatures. This is since computing *some-but-not-all* implicatures is a precondition for deriving the Maximality implicature à la Magri.[14]

Second, as Križ (2015) has argued, Magri's account is not rich enough to offer a viable account for Non-maximality, and as such it provides no explanation for why Non-maximality is tied to Homogeneity. In other words, it is unable to operationalize the idea I argued for in §2.2, namely that at least one source of Non-maximality is the absence of the Maximality implicature. This is so for a relatively simple reason: within his account, *the kids laughed* could have two possible interpretations—an existential one (**some**) if the **some** $\wedge\neg$ **all** alternative is not negated, or a universal one (**all**) if it does. However, except in extreme cases,[15] Non-maximality normally gives rise to readings which are stronger than **some** but weaker than **all** (for a more detailed discussion and problematization see Križ 2015). Since it does not provide an analysis of Non-maximality to begin with, it doesn't benefit after all from the potential advantage of the Implicature account in predicting the asymmetries in Non-maximality we discussed in §2.2.

Here is the situation we are now facing: On one hand, the general architecture of the Implicature account is motivated by the asymmetries in acquisition and Non-maximality we have discussed; on the other hand, the Implicature account proposed by Magri (2014) is not only conceptually problematic, but in fact it fails to fully account for the acquisition and Non-maximality facts.

My goal in this paper is to revive the Implicature account by adopting the general architecture of Magri (2014)'s account, while providing a different implementation which will not suffer from the issues discussed above. Note that a common core of the problems discussed here for his account—the conceptual one and the empirical ones alike—is the choice of relying on a *some-but-not-all* implicature. The implementation

[14] A possible response to this objection against Magri's (2014) account might be that those children compute implicatures only if the resulting meaning is a complete answer to a reasonable question: for instance, *all the hearts are red* is a complete answer to the question *which hearts are red?*, but *some but not all of the hearts are red* is not. See Singh et al. (2016) for a related proposal.

[15] Such an extreme case has been discussed by Malamud (2012, ex. (2)):

(i) a. *Context: Mary has a large house with over a dozen windows in different rooms. She locks up and leaves to go on a road trip with her friend Max, forgetting to close just a few of the many windows in various rooms. A few minutes into the ride, Max says, "There is a thunderstorm coming. Is the house going to be OK?" Mary replies:*

b. Oh my, we have to go back – the windows are open!

I will suggest avoids all of these problems by taking a different route for computing the implicature, building on the Implicature account of Free Choice. As a preliminary motivation for the account I will propose in §4, in the next section I present an analogy between the phenomena of Homogeneity and Free Choice.

3 Analogy between Free Choice and Homogeneity

If Homogeneity is to be given an Implicature account as Magri (2014) proposed and as the asymmetries discussed above suggest, one may wonder if there's any other phenomenon which behaves in a similar way and that we have reasons to believe involves an implicature. In what follows I argue that indeed Free Choice is such a phenomenon, thereby motivating an Implicature account of Homogeneity which is modeled after the Implicature account of Free Choice. The puzzle of Free Choice (von Wright 1968; Kamp 1974) is the fact that a sentence of the form *allowed (a or b)*, as in (17), seems to be equivalent to *allowed a and allowed b*, even though given standard semantics for *allowed* and *or* we would expect a much weaker meaning, equivalent to *allowed a or allowed b*.[16]

(17) Mary is allowed to read War and Peace or Anna Karenina.
- a. $\approx$ Mary is allowed to read W&P and allowed to read AK.
- b. $\not\approx$ Mary is allowed to read W&P or allowed to read AK (she may be prohibited to read one of them).

Beginning with Kratzer and Shimoyama (2002), an influential view has been that the standard semantics is correct, and that the Free Choice inference from *allowed (a or b)* to *allowed a and allowed b* is an implicature. Alonso-Ovalle (2005) pointed out that maintaining the standard semantics has the advantage of predicting the correct behavior under negation:

(18) Mary is not allowed to read War and Peace or Anna Karenina.
- a. $\not\approx$ It's not the case that Mary is both allowed to read W&P and allowed to read AK (she may be allowed to read one of them).
- b. $\approx$ It's not the case that Mary is allowed to read W&P or allowed to read AK (she is allowed to read none of them).

Taking into account Alonso-Ovalle's point about the interpretation of the negative sentence in (18), the problem of Free Choice is no longer merely a problem assuming the standard semantics for *allowed* and *or*; a non-standard semantics where the Free Choice inference follows from the basic semantics will run into the problem of explaining (18).[17] The problem can be characterized as in (19), which has a by-now-familiar shape. Based on the judgment in (17), one might suggest that (19a) holds.

[16] Free Choice effects have been argued to arise even when disjunction takes scope above the modal (Zimmermann 2000). See Klinedinst and Rothschild (2012); Bar-Lev and Fox (2019) for the claim that such examples should be analyzed differently from the case discussed here where disjunction has narrow scope.

[17] See Aloni (2007, 2016); Barker (2010); Starr (2016); Willer (2017), a.o., for proposals relying on such a non-standard semantics and their treatment of the behavior under negation pointed out by Alonso-Ovalle.

And based on the judgment in (18), one might suggest that (19b) holds. Given the rule of contraposition and these two assumptions, we conclude (19c), which is obviously nonsensical if taken at face value: *Mary is allowed to read W&P or AK* is not intuitively equivalent to *Mary is allowed to read W&P*.

(19) **The puzzle of Free Choice**:
  a. $\Diamond$ Mary read W&P or AK $\Rightarrow$ $\Diamond$ Mary read W&P
  b. $\neg\Diamond$ Mary read W&P or AK $\Rightarrow$ $\neg\Diamond$ Mary read W&P
  c. $\therefore$ $\Diamond$ Mary read W&P or AK $\Leftrightarrow$ $\Diamond$ Mary read W&P

The puzzle of Free Choice is then completely parallel to the puzzle of Homogeneity in (5), repeated here.

(20) **The puzzle of Homogeneity**: [=(5)]
  a. The kids laughed $\Rightarrow$ Kelly laughed
  b. $\neg$ The kids laughed $\Rightarrow$ $\neg$ Kelly laughed
  c. $\therefore$ The kids laughed $\Leftrightarrow$ Kelly laughed

The Implicature view of Free Choice proposed by Kratzer and Shimoyama (2002) solves the problem by denying (19a) and accounting for this inference as an implicature. The Implicature view of Homogeneity proposed by Magri (2014) similarly solves the problem by denying (20a) and accounting for this inference as an implicature. In the next section I will take the analogy one step forward, and propose that what has been argued by Fox (2007) to be the underlying property responsible for the Free Choice effect is responsible for the Homogeneity effect as well. Before I do that, I'd first like to strengthen the analogy and show that it extends to several other environments.

In the case of Free Choice, in both positive and negative sentences (*Mary is/isn't allowed to read W&P or AK*) the result of replacing disjunction with the individual disjuncts (e.g., *Mary is/isn't allowed to read W&P*) is inferred; in the case of Homogeneity, in both positive and negative sentences (*the kids laughed/didn't laugh*) the result of replacing a plurality-denoting constituent with a constituent denoting part of this plurality (e.g., *Kelly laughed/didn't laugh*) is inferred. This characterization of the facts extends to non-monotonic environments, as in (21) and (22). Suppose War and Peace is one of the books. For (21) to be true it must be the case that exactly one student read War and Peace (ignoring the availability of non-maximal readings); and for (22) to be true it must be the case that exactly one student is allowed to read War and Peace.[18]

(21) Exactly one student read the books. $\approx$
  a. One student read *all* of the books, and
  b. No more than one student read *any* of the books.

[18] Note that the same point holds for (21) and (22) without *exactly*, in which case the 'negative parts' of the meaning in (21b) and (22b) are arguably cancellable implicatures. While I will not propose an explicit account of sentences like (21) or its counterpart without *exactly*, given the account to be proposed in §4 such an account will be dependent on an account for the parallel Free Choice data, an issue which is discussed in Bar-Lev (2018).

(22) Exactly one student is allowed to read War and Peace or Anna Karenina. ≈

a. There is one student who is both allowed to read W&P *and* allowed to read AK, and
b. No more than one student is allowed to read W&P *or* AK.

VP ellipsis constructions where the antecedent and the elided material are in environments of different monotonicity show the same pattern: For (23) to be true it must be the case that Mary but not John read War and Peace (once again ignoring Non-maximality); and for (24) to be true it must be the case that Mary but not John is allowed to read War and Peace.[19, 20]

(23) Mary read the books, but John didn't. ≈

a. Mary read *all* of the books, and
b. John didn't read *any* of the books.

(24) Mary is allowed to read War and Peace or Anna Karenina, but John isn't. ≈

a. Mary is both allowed to read W&P *and* allowed to read AK, and
b. John isn't allowed to read W&P *or* AK.

The analogy between Free Choice and Homogeneity I presented here calls for a unified perspective on the two problems. In what follows I will take up this task and propose an account of Homogeneity which mimics Fox's (2007) account of Free Choice.[21] Before moving on to build our analysis of Homogeneity based on the analogy, a word of caution is in order. There are some differences between Free Choice and

---

[19] Such data speak against Ambiguity approaches to Homogeneity (Krifka 1996) and Free Choice (Aloni 2007) on some plausible assumptions about ellipsis licensing. See Spector (2013); Križ and Spector (2017) for a more involved Ambiguity account of Homogeneity which accounts for such data.

[20] Definite plurals and Free Choice disjunction in Strawson-Downward Entailing environments such as the antecedent of a conditional do not conform to the same inference pattern. The analogy however still persists in such environments: for both one may construct examples where a locally weak (existential/disjunctive) interpretation is preferred as well as examples where a locally strong (universal/conjunctive) interpretation is preferred:

(i) Locally strong interpretation preferred:

a. If you read the books you will succeed.
b. If you are allowed to read War and Peace or Anna Karenina you have a difficult choice to make.

(ii) Locally weak interpretation preferred:

a. If you touch the statues you'll be kicked out. (after Križ 2015, p. 27, ex. (89))
b. If you are allowed to have ice cream or cake you have nothing to complain about.

That said, there is a difficult to pinpoint intuition that Free Choice disjunction tends towards the weak interpretation in such contexts. If there's something to that intuition, it could be due to the possibility of a scope construal where disjunction takes scope above the existential modal, a possibility which will deliver a disjunctive meaning on the Implicature account of Free Choice and one which will have no parallel in the case of definite plurals given the account I will propose. For discussions of these issues I thank Paolo Santorio and Frank Staniszewski.

[21] Of course, one may accept the argument that a unified account is desired but reject the choice of an Implicature account to this end. Such a move has been recently made by Goldstein (2018) who analyzes Free Choice based on Trivalence accounts of Homogeneity. I will take the Implicature route, based on the arguments for asymmetry discussed in §2.

Homogeneity which are worth mentioning: (i) Non-maximality has no parallel in the case of Free Choice disjunction, and (ii) Maximality and Free Choice inferences are not similarly cancellable. We should be careful then to account for the differences as well as for the similarities. I set aside the differences for now and return to discuss them in §7, where I argue that the underlying property unifying the two phenomena in my account still leaves room for variation.

4 Proposal

4.1 The proposal in a nutshell (in analogy with Free Choice)

4.1.1 Fox's Implicature account of Free Choice

The account of Free Choice proposed in Fox (2007) has three main ingredients. I will go over them briefly and then present how they are going to translate into my account of Homogeneity. First, the basic semantics assumed is what's predicted by the standard semantics of *allowed* and *or*, which is a disjunctive meaning.

(25) **Basic weak semantics**:
⟦Mary is allowed to read W&P or AK⟧ ⇔
$\diamond$ Mary read W&P $\vee$ $\diamond$ Mary read AK

Second, the relevant property responsible for the Free Choice effect according to Fox is having a set of alternatives which is not closed under conjunction. The reason why Free Choice disjunction gives rise to such a set of alternatives has to do with the syntax of structures of the form *allowed (a or b)*. Standardly, disjunction gives rise to alternatives where disjunction is replaced with the disjuncts and with conjunction (see Sauerland 2004; Katzir 2007). These replacements yield in the case of *allowed (a or b)* the alternatives *allowed a*, *allowed b*, and *allowed (a and b)*. But the conjunction of the first two alternatives is not derived by this procedure: their conjunction *(allowed a) and (allowed b)* is weaker than *allowed (a and b)*, and is absent from the set of alternatives. Fox (2007) has argued that non-closure under conjunction is a necessary property for a sentence with a disjunctive meaning to end up having a conjunctive one; this property distinguishes simple disjunction (*a or b*) from Free Choice disjunction.

(26) **Non-closure under conjunction**:
a. *Alt* $\supseteq$ {$\diamond$ Mary read W&P $\vee$ $\diamond$ Mary read AK,
$\diamond$ Mary read W&P, $\diamond$ Mary read AK}
b. *Alt* $\not\ni$ $\diamond$ Mary read W&P $\wedge$ $\diamond$ Mary read AK

Finally, Fox utilizes an exhaustivity mechanism which given the basic weak semantics and the set of alternatives which is not closed under conjunction derives the conjunctive meaning. Fox's own mechanism utilizes two applications of the exhaustivity operator $\mathcal{E}xh$; since I will assume Bar-Lev and Fox's (2017) modification of $\mathcal{E}xh$, one application will suffice.

(27) **Strengthening**:
$[\![\mathcal{E}xh_{Alt}(\text{Mary is allowed to read W\&P or AK})]\!] \Rightarrow$
$\Diamond$ Mary read W&P $\wedge$ $\Diamond$ Mary read AK

4.1.2 A parallel Implicature account of Homogeneity

The account of Homogeneity I propose here has the following three ingredients, in parallel to Fox's account of Free Choice. First, I will assume a basic weak semantics following Magri (2014), such that if the kids are Kelly and Jane, the basic semantics of *the kids laughed* will only require that at least one of them laugh.

(28) **Basic weak semantics**:
$[\![\text{The kids laughed}]\!] \Leftrightarrow$ Kelly laughed $\vee$ Jane laughed

Second, the set of alternatives of *the kids laughed* will contain alternatives equivalent to *Kelly laughed*, *Jane laughed*, but crucially not conjunctions of such alternatives. It will thus be a set of alternatives which is not closed under conjunction, namely share the property responsible for the Free Choice effect in Fox's account (though, as will become clear later, the reason for the absence of conjunctive alternatives will be quite different than in the case of Free Choice).

(29) **Non-closure under conjunction**:
a. Alt = {Kelly laughed $\vee$ Jane laughed,
Kelly laughed, Jane laughed}
b. $Alt \not\ni$ Kelly laughed $\wedge$ Jane laughed

Third, by applying the exhaustivity mechanism which derives Free Choice inferences, we will derive the Maximality implicature:

(30) $[\![\mathcal{E}xh_{Alt}(\text{The kids laughed})]\!] \Leftrightarrow$ Kelly laughed $\wedge$ Jane laughed

In what follows I make the proposal explicit, presenting each of these ingredients in turn.

4.2 Basic weak semantics: an existential pluralization operator

Where does the weak meaning of *the kids laughed* come from? For Magri (2014), it is the definite plural itself which is interpreted as an existential quantifier. I will take a different path, following arguments by Križ (2015) against treating the definite plural as an existential (and more generally against the definite plural being the source of Homogeneity). I assume that the definite plural *the kids* denotes, as is standard, the plurality containing all kids; if the kids are Kelly and Jane, then $[\![\text{the kids}]\!] = Kelly \sqcup Jane$. Instead of blaming the definite plural, I follow Schwarzschild (1994); Gajewski (2005) who proposed that Homogeneity has to do with a pluralization operator which applies to the predicate; they implemented this idea within a Trivalence account of Homogeneity, while here I propose an implementation within an Implicature account.

Against common wisdom which dictates that pluralization is carried out by utilizing a universal distributivity operator (Link 1987; Schwarzschild 1996, a.o.) or a star operator (Link 1983; Schwarzschild 1994, a.o.), I utilize the existential pluralization operator ∃-PL, defined in (31a) (I postpone a discussion of the stipulative nature of this assumption to §4.5). The meaning of sentences involving plural predication which is commonly attributed to the workings of the pluralization operator alone will fall out from a conspiracy of ∃-PL together with the exhaustivity mechanism to be explicated later. ∃-PL takes a domain variable D, a property P and an individual x, and returns the proposition that some individual in the domain D which is part of x satisfies the property P.[22] Since the discussion in this paper is mostly restricted to distributive predicates, I only consider the atomic parts of x, as in (31b) (see the discussion of Homogeneity with non-distributive predicates in §8.1 where this stipulation is disposed of).

(31) a. $[\![\exists\text{-PL}]\!] = \lambda D_{et}.\lambda P_{\langle e,st\rangle}.\lambda x_e.\lambda w_s.\exists y \in D \cap Part_{AT}(x)[P(y)(w) = 1]$
b. $Part_{AT}(x) = \{y : y \sqsubseteq_{AT} x\}$ (i.e., the set of atomic parts of x)

∃-PL, I assume, is obligatorily present whenever there is predication of a plurality-denoting constituent. This assumption can be seen as a restatement of Schwarzschild's (1994) assumption that plural VPs are obligatorily starred, which has been reformulated in Kratzer (2007) as the claim that "sister constituents of plural DPs are pluralized". The LF of *the kids laughed* in (1a) (before applying the strengtehing mechanism) is hence (32).

(32) LF: [The kids] [∃-PL$_D$ laughed]

The subscript D on ∃-PL is a domain variable, which is assumed here to restrict any natural language quantifier (see Westerståhl 1984; von Fintel 1994; Stanley and Gendler Szabó 2000; Chierchia 2013 among many others); the importance of this variable to the proposal will become clear in §4.3. I will assume throughout the paper that D must be assigned a set containing all parts of the definite plural's denotation. (32) will then yield the proposition that at least one of the kids laughed.

(33) $[\![(32)]\!](w) = 1$ *iff* $\exists y \in D \cap Part_{AT}([\![\text{the kids}]\!])[y \text{ laughed in } w]$

Assuming again that $[\![\text{the kids}]\!] = Kelly \sqcup Jane$, we get the desired basic weak meaning.

(34) $[\![(32)]\!] \Leftrightarrow$ Kelly laughed $\vee$ Jane laughed

4.3 Subdomain alternatives and non-closure under conjunction

Recall that what's responsible for the Free Choice effect according to Fox is that Free Choice disjunction gives rise to a set of alternatives which is not closed under conjunction. The gist of my proposal is that having a set of alternatives which is

[22] One-place predicates are assumed to be of type $\langle e, st\rangle$ (since the exhaustivity mechanism I will introduce in §4.4 operates on propositions). The domain argument of ∃-PL is assumed to be of type et for simplification.

not closed under conjunction is the common core of Homogeneity and Free Choice and the underlying reason for their analogous behavior. In order to derive a set of alternatives which is not closed under conjunction, given the basic meaning which is a disjunction as in (34), I will propose: (i) that alternatives equivalent to the disjuncts of this disjunction (*Kelly laughed*, *Jane laughed*) are generated; and (ii) that their conjunction is not. Let us discuss each of these assumptions in turn.

Malamud (2012) (and following her Križ and Spector 2017) proposed that for interpreting sentences containing definite plurals we reason about propositions derived on the basis of replacing the definite plural's denotation with its parts; in the case at hand, those will be propositions such as *Kelly laughed* and *Jane laughed*. Malamud's insight is that reasoning about such propositions is fruitful for explaining Non-maximality (see Križ and Spector 2017 for a modification of her system that captures Homogeneity as well). I will make use of this insight in my account of Non-maximality in §5; here I'd only like to point out that by taking Malamud's idea together with Magri's assumption that the basic semantics is existential, we derive a set of propositions with precisely the property we are after, namely a set which is not closed under conjunction.[23]

(35) a. Kelly laughed ∨ Jane laughed
$\lambda w.\exists y \in Part_{AT}([\![\text{the kids}]\!])[y \text{ laughed in } w]$
b. Kelly laughed $\lambda w.\exists y \in Part_{AT}(\text{Kelly})[y \text{ laughed in } w]$
c. Jane laughed $\lambda w.\exists y \in Part_{AT}(\text{Jane})[y \text{ laughed in } w]$

How do we derive such propositions in the current system? An obvious choice would be to replace *the kids* with its parts. However, this choice involves several complications. First, since alternatives are syntactic objects, it will have to be assumed that every part of the denotation of *the kids* has a corresponding (salient) name to be substituted for *the kids*. This may be a reasonable assumption when *the kids* denotes a plurality of two, but it ceases to be natural once we move to, say, a plurality of 500 kids. Second, as Križ and Spector (2017) pointed out, this choice leads to problems when the definite plural does not simply denote an individual but contains a variable which makes its denotation assignment dependent, as in (36):

(36) Every kid read their books.

Third, this choice will make the extension of the system to non-distributive predicates (to be discussed in §8.1) very difficult.

For these reasons, I choose to avoid this move and utilize instead a different mechanism, which has been employed elsewhere for similar purposes. For their treatment of NPIs, Krifka (1995); Chierchia (2013) employed alternatives resulted from replacing the domain of quantification with its subdomains. I utilize this mechanism for deriving the alternative propositions here, and this is where the domain variable

[23] Malamud's idea by itself does not guarantee it; on her proposal, the set of propositions is closed under conjunction.

restricting ∃-PL comes into play. Specifically, the alternatives are derived by replacing the domain variable with its subsets:[24]

(37) $Alt(\exists\text{-PL}_D) = \{\exists\text{-PL}_{D'} : D' \subseteq D\}$

The set of alternatives of (32) is thus:

(38) $Alt((32)) = \{$[The kids] [$\exists$-PL$_{D'}$ laughed] $: D' \subseteq D\}$

Now take for example a subset D' which is a set containing Kelly but not Jane, and D'' which is a set containing Jane but not Kelly. Then the alternatives generated from replacing D with these subsets are the disjuncts of the disjunction the prejacent is equivalent to.[25] We thus derive the desired set of alternatives which is not closed under conjunction.

(39) a. ⟦[The kids] [∃-PL$_D$ laughed]⟧ ⇔ Kelly laughed ∨ Jane laughed
b. ⟦[The kids] [∃-PL$_{D'}$ laughed]⟧ ⇔ Kelly laughed
c. ⟦[The kids] [∃-PL$_{D''}$ laughed]⟧ ⇔ Jane laughed

In order to simplify the discussion to follow, we'll rely on the assumption mentioned above that D must be a set containing all parts of the definite plural's denotation and use a definition of the prejacent and the set of alternative propositions of *the kids laughed* based on the set of propositions of the form *x laughed* where *x* is one of the kids. The basic meaning of *the kids laughed* which I call *Prej* is the disjunction of all members in that set, (40a), and the set of its alternative propositions (those which matter, see fn. 25) which I call *SubAlt* is its closure under disjunction, (40b).

(40) a. $Prej = [\![\text{the kids } \exists\text{-PL}_D \text{ laughed}]\!] = \bigvee\{p : \exists y \in Part_{AT}([\![\text{the kids}]\!])[p = \lambda w.y \text{ laughed in } w]\}$
b. **Alternative propositions of *the kids* ∃-*PL*$_D$ *laughed***:
$SubAlt = \{p : \exists y \in Part_{AT}([\![\text{the kids}]\!])[p = \lambda w.y \text{ laughed in } w]\}^{\vee}$

Recall that for the set of alternatives not to be closed under conjunction, it is crucial that there be no alternative whose meaning is equivalent to the conjunction *Kelly laughed* ∧ *Jane laughed*. Namely, sentences such as *each/all of the kids laughed* cannot be alternatives to *the kids laughed*. This, I assume, is due to the fact that the DP in *each/all of the kids laughed* is syntactically more complex than in *the kids laughed*.

[24] I allow myself here a common abuse of notation: D is a syntactic object and not a set. We can define the alternatives more carefully by assuming that D carries an index i, and that the alternatives carry other indices as in (i). Assuming that the assignment function g is surjective we will end up with all subsets of $g(i)$ as alternatives (and potentially other sets as well, which however do not make any difference for our purposes). I thank Roger Schwarzschild for a discussion of this issue.

(i) $Alt(\exists\text{-PL}_{D_i}) = \{\exists\text{-PL}_{D_j} : j \in \mathbb{N}\}$

[25] Each of the subsets of D will either yield a proposition equivalent to one of those in (39), or lead to vacuous quantification (given any subset of D which contains neither Kelly nor Jane). One has to ensure that the latter kind of alternatives is somehow treated differently than the former: they may either not be considered as alternatives at all, or be trivially excluded by the exhaustivity mechanism. I do not intend to propose here an explicit way to ensure it, and henceforth simply ignore such alternatives. (A similar issue arises for Chierchia 2013, e.g., for the analysis of *John is allowed to eat any cookie*.)

Given the independently motivated view of alternativehood in Katzir (2007); Fox and Katzir (2011), the generation of alternatives cannot lead to syntactically more complex structures (in §4.5 I will return to discuss this issue in greater detail and with particular focus on floating quantifiers as potential universal alternatives).[26] Note that the reason for non-closure under conjunction is quite different between Free Choice and Homogeneity: in the case of Free Choice we generate an alternative where disjunction is replaced with conjunction, but (due to the existential modal) this alternative is not the conjunction of other alternatives; in the case of Homogeneity I assume that there is no scalar alternative to begin with (a more nuanced view will be discussed in §8.3).

4.4 Strengthening

As has been mentioned, Fox (2007) argued that the ability of Free Choice disjunction to be strengthened into conjunction results from the non-closure under conjunction of its set of alternatives. This has since turned out to be an empirical generalization, motivated not only on the basis of Free Choice phenomena but for certain simple disjunctions as well: disjunctions having disjunctive alternatives but lacking a conjunctive alternative may end up as conjunctions. Bowler (2014), for example, has argued that Warlpiri *manu*, which behaves like disjunction in Downward Entailing environments but like conjunction in Upward Entailing ones, is a disjunction which lacks a lexical alternative corresponding to *and*. Singh et al. (2016) argued that some children's conjunctive interpretation of disjunction results from their inability to access a conjunctive alternative.[27] Given the assumptions made above, namely that the basic meaning of *the kids laughed* is a disjunction, and that its set of alternatives contains disjunctive alternatives but no conjunctions thereof, Homogeneity falls within the scope of this generalization.[28] In this section I explicate the specific mechanism of strengthening by which I assume those disjunctions end up as conjunctions, and apply it to strengthen the basic existential meaning of *the kids laughed* into a universal one.

For concreteness, I adopt Bar-Lev and Fox's (2017) modification of the system of implicature calculation in Fox (2007). In order to illustrate how the system works, capitalizing on the importance of non-closure under conjunction for deriving the Maximality implicature, I will apply it to the two sentences in (41). In the current system their basic meanings are both equivalent to the disjunction in (42) (assuming that the kids are Kelly and Jane).

[26] Matthew Mandelkern (p.c.) pointed out that structural complexity isn't enough for the current purposes. In order to avoid alternatives like *every kid laughed* I will have to assume that a sort of semantic complexity plays a role: *the kids* cannot be replaced with *every kid* which is of a higher type.

[27] For further analyses relying on the same idea see Bar-Lev and Margulis's (2014) treatment of Hebrew *kol*, Meyer's (2016) treatment of English *or else*, and Oikonomou's (2016) theory of imperatives.

[28] As a reviewer points out, given this view one might expect some cross-linguistic variation with respect to Homogeneity effects, assuming that some languages could have both existential and universal pluralization operators. However, I am not aware of any language which lacks Homogeneity effects altogether. In §8.3 I will entertain the possibility that even English has a universal pluralization operator, albeit one which is only licensed by certain DPs. Note however that the licensing conditions for this operator could vary from one language to another. In fn. 52 I consider such variation as a possible source for the difference between English and French in whether definite plurals containing numerals show Homogeneity effects.

(41) a. **Simple disjunction**: Kelly or Jane laughed.
b. **Homogeneity**: The kids laughed.

(42) *Prej* = Kelly laughed ∨ Jane laughed

Their sets of alternatives however are different: while (41a) gives rise to a conjunctive alternative on top of the alternatives in *SubAlt*, (41b) only has *SubAlt* alternatives.

(43) a. $Alt((41a)) = \{$Kelly laughed ∨ Jane laughed, Kelly laughed, Jane laughed, Kelly laughed ∧ Jane laughed$\}$
b. $Alt((41b)) = SubAlt = \{$Kelly laughed ∨ Jane laughed, Kelly laughed, Jane laughed$\}$

Fox (2007) proposed the notion of Innocent Exclusion, designed to avoid deriving a contradictory strengthened meaning for any disjunction which has the disjuncts as members in its set of alternatives. Since both *Kelly laughed* and *Jane laughed* are stronger than *Kelly laughed* ∨ *Jane laughed*, a naïve view of implicature calculation would predict that both should be negated; but this leads to the contradiction (*Kelly laughed* ∨ *Jane laughed*) ∧ ¬ *Kelly laughed* ∧ ¬ *Jane laughed*. Innocent Exclusion avoids such contradictions by following the procedure in (44).

(44) **Innocent Exclusion procedure**:
a. Take all maximal sets of alternatives that can be assigned false consistently with the prejacent.
b. Only exclude (i.e., assign false to) those alternatives that are members in all such sets—the **Innocently Excludable** (=IE) alternatives.

Formally, the set of IE alternatives is defined as follows:

(45) Given a sentence p and a set of alternatives C:
$IE(p,C) = \bigcap\{C' \subseteq C : C'$ is a maximal subset of C, s.t. $\{\neg q : q \in C'\} \cup \{p\}$ is consistent$\}$

Since the falsity of *Kelly laughed* and the falsity of *Jane laughed* together is not consistent with the prejacent *Kelly laughed* ∨ *Jane laughed*, none of the maximal sets on the first step of the procedure will contain both, and consequently none of these alternatives will be IE on the second step. The conjunctive alternative *Kelly laughed* ∧ *Jane laughed* on the other hand ends up in all such maximal sets, and hence it is IE. In other words, while for (41a) there is an IE alternative (the conjunctive one), for (41b) there is none: all alternatives are non-IE.

(46) a. $IE((41a))(Alt((41a))) = \{$*Kelly laughed* ∧ *Jane laughed*$\}$
b. $IE((41b))(Alt((41b))) = \varnothing$

Bar-Lev and Fox (2017) add another procedure on top of Innocent Exclusion, called Innocent Inclusion, motivated by considerations of Free Choice phenomena.[29]

[29] Bar-Lev and Fox (2017) argued that the direct derivation of Free Choice provided by assuming Innocent Inclusion is advantageous over Fox's (2007) indirect derivation which relies on an iterative application of

(47) **Innocent Inclusion procedure**:

a. Take all maximal sets of alternatives that can be assigned true consistently with the prejacent and the falsity of all IE alternatives.

b. Only include (i.e., assign true to) those alternatives that are members in all such sets—the **Innocently Includable** (=II) alternatives.

Formally, the set of II alternatives is defined as follows:

(48) Given a sentence p and a set of alternatives C:
$II(p,C) = \bigcap\{C'' \subseteq C : C''$ is a maximal subset of C, s.t.
$\{r : r \in C''\} \cup \{p\} \cup \{\neg q : q \in IE(p,C)\}$ is consistent$\}$

Are *Kelly laughed* and *Jane laughed* II? It depends on what the truth of the prejacent together with the falsity of the IE alternatives entails. In the case of (41a), it would entail the falsity of *Kelly laughed* $\wedge$ *Jane laughed*. Given this result, none of the maximal sets on the first step of the Inclusion procedure will contain both *Kelly laughed* and *Jane laughed*, and as a result none of them ends up II. In the case of (41b), on the other hand, where the only alternatives are disjunctive alternatives and hence nothing is IE, there is only one maximal set of alternatives that can be assigned true consistently with the prejacent, namely the whole set of alternatives. It follows then that all the alternatives are Innocently Includable. By assigning *Kelly laughed* and *Jane laughed* truth, we derive the desired conjunctive result *Kelly laughed* $\wedge$ *Jane laughed*. As desired, we only derive this conjunction for (41b); for (41a) we derive its negation, due to the conjunctive alternative being IE.

(49) a. $II((41a))(Alt((41a))) = \{$*Kelly laughed* $\vee$ *Jane laughed*$\}$

b. $II((41b))(Alt((41b))) = Alt((41b)) = SubAlt$

The exhaustivity operator $\mathcal{E}xh^{\text{IE+II}}$ incorporates the two procedures, and is defined as follows (from now on I will omit the superscript and refer to it simply as $\mathcal{E}xh$):

(50) **Innocent-Exclusion+Innocent-Inclusion-based exhaustivity operator**:
$$[\![\mathcal{E}xh^{\text{IE+II}}]\!](C)(p)(w) \Leftrightarrow \forall q \in IE(p,C)[\neg q(w)] \wedge \forall r \in II(p,C)[r(w)]$$

When applied to (41a) and (41b), we get:

(51) a. Result for Simple disjunction:
$[\![\mathcal{E}xh]\!](Alt((41a)))((41a)) =$
(*Kelly laughed* $\vee$ *Jane laughed*) $\wedge \neg$(*Kelly laughed* $\wedge$ *Jane laughed*)

b. Result for Homogeneity:
$[\![\mathcal{E}xh]\!](Alt((41b)))((41b)) = \bigwedge Alt((41b)) = \bigwedge SubAlt =$
Kelly laughed $\wedge$ *Jane laughed*

an Innocent Exclusion-based exhaustivity operator $\mathcal{E}xh^{\text{IE}}$: assuming Innocent Inclusion provides a global derivation for Universal Free Choice inferences (argued to be needed in Chemla 2009) which is unavailable with iterative applications of $\mathcal{E}xh^{\text{IE}}$, and furthermore can account for the interaction of Free Choice disjunction with *only*. In Bar-Lev (2018); Bar-Lev and Fox (2020) further evidence for Innocent Inclusion is provided based on Simplification of Disjunctive Antecedents and cases of Free Choice where a universal quantifier intervenes between the existential modal and disjunction.

To guarantee that the Maximality implicature is not an optional one, a crucial assumption I make is that applying the exhaustivity operator, at least at matrix position, is obligatory (Magri 2009). (My account of Non-maximality in §5, which essentially makes the Maximality implicature optional in certain contexts, still relies on this assumption. Optionality will sneak in, just as in Magri 2009, by manipulating the set of alternatives.)

Putting all the pieces together, the LF we get for *the kids laughed* in (1a) is as follows:

(52) LF of (1a): $\mathcal{E}xh_C$ [[The kids] [∃-PL$_D$ laughed]]
(where $C = \{$[The kids] [∃-PL$_{D'}$ laughed] $: D' \subseteq D\}$)

In the general case, the result of exhaustification is the conjunction of all the alternatives resulted from replacing the domain variable D with subsets of D which have a non-empty intersection with the set of atomic parts of ⟦the kids⟧ (see fn. 25); this boils down to the conjunction of all members of *SubAlt* assuming that D contains all the kids.

(53) $[\![(52)]\!](w) = 1$ *iff* $\forall D' \subseteq D[D' \cap Part_{AT}([\![\text{the kids}]\!]) \neq \varnothing \rightarrow$
$\exists y \in D' \cap Part_{AT}([\![\text{the kids}]\!])[y \text{ laughed in } w]]$ *iff*
$\forall y \in D \cap Part_{AT}([\![\text{the kids}]\!])[y \text{ laughed in } w]$

By exhaustification then we get the desired maximal interpretation according to which all the kids laughed. Importantly, whether we apply $\mathcal{E}xh$ for the LF of *the kids didn't laugh* in (1b) doesn't matter for the truth conditions. As long as it doesn't apply under negation (see Fox and Spector 2018) the sentence will only be true if no kid laughed.

(54) LF of (1b): ($\mathcal{E}xh_C$) NEG [[The kids] [∃-PL$_D$ laughed]]

(55) $[\![(54)]\!](w) = 1$ *iff* $\neg\exists y \in D \cap Part_{AT}([\![\text{the kids}]\!])[y \text{ laughed in } w]$

4.5 On floating quantifiers and the lack of a conjunctive alternative

The account just presented relies on two assumptions which may seem particularly stipulative: One is that pluralization is carried out by an existential operator, and another is the assumption that the LF of a sentence featuring this operator, such as *the kids laughed*, does not have an alternative LF with a universal meaning. I would like to first emphasize that I take these two stipulations to be the (perhaps unfortunate) cost of capturing the asymmetry between positive and negative sentences: This asymmetry points towards a basic existential meaning which is strengthened into a universal one; to capture it, one has to posit some existential operator, and furthermore assume the absence of a universal alternative since otherwise strengthening into a universal meaning would be blocked. While I do not intend to deny the stipulative nature of these two assumptions, I would like to provide here a (weak) defense of them against immediate challenges arising from the existence of universal floating quantifiers like *each* and *all*.

The common assumption that pluralization operations have universal force seems natural given the existence of universal floating quantifiers, which are often taken to be their overt counterparts (see Champollion 2016a,b). In contrast, the existential pluralization operator ∃-pl seems to have no overt counterparts, and as a result positing it may seem like an entirely ad hoc stipulation (I thank a reviewer for raising this objection). I would like to argue, however, that it is no worse than stipulations made on competing views of Homogeneity. Covert pluralization operators and overt floating quantifiers cannot be taken to be semantically identical quite independently of one's view of Homogeneity, since in the vast majority of cases floating quantifiers remove Homogeneity.[30] On accounts where covert pluralization is responsible for Homogeneity (Schwarzschild; Gajewski), the difference can be stated in terms of the presence or absence of an excluded-middle presupposition in their semantics; on other accounts different stipulations are needed in order to set floating quantifiers apart from covert pluralization operators. Stipulating that the difference is in quantificational force does not strike me as more ad hoc than any other stipulation.[31]

Even if one accepts that the covert pluralization operator could be existential, one might still expect that we would be able to substitute it for overt floating quantifiers like *each* and *all* when we derive alternatives, based on the common assumption that they occupy the same syntactic position. Under this assumption there seems to be no reason to think the structure of *the kids each/all laughed* should be more complex than that of *the kids* ∃-*pl*$_D$ *laughed* (recall that complexity was the reason given above for why *each/all the kids laughed* is not an alternative of *the kids laughed*). In other words, the implicature account presented here is incompatible with the assumption that *each* and *all* are overt counterparts of the covert pluralization operator. I would like to point out, however, that it is not outlandish to assume that floating quantifiers aren't scalemates of ∃-pl. On one prominent view of their syntax (due to Sportiche 1988), floating quantifiers originate within DP; this would set them apart from ∃-pl which is a modifier of VP. On this view, the underlying structure of *the kids each/all laughed* is the same as that of *each/all of the kids laughed*, and under this assumption the former sentences can be ruled out as alternatives, just like the latter, on grounds of complexity.[32] We are then led to think of the difference between floating quantifiers

[30] The only exception to this generalization I'm aware of is Russian *po* (Križ 2017), a floating quantifier that maintains Homogeneity. An analysis of *po* one might consider based on the implicature view is that, just like ∃-pl, it is an existential quantifier with no universal counterparts. I do not aim to provide here a satisfactory analysis of *po*; further research is needed in order to determine whether such a direction could account for all the curious properties of *po* discussed by Križ.

[31] A reviewer further points out that the assumption that Homogeneity depends on the presence of a covert pluralization operator is at odds with the claim that some languages do not allow distributive readings with predicates which contain indefinites or numerals (e.g., *earn exactly 100 euros*, see Flor et al. 2017) in the absence of overt quantificational operators; this would seem surprising on the current view given that these languages aren't exempt from Homogeneity effects. First, I would like to point out that this concern would apply equally well to the standard view of Homogeneity due to Schwarzschild (1994); Gajewski (2005), which shares the assumption that covert pluralization is responsible for Homogeneity. Second, even English covert pluralization doesn't allow distributive readings for such predicates very easily (see Dotlačil 2010, §2; Champollion to appear, §2.3 and references therein), a fact which is not well understood. A way to reconcile such facts with obligatory covert pluralization, following Schwarzschild (1994), is to stipulate that indefinites and numerals prefer to outscope covert pluralization (perhaps this preference is stronger in the languages discussed by Flor et al.). See however some issues with such a view in Dotlačil (2010, §3.4.1). I would have to leave a more in depth discussion of these important issues to another occasion.

[32] Further complication of our commitment to the absence of conjunctive alternatives comes from Homogeneity removal by quantifiers like *all* which, as we will discuss in §8.3, can be analyzed by assuming

and covert pluralization as follows: While floating quantifiers are syntactically part of the subject of the predication, ∃-PL is how (plural) predication is carried out (this is again following the view in Schwarzschild 1994; Kratzer 2007 according to which the pluralization operator obligatorily applies to the sister of plural DPs).

It would undoubtedly be desirable to find some independent evidence both for an existential pluralization operator and for syntactic differences between covert pluralization and floating quantification; unfortunately, I can offer no such conclusive evidence at this point. One source of evidence for a covert existential pluralization operator can perhaps be found in recent work by Kobayashi and Rouillard (2019) who utilize an existential pluralization operator to explain some puzzling differences between answers to questions with singular and with plural *wh*-phrases.

4.6 Summary of the proposal

By modeling the Implicature account of Homogeneity after that of Free Choice we've been able to account for their similar behavior. As I have pointed out, the emerging view is a mashup of the general architecture of Magri's (2014) Implicature account, and the reasoning about alternatives of the form proposed by Malamud (2012). In the next section I propose an extension of the proposal to cover Non-maximality, pointing out that by choosing to derive parallel alternatives to those of Free Choice disjunction, we are able to provide a similar analysis to Malamud (2012); Križ and Spector (2017) for Non-maximality. However, I will argue that the current approach is advantageous since it provides a more natural explanation for Non-maximality as following from the context-sensitivity of implicature calculation, and since it predicts asymmetries between positive and negative sentences with respect to Non-maximality.

5 Non-maximality

Recall that Non-maximality is context dependent, as is evident from the availability of a non-maximal reading in (56) in contrast with (57) (repeated from (3) and (4), respectively):

(56) a. *Context: There was a clown at my kid's birthday party. Someone asks me if they gave a funny performance. I reply:*
b. The kids laughed. [=(3)]

(57) a. *Context: Someone asks me who among the five adults and the ten kids at the birthday party laughed. I reply:*
b. The kids laughed. [=(4)]

a universal counterpart of ∃-PL. I will entertain there the possibility that such an operator exists after all but is only licensed by quantificational DPs, which is why replacing ∃-PL in *the kids laughed* with it would only be able to derive a licit LF if we at the same time replace *the kids* with the more complex *each/all of the kids* (or other quantifiers, which are at least semantically more complex, see fn. 26). The general picture, i.e., the claim that conjunctive alternatives are not derivable without moving to more complex structures, will still be maintained.

Non-maximality poses a problem for accounts in which the truth of *the kids laughed* entails the truth of *Kelly laughed*. This is evidently a problem for the Trivalence account in which this entailment holds, and it requires positing extra machinery in order to weaken the meaning given a context (see Križ 2015, 2016 for an ingenious proposal). On the Ambiguity account the situation is on the face of it a little better: *Kelly laughed* does not follow from all the ambiguity resolutions of *the kids laughed*. On the Implicature account similarly *Kelly laughed* only follows from *the kids laughed* as an implicature; the meaning before exhaustification does not entail it.[33] Some mechanism is still needed, for both the Ambiguity and the Implicature approaches, by which the only meaning derived in contexts like (57a) will be maximal, while in contexts like (56a) a non-maximal reading will be derived, but one which is still stronger than *some of the kids laughed* (recall the objection made on this basis to Magri's view in §2.3). Mechanisms of this sort have been proposed by Malamud (2012) and following her by Križ and Spector (2017). The goal of this section is to provide such a mechanism which follows their lead but is couched within the theory of Homogeneity proposed in the previous section, and in which as hypothesized in §2.2 Non-maximality results from not deriving the Maximality implicature.

I have noted in the previous section that the current account involves reasoning about a similar set of propositions to those utilized by Malamud (2012); Križ and Spector (2017). Malamud's insight was that reasoning about such propositions paves the way for deriving non-maximal readings. Building mainly on the formulation of the approach to Non-maximality in Križ and Spector (2017) which is based on Malamud's insight (but which solves some problems her account faces), I will argue that the current account predicts such readings to arise given what we know independently about implicature calculation, which is that some of the alternatives may be ignored given certain contexts. I will first show in §5.1 that by ignoring some of the alternatives (which I will call *pruning*) we generate for computing the Maximality implicature we can derive a set of potential readings similar to Križ and Spector's. In §5.2 the context dependency of non-maximal readings will be reduced to the context-sensitivity of implicature calculation, specifically to constraints on pruning given a context. The main advantage of the Implicature account over both the Trivalence and the Ambiguity accounts with respect to Non-maximality, namely the marginality of non-maximal readings in negative sentences (§2.2), is taken up in §5.3.

5.1 How to derive non-maximal readings

A natural way to think about Non-maximality in the current framework for Homogeneity is as following from the context-sensitivity of implicature calculation. It is known at least since Horn (1972) that whether an implicature is derived or not is dependent on the context, and contextual factors have been argued to play a crucial role in restricting possible choices of pruning (i.e., ignoring some alternatives in the original set of alternatives; see, e.g., Magri 2009; Fox and Katzir 2011; Katzir 2014; Crnič et al. 2015; Singh et al. 2016). Reducing the context-dependency of Non-maximality

[33] The Ambiguity and Implicature accounts are on equal footing here essentially because the Implicature account is a sub-species of the Ambiguity account, see fn. 4.

to constraints on such choices given a context is then desirable.[34] Before discussing how the context in fact constrains possible choices of pruning, let us demonstrate that having the ability to prune, i.e., to substitute the whole set of alternatives for a certain subset, we can derive non-maximal readings. I return to the question of how the context affects the availability of these readings in §5.2.

In order to simplify the discussion, let us use a toy example where there are 3 kids at the birthday party: Kelly, Jane and Bill. Our goal now is to show that by pruning we can derive a reading where *the kids laughed* ends up equivalent to *at least two of the kids laughed*.[35] The basic meaning of *the kids laughed*, given that the kids are Kelly, Jane and Bill, is the disjunction in (58), and *SubAlt* is the set of all of its disjuncts:

(58) *Prej* = Kelly laughed ∨ Jane laughed ∨ Bill laughed

(59) *SubAlt* = {Kelly laughed ∨ Jane laughed ∨ Bill laughed, [= *Prej*]
Kelly laughed ∨ Jane laughed,
Kelly laughed ∨ Bill laughed,
Jane laughed ∨ Bill laughed,
Kelly laughed,
Jane laughed,
Bill laughed}

As we have seen, by applying exhaustification we will derive the conjunction of all of the alternatives in *SubAlt*, i.e., the maximal reading according to which all the kids laughed.

(60) $[\![\mathcal{E}xh]\!](SubAlt)((58)) = \bigwedge SubAlt =$
Kelly laughed ∧ Jane laughed ∧ Bill laughed

Suppose now that we choose to prune all the alternatives which are built of only one atomic disjunct (e.g., *Kelly laughed*), and have the following subset of *SubAlt* as our set of alternatives instead:

(61) *SubAlt′* = {Kelly laughed ∨ Jane laughed ∨ Bill laughed, [= *Prej*]
Kelly laughed ∨ Jane laughed,
Kelly laughed ∨ Bill laughed,
Jane laughed ∨ Bill laughed}

As the reader may verify, choosing this set of alternatives does not make any of the alternatives IE. All the alternatives are rather II, hence all of them are assigned true by applying $\mathcal{E}xh$. However, their conjunction is weaker than *all the kids laughed*, and is in fact equivalent to *at least two of the kids laughed*.[36]

[34] I should note that Malamud (2012) sets such reduction as one of her goals. However, the connection between a general theory of pruning and her analysis of Non-maximality isn't made transparent.

[35] In reality, when the number of kids is 3 examples where we understand *the kids laughed* non-maximally are hard to come by and perhaps non-existent. Presumably, this is since it's difficult to conceive of the difference between 2 kids laughing and 3 kids laughing as irrelevant, whereas it's easier to think of the difference between say 29 kids laughing and 30 kids laughing as irrelevant.

[36] This is not a quirk of having 3 kids in the context. Given any set of kids $\{x_1, \ldots, x_n\}$ we'd have for *the kids laughed* the basic disjunctive meaning (x_1 laughed ∨ … ∨ x_n laughed). For any j, $1 \leq j \leq n$, *at*

(62) $[\![\mathcal{E}xh]\!](SubAlt')((58)) = \bigwedge SubAlt' =$
(Kelly laughed ∨ Jane laughed) ∧ (Kelly laughed ∨ Bill laughed) ∧ (Jane laughed ∨ Bill laughed) =
(Kelly laughed ∧ Jane laughed) ∨ (Kelly laughed ∧ Bill laughed) ∨ (Jane laughed ∧ Bill laughed)

We've seen then that we can derive two different readings by making different choices about the set of alternatives. By exhaustifying over the whole set we derive the maximal reading *all the kids laughed*, which is what is desired for the context in (57a), and by exhaustifying over its subset in (61) instead we derive a non-maximal reading equivalent to *at least two of the kids laughed*, which is what's desired for the context in (56a). Of course, there are many other readings derivable by choosing different subsets of the whole set of alternatives (the number grows the more kids there are in the context). Some of the readings we can in principle derive by choosing subsets of the set of alternatives are however never attested; those are readings derived by making some of the alternatives IE. For example, suppose we choose to prune all alternatives except *Kelly laughed*. This alternative would then end up IE, and the result of exhaustification would be *(Jane laughed* ∨ *Bill laughed)* ∧ ¬ *Kelly laughed*. But there is no context in which the meaning of *the kids laughed* entails that one of them didn't: it may be *compatible* with some of them not laughing (Non-maximality), but there is no evidence that it can be *entailed* by it.[37]

Such pathological choices of pruning luckily have to be ruled out for independent reasons. These choices involve breaking of the symmetry (*Generalized Symmetry* as Katzir 2014 refers to it) which holds between the disjunctive alternatives of disjunction, which is known to never actually happen (see Fox and Katzir 2011; Katzir 2014; Crnič et al. 2015; Singh et al. 2016). In other words, the fact that *Kelly or Jane laughed* can never mean that Kelly laughed but Jane didn't requires ruling out this kind of pruning. I will follow Crnič et al.'s (2015) proposal according to which pruning can only lead to weakening ($Alt(S)$ is the complete set of alternatives of S):[38]

(63) **Constraint on pruning** (1st version):
$\mathcal{E}xh_C$ S is licensed for $C \subseteq Alt(S)$ only if $[\![\mathcal{E}xh_{Alt(S)}\ \mathrm{S}]\!] \Rightarrow [\![\mathcal{E}xh_C\ \mathrm{S}]\!]$.

Once we have the weakening constraint in place we correctly rule out any pruning which would lead to some of the subdomain alternatives being IE, since no proposition which entails that some of the kids didn't laugh could be entailed by the proposition that all the kids laughed (which is the result of having no pruning). The result is that we can only derive readings which are conjunctions of members of *SubAlt*, if such

least j kids laughed is equivalent to the conjunction of all alternatives built of $n+1-j$ atomic disjuncts; hence exhaustifying over a set of alternatives containing these alternatives but no alternative stronger than any of them will yield the meaning *at least j kids laughed*.

[37] One can never object to the claim that the kids laughed by asserting that one of them did:

(i) A: The kids laughed.
B: #That's not true, Kelly did!

[38] See their paper for a possible intuition behind this idea. Crnič et al.'s constraint (their ex. (37)) is stronger than (63) for reasons which are irrelevant for our purposes.

conjunctions result from a licit pruning choice.[39] As it turns out, any proposition that could be derived by conjoining at least two members of *SubAlt* which do not stand in an entailment relation to each other is derivable with licit pruning. However, there turn out to be no licit pruning choices which lead to readings equivalent to members of *SubAlt*, with the sole exception being readings equivalent to the prejacent (which may result from pruning all the alternatives). Therefore the set of derivable *Readings* of *the kids laughed* in our system is the closure under conjunction of *SubAlt*, barring all members of *SubAlt* except the prejacent:[40]

(64) **Derivable readings of *the kids laughed***:
$\mathit{Readings} = \{p : p \in \mathit{SubAlt}^{\wedge} \setminus (\mathit{SubAlt} \setminus \{\mathit{Prej}\})\}$

5.2 Restricting non-maximal readings: a novel view of pruning

Note that the weakening constraint would not help us match between possible readings and contexts, of which there is no mention in its definition.[41] What we are still missing are ways to restrict these readings to only those contexts in which we actually find them. If we allow free pruning, we will derive non-maximal readings even in contexts where only a maximal reading is attested. The question is essentially what's responsible for the context-sensitivity of implicature calculation, or, in other words: How does the context restrict possible pruning choices? A possible view is that pruning is a contextual restriction of the set of formal alternatives to only those that are relevant in the context (e.g., Fox and Katzir 2011; Katzir 2014). I will first discuss why this view does not suffice in order to derive the correct Non-maximality facts, and suggest a shift in perspective on pruning in order to fix the problem.

As standard, I take the set of relevant propositions to be dependent on a contextually supplied partition Q of the set of possible worlds.

[39] The reader may verify that given that no licit pruning choice can lead to IE alternatives, any licit pruning choice would lead to Inclusion of all the alternatives in *SubAlt* which remain after pruning.

[40] Assuming that the kids are Kelly, Jane and Bill, and that a = Kelly laughed, b = Jane laughed, and c = Bill laughed, we get 12 distinct members in the set *Readings*:

(i) a. $\mathit{Readings} = \{a \wedge b \wedge c, a \wedge b, a \wedge c, b \wedge c, a \wedge (b \vee c), b \wedge (a \vee c), c \wedge (a \vee b), (a \vee b) \wedge (a \vee c) \wedge (b \vee c), (a \vee b) \wedge (a \vee c), (a \vee b) \wedge (b \vee c), (a \vee c) \wedge (b \vee c), a \vee b \vee c\}$
b. Members of $\mathit{SubAlt}^{\wedge}$ which are not in *Readings*: $\{a, b, c, a \vee b, a \vee c, b \vee c\}$

As far as distributive predication is concerned, the set of possible readings we get turns out almost identical to Križ and Spector's 'candidate meanings', which is simply $\mathit{SubAlt}^{\wedge}$ (the alternative view of pruning proposed in Bar-Lev 2018, §3.A makes identical predictions to theirs). It is indeed impossible to find cases where *the kids laughed* has a non-maximal reading equivalent to a member of *SubAlt*, for instance where it means that one specific kid, say Kelly, laughed. While other accounts (e.g., Križ 2015, 2016; Križ and Spector 2017) do predict such readings to be possible, it is not clear that this can be taken as an advantage of the current system: *the kids laughed* is still predicted to have a meaning according to which two specific kids, say Kelly and Jane, laughed, even when there are 20 kids. Such readings seem equally unavailable, and a more principled explanation for why they are is called for.

[41] See Katzir (2014) for an alternative way to ensure that context does not break generalized symmetry, which does relate to the context; Bar-Lev (2018, §3.A) develops an alternative to what I'm about to propose which is based on Katzir's proposal.

(65) **Relevance of propositions given Q**:
A proposition p is relevant to a partition Q *iff* $\exists Q' \subseteq Q[p = \bigcup Q']$

I assume following Malamud (2012); Križ (2015, 2016); Križ and Spector (2017) that what's responsible for the availability of non-maximal readings is the identity of Q (what Malamud 2012 calls the Question Under Discussion and Križ 2015 calls the Issue). Let us take the Q in the context of (56) to be the following partition (and recall that this is to be taken as a toy example for illustration purposes only, see fn. 35):

(66) $$Q = \begin{cases} i_1 & \text{At most one kid laughed (= unfunny clown),} \\ i_2 & \text{Two or three kids laughed (= funny clown)} \end{cases}$$

We would like all alternatives built of two atomic disjuncts (e.g., *Kelly laughed* ∨ *Jane laughed*) to be unprunable: as we have seen in (62), their conjunction is what's responsible for the desired meaning for this context, and pruning them (together with the alternatives built of one atomic disjunct) would yield a meaning which is too weak and irrelevant, i.e., *at least one kid laughed* which identifies no union of cells in Q. Unfortunately, none of the two-disjuncts alternatives is relevant: there is no subset of Q which is equivalent to the proposition *Kelly laughed* ∨ *Jane laughed*, for instance. So given the view we have been entertaining we'd expect all of these alternatives to be prunable, even though their presence in the set of alternatives is required to derive a relevant output: given the set of alternatives *SubAlt′* we derive a meaning equivalent to *at least two kids laughed*, i.e, a proposition which identifies cell i_2.

I hence suggest a shift in perspective: to decide what prunings are licit we don't just look at each alternative in isolation and ask whether it's relevant, but rather look at different pruning choices and ask whether they yield relevant propositions when we apply $\mathcal{E}xh$. This allows us to restrict pruning to a minimum that's necessary for deriving a relevant meaning. Combining this idea with Crnič et al.'s (2015) constraint as in (67), pruning turns out to be a way to identify the union of cells in Q which is 'closest' to the (possibly irrelevant) proposition identified by exhaustification over the whole set of alternatives, i.e., the smallest union of cells entailed by that proposition (if we can indeed identify that union of cells by pruning).[42]

(67) **Constraint on pruning** (2nd and final version):
$\mathcal{E}xh_C$ S is licensed for $C \subseteq Alt(S)$ given a contextually provided partition Q only if C is a maximal subset of $Alt(S)$, s.t.
a. $[\![\mathcal{E}xh_C \text{ S}]\!]$ is relevant to Q, and
b. $[\![\mathcal{E}xh_{Alt(S)} \text{ S}]\!] \Rightarrow [\![\mathcal{E}xh_C \text{ S}]\!]$ (following Crnič et al. 2015)

As a result of this constraint, given Q in (66) we can prune the atomic alternatives *Kelly laughed*, *Jane laughed* and *Bill laughed* from the set of alternatives, namely use *SubAlt′* instead of *SubAlt*, since a relevant meaning (*at least two of the kids laughed*) is derived given *SubAlt′*, and no enlargement of this set within *SubAlt* yields a relevant meaning. However, we will not be able to prune any other alternatives, namely use any proper subset of *SubAlt′*, since such a set will no longer be a maximal set in *SubAlt*

[42] As such this constraint closely resembles the idea behind Križ's (2016) account of Non-maximality.

which yields a relevant meaning. Hence *at least two of the kids laughed* would be the weakest meaning we could derive for *the kids laughed* in that context, as desired. As for the context in (57), we would not be able to prune anything, since the question made salient in this context (*who laughed?*) contains a cell equivalent to *all the kids laughed*. Since by exhaustifying over the whole set of alternatives *SubAlt* we derive this relevant meaning, no pruning will be allowed and as a result the only reading will be the maximal one.[43]

We can formulate this view of pruning in the following way (thanks to suggestions by Kai von Fintel and Danny Fox):

(68) a. **Maxim of Relevance**: Every utterance must be relevant to Q.
b. **Weakening**: Pruning can only weaken the meaning. (Crnič et al. 2015)
c. **Minimal pruning**: Don't prune more than necessary to satisfy (68a).

The resulting theory of pruning can be portrayed as a way to find an equilibrium between the notions of Informativity and Relevance: We can prune in order to weaken the meaning and get a relevant proposition, but we must choose the most informative such proposition. Constraining pruning in such a way might, furthermore, be seen as natural given the view in Fox (2017) that $\mathcal{E}xh$ is a formal device the goal of which is to bridge the gap between questions qua sets of propositions (not necessaily partitions) and questions qua partitions: given a set of propositions $\mathcal{E}xh$ can identify cells in the question-partition. Pruning is needed in a situation where given the set of propositions we derive as formal alternatives to S (i.e., $Alt(S)$), applying $\mathcal{E}xh$ to S does not help us identify a cell in the contextually provided partition (Q). Pruning is then a way to let $\mathcal{E}xh$ identify a cell in Q nonetheless. If this is indeed the reason for allowing pruning, it is reasonable to only allow the minimum pruning necessary to achieve this goal, i.e., to identify a cell in the question-partition Q.

Our predictions at this point are as follows:

(69) p is a possible reading of *the kids laughed* given a contextually provided partition Q *iff* p is the strongest among the members of *Readings* in (64) relevant to Q.[44,45]

This characterization is closely connected to Malamud's and Križ and Spector's, and yields almost identical predictions for non-maximal readings with distributive

[43] For the context in (57) in which the question made salient is *who laughed?*, it would suffice to ask for each alternative in isolation whether it's relevant, namely the view initially entertained here would be enough and one would not need the more cumbersome constraint on pruning proposed here. This is however not in general a property of contexts forcing a maximal reading. If the salient question is *did all the kids laugh?*, a maximal reading will be the only one possible for *the kids laughed* but propositions built of one atomic disjunct (e.g., *Kelly laughed*) will not be relevant in isolation.

[44] There can be at most one such proposition given Q: Since Relevance as well as *Readings* are closed under conjunction, the conjunction of all relevant members of *Readings* is guaranteed to be *the* maximally strong proposition among the members of *Readings* which are relevant to Q.

[45] The constraint on pruning in (67) guarantees that the largest set in $Alt(S)$ which leads to a relevant meaning will be picked out. The reason why this also makes sure that the set which leads to the strongest meaning among the relevant propositions in *Readings* will be picked out is that the following holds for any two subsets of *SubAlt*, C' and C'', which are licit choices of pruning given Crnič et al.'s constraint: if $C' \supset C''$ then $[\![\mathcal{E}xh_{C'}$ the kids $\exists$-PL$_D$ laughed$]\!] \Rightarrow [\![\mathcal{E}xh_{C''}$ the kids $\exists$-PL$_D$ laughed$]\!]$.

predicates in positive sentences (for the differences see fn. 40). One advantage of the current view of Non-maximality is that the only novelty it involves is restricting pruning to a minimum that is necessary to derive a relevant meaning (which is arguably motivated given $\mathcal{E}xh$'s goal, see above); otherwise it posits no machinery which isn't needed elsewhere. A more substantive difference is with respect to Non-maximality in negative sentences, to which we now turn.

5.3 Accounting for asymmetry

Recall that I have argued in §2.2 for an asymmetry between Non-maximality in positive and negative sentences: while non-maximal readings in positive sentences are entirely natural, in negative sentences they are more difficult to get, especially when the scope of the definite plural is fixed below negation. I have exemplified the point using (10), repeated here:

(70) *Context: the kids are required to take at least 2 of the 4 vitamins I gave each of them.* [=(10)]

a. All of the kids took their vitamins.
≈ Every kid took at least 2 vitamins. (non-maximal)

b. None of the kids took their vitamins.
≈ No kid took any vitamin (i.e., not even one vitamin). (only maximal)

The route to Non-maximality developed above indeed has an asymmetric nature: The derivation of non-maximal readings via this route relies on the ability to derive an implicature which is weaker than the Maximality implicature we derive using the whole set of alternatives. This reasoning obviously only applies to cases where there is an implicature to be derived; but within the current account Homogeneity for truly negative sentences follows directly from the basic semantics rather than as an implicature.[46]

If pruning was the only way to derive Non-maximality, we would in fact expect non-maximal readings to be entirely impossible for (70b). Recall however that this would not be a good prediction: As I have mentioned in §2.2, (70b) is not obviously false or misleading if we find one kid who took one vitamin. In the remaining of this section I will hence develop the following nuanced view: there is a way to derive Non-maximality in (70b), but the conditions under which it is available are more constrained than those of the implicature-based route to Non-maximality which are responsible for the non-maximal readings of (70a); this is why we tend to understand (70b) maximally, but this is merely a tendency and not a categoric judgment.

[46] Recall that for simple negative sentences which involve no fixing of the scope of the definite plural below negation (e.g., with a bound variable, as in (70b)) we would still expect non-maximal readings due to the possibility of negation taking narrow scope. Indeed we have seen that non-maximal readings for (9), repeated here, are readily available to some speakers.

(i) a. *Context: There was a clown at my kid's birthday party. Someone asks me if they gave a funny performance. I reply:*

b. The kids didn't laugh. [=(9)]

What other routes to Non-maximality could there be?[47] Brisson (1998, 2003) proposed to derive Non-maximality by relying on the theory of covers due to Schwarzschild (1996). While one might be reluctant to rely on Brisson's own mechanism (due to the central role of 'ill-fitting covers' in it), I would like to employ another way to utilize covers for deriving non-maximal readings which falls out of the theory of covers (without relying on 'ill-fitting' ones).[48] Suppose that, before applying ∃-PL, the predicate is restricted by a cover variable which is assigned a set of contextually relevant pluralities (this is in line with the implementations of the theory of covers in Heim 1994; Beck 2001, which I will adopt in §8.1).[49] The meaning of (70b) would then be, roughly, *none of the kids took any contextually relevant plurality of their vitamins*. Suppose pluralities consisting of only one vitamin are irrelevant and bigger pluralities are; in this case the meaning we get will be *none of the kids took two or more of their vitamins*. This is a non-maximal reading compatible with there being kids who take one vitamin. Note that (unlike Brisson's mechanism) this route to Non-maximality is unavailable for positive sentences with distributive predicates. For *the kids laughed*, as long as every atomic kid is part of some plurality of kids which is in the cover, the exhaustified meaning (with no pruning) would entail that all the kids laughed (since if P is a distributive predicate and it is true of some individual x is part of, then P is true of x).[50]

We have then two routes to Non-maximality at our disposal: one provided by pruning, which can only apply in positive sentences, and another provided by covers, which can only apply in negative sentences. The asymmetry between positive and negative sentences I have argued for in §2.2 can then be explained as follows: It is easy to avoid the derivation of implicatures which are irrelevant, since the strength of the implicature derived is directly determined by the QUD; hence very little contextual manipulation is needed to derive Non-maximality in positive sentences. In contrast, it is difficult to override a default cover, one which contains all the atomic parts of the subject of predication, and replace it with one which doesn't.[51] So Non-maximality which follows from restriction by a cover (ignoring some pluralities) requires more

[47] One might consider deriving Non-maximality by letting *the kids* refer to a plurality which doesn't contain all the kids. This idea has however been forcefully and convincingly attacked by Križ (2015, §3.1.5), and I will hence not assume this plays any relevant role.

[48] The importance of covers for a theory of Homogeneity is elaborated on in Bar-Lev (2018, §4–5), Bar-Lev (2019). See also §8.1 below.

[49] Note that for this to work we'll have to get rid of our assumption that we only consider atomic pluralities in the definition of ∃-PL, which we'll do in §8.1.

[50] One might wonder whether applying $\mathcal{E}xh$ under negation could provide us with yet another route to Non-maximality in negative sentences, given that it has been argued to apply under negation under certain circumstances (see Magri 2009; Fox and Spector 2018). I will stick to a simplified view according to which it never does, because I do not think the examples discussed in this section which show clear non-maximal readings under negation (especially (72)) have the hallmarks of cases involving embedded $\mathcal{E}xh$ under negation discussed by Fox and Spector; however, it's possible that an LF with embedded $\mathcal{E}xh$ is available but is extremely dispreferred, and its availability contributes to obscuring the asymmetry between positive and negative sentences.

[51] Manipulating covers has indeed been shown to be difficult (see the debate in Gillon 1987, 1990; Lasersohn 1989 about the limited availability of so-called intermediate readings which require non-default covers), though little is known about what determines the choice of a cover. Full discussion of this issue is outside the scope of this paper.

contextual manipulation than Non-maximality which follows from restricting the domain of the exhaustivity operator (pruning, i.e., ignoring some propositions).

The fact that (70b) prefers a maximal reading is then explained as the result of taking pluralities of fewer than 2 vitamins to be relevant by default, under the reasonable assumption that taking one vitamin is better than taking none. The fact that this assumption may be up for debate is arguably what makes a non-maximal reading for (70b) dispreferred but not entirely impossible. A prediction that this view makes is that with enough contextual manipulation we may be able to find natural occurrences of non-maximal readings under negation. Specifically, if we make small pluralities entirely irrelevant, such readings are predicted to arise. This seems to be borne out when we consider cases like (71) and (72) (I thank Benjamin Spector for pointing out to me the existence and relevance of this kind of examples). In both (71b) and (72b), small pluralities (of less than 4 poisoned peas or 20 pies) are almost explicitly made irrelevant by the contexts: eating less than 4 poisoned peas is of no consequence, as well as eating less than 20 pies. While I'm not entirely sure that (71b) has absolutely no inference that no one ate any poisoned peas (which is perhaps because when poisoned peas are involved we might still want to be on the safe side and consider each pea relevant), the judgment reported for (72b) seems unobjectionable: There indeed seems to be no inference in this case that no contestant ate any of the pies.

(71) a. *Context: Each of the guests in the party was served a plate containing 10 poisoned peas. Eating 3 of those peas is entirely harmless, but eating 4 or more leads to certain death.*

b. None of the guests ate their poisoned peas.
≈ No guest ate more than 3 poisoned peas. (Benjamin Spector, p.c.)

(72) a. *Context: Each contestant in a pie eating contest is provided with 25 pies. In order to move on to the next round a contestant must eat at least 20 pies (they get more points for eating more).*

b. No contestant ate their pies.
≈ No contestant ate at least 20 pies. (Emily Clem, p.c.)

5.4 Interim summary

We have seen that the Implicature account of Homogeneity proposed in §4 predicts non-maximal readings to arise as the result of pruning. To capture the desired interaction of Non-maximality with different contexts, I proposed a novel view of pruning where we cannot prune more alternatives than necessary to achieve a relevant meaning. On top of utilizing independently needed machinery to account for Non-maximality, the current account has the advantage of predicting the tendency of non-maximal readings to be unavailable for definite plurals embedded under negation, and it can explain cases where they do seem to be available based on the theory of covers. In the next section I discuss a different phenomenon related to Homogeneity and Non-maximality, namely 'gappy' (neither-true-nor-false) judgments.

6 Notes on Gappiness

6.1 Gappiness as a side effect of Non-maximality

Much of the appeal of the Trivalence account comes from the characterization of judgments in (73) (see Löbner 1987, 2000), which I call Gappiness. A sentence like *Mary read the books* is judged neither-true-nor-false in a 'non-homogeneous' situation where Mary read some but not all of the books (I defer the discussion of the parallel judgment for the negative sentence *Mary didn't read the books* to §6.2). If trivalence is indeed involved, the gappy judgment can be seen as reflecting a presupposition failure.

(73) Mary read the books.
- a. **True** if Mary read *all* of the books.
- b. **False** if Mary read *none* of the books.
- c. **Neither-true-nor-false** if she read *some but not all* the books.

I would like to argue however that such judgments have nothing to do with trivalence. These judgments might be a good reason to believe that trivalence is involved only insofar as we can be convinced that gappiness judgments are stable across different triggers of Homogeneity. If we could find cases where there is Homogeneity but no gappiness, the case for a Trivalence account from Gappiness will be undermined. And indeed we find such cases, showing that Gappiness is dependent on Homogeneity but not the other way around. Evidence for this claim comes from expressions like *the 4 books* and *Anna Karenina and War and Peace*, which give rise to Homogeneity as in (74) and (75), but no Gappiness seems to be involved, as in (76) and (77).

(74)
- a. Mary read the 4 books.
  ≈ Mary read *all* of the 4 books.
- b. Mary didn't read the 4 books.
  ≈ Mary didn't read *any* of the 4 books.

(75)
- a. Mary read Anna Karenina and War and Peace.
  ≈ Mary read *both* Anna Karenina and War and Peace.
- b. Mary didn't read Anna Karenina and War and Peace.
  ≈ Mary didn't read *either* of Anna Karenina and War and Peace.

(76) Mary read the 4 books.
- a. **True** if Mary read *all* of the 4 books.
- b. **False** otherwise.

(77) Sue read Anna Karenina and War and Peace.
- a. **True** if Mary read *both* Anna Karenina and War and Peace.
- b. **False** otherwise.

Križ (2015, 2016) points out that such expressions show Homogeneity but don't give rise to non-maximal readings, but doesn't discuss the fact that they also do

	the books	*the 4 books*, *AK and W&P*	*all of the books*
Homogeneity	✓	✓	✗
Non-maximality	✓	✗	✗
Gappiness	✓	✗	✗

Table 2 Homogeneity, Non-maximality and Gappiness properties of different expressions.

not give rise to Gappiness like *the books*.[52] The fact that Gappiness disappears but Homogeneity doesn't makes the identification of Homogeneity with Gappiness (and hence trivalence) inadequate. If Gappiness is the result of having a trivalent system in the background, why did it disappear? And if the numeral, for example, miraculously made it disappear, how come we still have Homogeneity? Homogeneity and Gappiness can be told apart then, but as far as I can see the generalization is that Non-maximality and Gappiness go hand in hand (see table 2).

(78) a. Expressions that give rise to Non-maximality give rise to Gappiness (*the books*).
b. Expressions that don't give rise to Non-maximality don't give rise to Gappiness (*AK and W&P*, *the 4 books*).

I thus suggest the following hypothesis:

(79) **Hypothesis**: Gappiness is a side-effect of Non-maximality; 'Neither true nor false' is in fact 'could be true and could be false': could be true if the context justified a non-maximal reading, and could be false if it didn't.

[52] Why there are no non-maximal readings for *the 4 kids* and *War & Peace and Anna Karenina* remains a mystery; let me however sketch a direction for an analysis, following insights by Križ (2015). Suppose mentioning the names of the individuals involved or their number in *the 4 kids laughed* and *I read War & Peace and Anna Karenina* makes it very likely that we think of the fine-grained questions *how many of the kids laughed?* or *which books did I read?* as being under discussion. Given the analysis in §5, such questions are indeed expected to only give rise to maximal readings. It should be noted that unlike most facts discussed in this paper, the Homogeneity properties of those expressions are subject to crosslinguistic variation. Križ (2015, §3.3.5) presents data showing that definite plurals containing numerals in French behave 'non-homogeneously', differently from the English behavior described in (74) and from French definite plurals without numerals.

(i) Il arriva à l'heure au rendez-vous, mais il ne parla pas avec les trois étudiants. Il parla seulement avec l'un d'entre eux.
'He arrived on time, but he didn't speak with the three students. He only talked to one of them.'
(Križ 2015, p. 84, ex. (31a))

As Benjamin Spector pointed out to me, the French data may be seen as particularly problematic for the current approach, since we have in this case a default universal reading under negation which is surprising if the pluralization operator is existential. One response to this problem could be to assume that French numeralized definite plurals should be analyzed as quantificational, assuming a covert universal quantifier applies to the definite description such that it ends up meaning *all 4 books* (as suggested to me by Manuel Križ, p.c.). Another response could be that the presence of the numeral provides a locus for focus, which prompts exhaustification under negation (Fox and Spector 2018). A third possibility builds on an idea we will entertain in §8.3 according to which English has a universal pluralization operator which is only licensed by certain DPs. Perhaps a difference between French and English is whether definite plurals containing numerals license this operator or not. I admit that none of these responses is entirely satisfying, and leave it as an unresolved problem.

This hypothesis is compatible with all existing accounts of Non-maximality (including Trivalence-based ones, e.g., Križ 2015, 2016): the availability of non-maximal readings is uncontroversially context-dependent, or the way we put it (following Malamud 2012; Križ 2015, 2016; Križ and Spector 2017) it is dependent on a contextually provided partition Q. What underlies Gappiness, I propose, is uncertainty about the identity of Q. In such a situation we plausibly consider several potential Qs. Since given some choices for Q we will get a non-maximal reading which is true in the situation and given other choices we will only get a reading which is false in the situation, we will judge the sentence as neither true nor false: it will not have the same truth value given different resolutions of what Q is.

This view predicts that if the identity of Q isn't in question, there will be no Gappiness. Under the assumption that the identity of Q is subject to simple contextual manipulations it is predicted that in the following contexts where a specific Q is given there will be no Gappiness.

(80) *Context: If Mary reads any five of the ten books she is guaranteed to pass the test; if she reads less than five she is guaranteed to fail.*
Mary read the books
a. **True** if Mary read at least five of the ten books
b. **False** otherwise

(81) *Context: To pass the test, Mary has to read all of the ten books on the reading list; if she reads less than ten she is guaranteed to fail.*
Mary read the books
a. **True** if Mary read all of the ten books
b. **False** otherwise

Whether this is a good prediction or not is not entirely clear to me. Note in any case that the assumption that by providing a specific context we can resolve the identity of Q is not a trivial one (see Križ 2015), hence the hypothesis proposed here is not forced to wed that prediction.

6.2 Asymmetry with Gappiness

The idea that Gappiness stems from Non-maximality is further supported by experimental data from Križ and Chemla (2015). Križ and Chemla examined gappy judgments for definite plurals in various contexts. Their results show (though they do not explicitly discuss this) that in a negative sentence as in (82b) the rate of neither-true-nor-false judgments drops and that of false judgments increases relative to the positive case in (82a).

(82) a. The squares are red.
Neither true nor false if some but not all of the squares are red
b. The squares aren't red.
?? **Neither true nor false** if some but not all of the squares are red

	Every (83a)	*No* (83b)
Completely true	33.6%	3.6%
Completely false	20.9%	75.0%
Neither	45.5%	21.4%

Table 3 Results of Križ and Chemla's (2015) experiment comparing gappy judgments for (83a) and (83b).

	Every (83a)	*No* (83a)
Non-maximal responses (Completely true+Neither)	**79.1%**	25.0%
Only-Maximal responses (Completely false)	20.9%	**75.0%**

Table 4 Interpretation of Križ and Chemla's (2015) results in table 3 according to the current account of Gappiness.

This reflects the slight asymmetry with regard to the acceptability of non-maximal readings in such cases discussed in §2.2; recall that some speakers do not accept non-maximal readings for negative sentences. Since Gappiness has been proposed above to be a side effect of Non-maximality, we predict this asymmetry to be more robust when an LF where disjunction takes scope below the definite plural is blocked, which as we have seen there significantly reduces the acceptability of non-maximal readings. And indeed, it is. Križ and Chemla investigated judgments for the following pair:[53]

(83) a. Every boy found his presents.
b. No boy found his presents.

The distribution of participants' responses when shown pictures of 'gappy' contexts (e.g., for (83a), where every boy found some of his presents but some boy didn't find all of his presents) is presented in table 3. First, note that 'neither' responses are much less common in the *no* case than in the *every* case. Moreover, 'completely true' responses are almost non-existent in the *no* case but relatively common in the *every* case, and 'completely false' judgements are the vast majority in the *no* case but a minority in the *every* case. This can be made sense of under the Implicature analysis combined with our view of Gappiness as resulting from Non-maximality: Since non-maximal readings are marginal under negation, so are the gaps in this environment. Under our analysis of Gappiness, the results could be seen as in table 4, lumping together 'completely true' and 'neither' judgments since both rely on the availability of non-maximal readings.[54] Seen this way, the contrast between the *no* case and the *every* case is striking: 75% only allow for maximal readings for the *no* case, while 79.1% allow for non-maximal ones for the *every* case. While this asymmetry is expected under the Implicature approach, it is surprising for other approaches.

To sum up, Križ and Chemla's findings serve both as one more argument in favor of an asymmetry between positive and negative sentences, and as an argument in favor of the view that Gappiness is a side effect of Non-maximality.

[53] The data reported here is taken from experiment C2 in Križ and Chemla (2015). I thank Manuel Križ for making the detailed (unpublished) data available to me.

[54] The difference presumably being whether they consider the possibility of having a Q for which the maximal reading is relevant ('neither') or not ('completely true').

7 On some differences between Free Choice and Homogeneity

As we mentioned at the end of §3, there are some differences between Free Choice and Homogeneity: Free Choice and Maximality inferences are not similarly cancellable, and Non-maximality does not seem to have a parallel with Free Choice. Now that we have an account of Homogeneity and Non-maximality, we can return to these issues. I wish to show that the picture is in fact more complex than might seem at first, and that the differences we find between the two phenomena can be reasonably explained despite the unification of Free Choice and Homogeneity under the current proposal.

One way in which they differ is in (84). We may conclude based on this example that Free Choice is cancellable and Homogeneity is not, or in other words that Maximality is an obligatory inference while Free Choice is not:

(84) a. #John read the books, I don't know which.
b. You are allowed to eat ice cream or cake, I don't know which.

Another way in which they differ points in the opposite direction. Since Non-maximality has no parallel in the case of Free Choice disjunction, it may seem as though Free Choice is an obligatory inference while Maximality is not. Further evidence for this comes from (85). Here we have the opposite cancellability pattern than in (84): The Maximality inference seems cancellable, (85a), and patterns differently from Free Choice, (85b), and like run-of-the-mill scalar implicatures, (85c).

(85) a. The kids (# all) laughed. In fact, Kelly and Bill did but Jane didn't.
b. #You are allowed to have ice cream or cake. In fact, you are allowed to have ice cream but not cake.
c. Some of the kids laughed. In fact, all of them did.

What should we make of this? Let me first dismiss the potential evidence from (84) and argue that it does not in fact involve an implicature cancellation and rather bears on a difference in scope possibilities between the two phenomena. I will then focus on proposing an explanation for the difference between Homogeneity and Free Choice with respect to Non-maximality and (85).

Bar-Lev and Fox (2017) have suggested that the Free Choice inference is in fact obligatory, assuming that $\mathcal{E}xh$ obligatorily applies at least at matrix position (following Magri 2009) and that the alternatives leading to the Free Choice inference cannot be pruned. What looks like the cancellation of the Free Choice inference in (84b) on this view involves in fact a different LF where disjunction takes wide scope above the existential modal (see Bar-Lev and Fox 2017, fn. 12); for such an LF the Implicature account predicts no FC inferences to begin with since its set of alternatives is closed under conjunction.[55] Since *John read the books* does not have any parallel way to avoid deriving the Homogeneity inference, we'd not expect it to behave the same way, hence the oddity of (84a) isn't surprising.

[55] Note that the availability of such an LF does not predict (85b) to be acceptable, since on this LF it would be expected to behave just like simple disjunction for which a similar cancellation is impossible:

(i) #Mary ate ice cream or cake. In fact, she ate ice cream but not cake.

What still requires an explanation then is the absence of a parallel to Non-maximality with Free Choice and the difference in cancellability between the two phenomena in (85). Both of these pieces of evidence suggest that Free Choice is obligatory while Homogeneity is not, which is surprising given the current account where the same mechanism underlies the two phenomena. Most strikingly, the assumption we have just mentioned that the alternatives leading to the Free Choice inference cannot be pruned is in stark contrast with what's at the heart of our account of Non-maximality, namely that the alternatives leading to the Homogeneity inference can be pruned.

However, there are independent reasons to expect such a contrast, due to the different nature of the alternatives in the two cases. Bar-Lev (2018, §2) proposed following Chemla and Bott (2014) that what's responsible for the obligatoriness of Free Choice inferences in contrast with run-of-the-mill scalar implicatures is the nature of the alternatives involved: while in the case of Free Choice those are alternatives derived by deleting parts of the prejacent (deletion alternatives), in other cases those are alternatives derived by substitution with other items from the lexicon (substitution alternatives). While substitution alternatives can be pruned, deletion alternatives cannot.[56] Note that subdomain alternatives are not deletion alternatives: we derive them by replacing the domain variable with other variables.[57] Hence they can be pruned, which opens the way for our account of Non-maximality. The paradigm in (85) where Homogeneity patterns with run-of-the-mill scalar implicatures and differently from Free Choice can be explained in the same fashion: the distinction between cancellable and uncancellable implicatures in (85) is along the lines of the distinction between inferences derived based on substitution alternatives and those derived based on deletion ones, namely between having the ability to prune and lacking such ability.

A similar response can be made to another concern about differences between Free Choice and Homogeneity. Given the tight connection between Free Choice and Homogeneity on our view, one may wonder why children's performance with Free Choice disjunction is adult-like (Zhou et al. 2013; Tieu et al. 2016), while as discussed in §2.1 their performance is non-adult-like with respect to Homogeneity (Tieu et al. 2015). Here too the culprit might be the nature of the alternatives: Zhou et al. (2013); Tieu et al. (2016) suggest that children's exceptional success with Free Choice (relative to other scalar implicatures) is due to acquiring access to deletion alternatives prior to substitution ones (as has been argued for independently by Gualmini et al. 2001; Barner and Bachrach 2010; Barner et al. 2011; Singh et al. 2016).

[56] One might entertain the possibility that this difference is due to the salience of deletion alternatives which are syntactically part of the prejacent, in contrast with substitution alternatives which are not.

[57] They do not clearly fall under substitution alternatives either, since they do not involve lexical replacements. Still, they involve substitution with elements which are not syntactically parts of the prejacent.

8 Beyond distributive predication

8.1 Homogeneity with non-distributive one-place predicates

Along with most of the literature on Homogeneity, I focused on Homogeneity with distributive predicates. The proposed account followed some trivalent accounts (Schwarzschild 1994; Gajewski 2005) in finding the culprit of Homogeneity in the pluralization operator. While for Gajewski (2005) Homogeneity results from pluralization due to positing a truth value gap in the definition of the distributivity operator as in (86), we captured it by a mechanism of strengthening an existential into a universal, to the effect in (87) (assuming no pruning and that D contains all parts of $[\![\text{DP}]\!]$).

(86) $$[\![\textsc{dist}]\!](P)(x)(w) = \begin{cases} 1 & \textit{iff} \quad \forall y \in Part_{AT}(x)[P(y)(w) = 1] \\ 0 & \textit{iff} \neg\exists y \in Part_{AT}(x)[P(y)(w) = 1] \\ \# & \text{otherwise} \end{cases}$$

(87) a. $[\![\text{DP } \textsc{dist} \text{ VP}]\!](w) = 1$ *iff* $[\![\mathcal{E}xh_C \text{ DP } \exists\text{-}\textsc{pl}_D \text{ VP}]\!](w) = 1$
b. $[\![\text{DP } \textsc{dist} \text{ VP}]\!](w) = 0$ *iff* $[\![\text{DP } \exists\text{-}\textsc{pl}_D \text{ VP}]\!](w) = 0$

Križ (2015) however observed that Homogeneity arises with some non-distributive predicates as well (but not with all of them, an issue we'll get back to towards the end of this section): both sentences in (88) aren't judged true if a plurality consisting of half of the kids lifted the piano together and no one else did.

(88) a. The kids lifted the piano.
b. The kids didn't lift the piano.

While Križ has taken such data as a reason to abandon the view due to Schwarzschild (1994); Gajewski (2005) according to which the culprit of Homogeneity lies in the pluralization operator, Bar-Lev (2019) argues that close scrutiny reveals that it rather supports a modification of that view, where instead of making the pluralization operator DIST responsible for Homogeneity we blame Link's (1983) $\star$ operator. Bar-Lev proposes a definition for $\star$ which can be stated as in (89).[58]

(89) $$[\![\star]\!](P)(x)(w) = \begin{cases} 1 & \textit{iff} \quad \forall y \sqsubseteq x[\exists z \sqsubseteq x[z \circ y \wedge P(z)(w) = 1]] \\ 0 & \textit{iff} \neg\exists y \sqsubseteq x[\exists z \sqsubseteq x[z \circ y \wedge P(z)(w) = 1]] \\ \# & \text{otherwise} \end{cases}$$

(where '$\circ$' is the overlap relation: $x \circ y$ *iff* $\exists z[z \sqsubseteq x \wedge z \sqsubseteq y]$)

I will not attempt to provide here a satisfactory discussion of Homogeneity with non-distributive predication, but rather only give the gist of a strategy for extending the proposal to cover Homogeneity with such predicates based on Bar-Lev (2019). This strategy is developed in Bar-Lev (2018, ch. 5), though implemented in a different (and more complex) way than what I will present here. The basic idea is as before: we

[58] See Champollion (2016a, ex. 32) for the equivalence between standard definitions of $\star$ and the truth conditions in (89).

want to derive what's usually attributed to the workings of the pluralization operator as a Maximality implicature. But instead of deriving the truth conditions standardly derived by applying DIST, we now aim to derive those standardly derived by applying $\star$. In other words, we'd like to derive the following:

(90) a. $[\![\text{DP} \star \text{ VP}]\!](w) = 1$ *iff* $[\![\mathcal{E}xh_C \text{ DP } \exists\text{-PL}_D \text{ VP}]\!](w) = 1$
b. $[\![\text{DP} \star \text{ VP}]\!](w) = 0$ *iff* $[\![\text{DP } \exists\text{-PL}_D \text{ VP}]\!](w) = 0$

In order to achieve this effect, we can redefine ∃-PL as an existential version of $\star$, as in (91a), instead of defining it as an existential version of DIST as we did in (31a).[59]

(91) a. $[\![\exists\text{-PL}]\!] = \lambda D_{et}.\lambda P_{\langle e,st\rangle}.\lambda x_e.\lambda w_s.$
$\exists y \in D \cap Part(x)[\exists z \in Part(x)[z \circ y \wedge P(z)(w) = 1]]$
b. $Part(x) = \{y : y \sqsubseteq x\}$ (i.e., the set of atomic and non-atomic parts of x)

Now let us see that this yields the right result when applied to the sentences in (88), starting with the negative sentence (88b).

(92) a. LF of (88b): ($\mathcal{E}xh_C$) NEG [[the kids] [$\exists\text{-PL}_D$ lifted the piano]]
b. $[\![(92a)]\!](w) = 1$ *iff* $\neg\exists y \in D \cap Part([\![\text{the kids}]\!])[y \text{ lifted the piano in } w]$
iff $[\![[\text{The kids}] [\star \text{ lifted the piano}]]\!](w) = 0$
'No plurality of kids lifted the piano'

Turning to the positive sentence (88a), we apply our strengthening mechanism as before. But now when we conjoin all the subdomain alternatives we get weaker truth conditions, which are identical to those standardly derived by applying $\star$.[60]

(93) a. LF of (88a): $\mathcal{E}xh_C$ [[the kids] [$\exists\text{-PL}_D$ lifted the piano]]
b. $[\![(93a)]\!](w) = 1$ *iff*
$\forall D' \subseteq D[D' \cap Part([\![\text{the kids}]\!]) \neq \varnothing \rightarrow \exists y \in D' \cap Part([\![\text{the kids}]\!])$

---

[59] Note that this move changes nothing if P is distributive, for the same reason that $[\![\star]\!](P)(x) \Leftrightarrow [\![\text{DIST}]\!](P)(x)$ whenever P is distributive. This means that *SubAlt* in (40b) remains the set of alternative propositions of *the kids* $\exists$-*PL*$_D$ *laughed* under the current modification. Taking into account non-distributive predication, the set of alternative propositions of a sentence of the form *DP* $\exists$-*PL*$_D$ *VP*, would be (again under the assumption that D contains all members of $Part([\![\text{DP}]\!])$):

(i) $\{p : \exists y \in Part([\![\text{DP}]\!])[p = \lambda w.\exists z \in Part([\![\text{DP}]\!])[z \circ y \wedge [\![\text{VP}]\!](z)(w) = 1]]\}$

Or, equivalently (as long as $Part([\![\text{DP}]\!])$ is the closure under $\sqcup$ of $Part_{AT}([\![\text{DP}]\!])$:

(ii) $\{p : \exists y \in Part_{AT}([\![\text{DP}]\!])[p = \lambda w.\exists z \in Part([\![\text{DP}]\!])[z \circ y \wedge [\![\text{VP}]\!](z)(w) = 1]]\}^{\vee}$

For example, here is the set of alternatives of *the kids lifted the piano* (the VP abbreviated as *l.t.p*) where $[\![\text{the kids}]\!]$ = Kelly ⊔ Jane, assuming that D contains Kelly and Jane (as well as Kelly ⊔ Jane, though its presence is inconsequential), D' contains Kelly but neither Jane nor Kelly ⊔ Jane, and D'' contains Jane but neither Kelly nor Kelly ⊔ Jane (compare with (39)).

(iii) a. $[\![[\text{The kids}] [\exists\text{-PL}_D \text{ l.t.p}]]\!] \Leftrightarrow$ Kelly l.t.p ∨ Jane l.t.p ∨ Kelly ⊔ Jane l.t.p
b. $[\![[\text{The kids}] [\exists\text{-PL}_{D'} \text{ l.t.p}]]\!] \Leftrightarrow$ Kelly l.t.p ∨ Kelly ⊔ Jane l.t.p
c. $[\![[\text{The kids}] [\exists\text{-PL}_{D''} \text{ l.t.p}]]\!] \Leftrightarrow$ Jane l.t.p ∨ Kelly ⊔ Jane l.t.p

[60] Non-maximality is, as before, the result of pruning alternatives which yields a weaker meaning.

$[\exists z \in Part(⟦\text{the kids}⟧)[z \circ y \wedge z \text{ lifted the piano in } w]]]$
iff $\forall y \in D \cap Part(⟦\text{the kids}⟧)[\exists z \in Part(⟦\text{the kids}⟧)$
$[z \circ y \wedge z \text{ lifted the piano in } w]]$
iff $⟦[\text{The kids}]\ [^\star \text{ lifted the piano}]⟧(w) = 1$
'Every plurality of kids overlaps some plurality of kids which lifted the piano'

For some non-distributive predicates such as *weigh exactly 20 kg* and *be light enough to carry* parallel meanings to (93b) would be generally too weak; such predicates give rise to stronger collective/distributive meanings in positive sentences. This has been a problem for accounts relying on pluralization using ⋆, for which a solution has been proposed by means of relativizing ⋆ to a *cover* (see Schwarzschild 1994; Heim 1994). This solution can be translated into the current system by letting the predicate be modified by a cover variable (as in Heim 1994; Beck 2001).[61] As the reader may suspect, this move may result in weaker readings under negation.[62] Bar-Lev (2019) however claims that this is as it should be: the predicates for which purely collective/distributive meanings are needed in positive sentences are precisely those predicates which have been identified by Križ (2015) as 'non-homogeneous' on their collective interpretation and for which weaker meanings are needed in negative sentences.

In the next sections I will discuss two more issues having to do with Homogeneity with non-distributive interpretations: §8.2 focuses on Homogeneity with co-distributive meanings, and §8.3 discusses the removal of Homogeneity by non-distributive plural quantifiers, most notably *all*. There are several more issues concerning Homogeneity with non-distributive predication which I will not discuss in this paper. First, sentences such as (88b) have been argued by Križ (2015) not to be true if a plurality of kids together with non-kids lifted the piano, a fact which is not explained so far. Second, scoping the definite plural above negation in (92a) will result in an overly weak meaning given our modified definition of ∃-PL in (91a). For discussion of both of these issues I refer the reader to Bar-Lev (2018, appendix A).

8.2 Homogeneity with two-place predicates

Bar-Lev (2019) points out that a view of Homogeneity based on the definition of ⋆ in (89) further allows a relatively simple explanation for Homogeneity with predicates having two plural DPs as arguments, which is not easily explained on most accounts of Homogeneity (see especially Gajewski 2005, pp. 142–149 and Križ 2015, pp. 58–61). While (94a) can be understood co-distributively, its negation in (94b) can only be true if no girl danced with any boy.

[61] Note that we'd end up with two domain variables: D which gives rise to sub-domain alternatives, and a cover variable which doesn't. This redundancy disappears in Bar-Lev (2018, ch. 5) where the cover variable is instead identified with the domain variable D, but that system is more complex in other respects which for the sake of brevity I do not present here.

[62] In §5.3 this was brought up as a possible source for Non-maximality in negative sentences.

(94) a. The girls danced with the boys.
≈ 'Every girl danced with some boy and every boy danced with some girl'
b. The girls didn't dance with the boys.
≈ 'No girl danced with any boy'

On a view of Homogeneity based on (89), it is quite straightforward to explain this behavior by utilizing a two-place counterpart of *, following common practice. We can define *'s two-place counterpart ** as in (95), assuming (as in (96)) that two pairs overlap if their first coordinates or their second coordinates overlap in order to maintain equivalence between the truth conditions in (95) and standard definitions of **.[63]

(95) $$[\![\star\star]\!](P)(x)(y)(w) = \begin{cases} 1 & \textit{iff} \;\; \forall\langle x',y'\rangle \sqsubseteq \langle x,y\rangle[\exists\langle x'',y''\rangle \sqsubseteq \langle x,y\rangle \\ & \qquad [\langle x'',y''\rangle \circ \langle x',y'\rangle \wedge P(x'')(y'')(w) = 1]] \\ 0 & \textit{iff} \; \neg\exists\langle x',y'\rangle \sqsubseteq \langle x,y\rangle[\exists\langle x'',y''\rangle \sqsubseteq \langle x,y\rangle \\ & \qquad [\langle x'',y''\rangle \circ \langle x',y'\rangle \wedge P(x'')(y'')(w) = 1]] \\ \# & \text{otherwise} \end{cases}$$

(96) $\langle x,y\rangle \circ \langle x',y'\rangle$ *iff* $x \circ x' \vee y \circ y'$

[63] I further rely on the following common assumption (see, e.g., Krifka 1986; Kratzer 2007):

(i) $\langle a,b\rangle \sqsubseteq \langle c,d\rangle$ *iff* $a \sqsubseteq c \wedge b \sqsubseteq d$

The assumption in (96) may seem entirely ad hoc and at odds with the assumption in (i): since we defined the overlap relation based on parthood in (89), and since the parthood relation is generalized to pairs in (i), one could expect the overlap relation to automatically generalize to pairs; but this generalized overlap would give us something different from (96) (i.e., the disjunction in (96) would be replaced with conjunction). Perhaps surprisingly, a way to make (96) fall out from a general definition of overlap is to admit a null element 0 which is part of every plurality in the domain of individuals D_e (for reasons to do so see Landman 2011; Buccola and Spector 2016; Bylinina and Nouwen 2018). Once this is done, a generalized overlap relation between two entities could be defined as having a part in common which is not the bottom element (see Rothstein 2017, p. 33), as in (ii).

(ii) a. $Bot(x)$ *iff* $\neg\exists y[y \sqsubset x]$
b. $x \circ x'$ *iff* $\exists x''[\neg Bot(x'') \wedge x'' \sqsubseteq x \wedge x'' \sqsubseteq x']$

Importantly, if $x \in D_e$, then $Bot(x)$ can only hold if $x = 0$, and if $x \in D_e \times D_e$, then $Bot(x)$ can only hold if $x = \langle 0,0\rangle$. Let me show now that assuming (i) while admitting a bottom element and assuming the notion of overlap in (ii), (96) follows (I will show this assuming the variables x, x', y, y' in (96) all range over members of D_e, which is what's relevant for our discussion). From right to left: Suppose one of the disjuncts on the right hand-side of (96) is true; for instance, assume $x \circ x'$. Then (given (ii)) there must be some x'' s.t. (a) $x'' \sqsubseteq x$, (b) $x'' \sqsubseteq x'$, and (c) $x'' \neq 0$. Now for any arbitrary y and y': $\langle x'',0\rangle \sqsubseteq \langle x,y\rangle$ because of (a) (and due to 0 being part of every member of D_e together with (i)), and similarly $\langle x'',0\rangle \sqsubseteq \langle x',y'\rangle$ because of (b). Since $\langle x'',0\rangle$ is not a bottom element of $D_e \times D_e$ (because of (c)), it also follows that $\langle x,y\rangle \circ \langle x',y'\rangle$. (Of course, the same conclusion can be reached assuming $y \circ y'$.) From left to right: if $\langle x,y\rangle \circ \langle x',y'\rangle$, then there must be a pair $\langle x'',y''\rangle$ distinct from $\langle 0,0\rangle$ which is part of both $\langle x,y\rangle$ and $\langle x',y'\rangle$. This means that there is an x'' and y'' s.t. (a) $x'' \sqsubseteq x \wedge x'' \sqsubseteq x'$; (b) $y'' \sqsubseteq y \wedge y'' \sqsubseteq y'$; and (c) $x'' \neq 0$ or $y'' \neq 0$. If $x'' \neq 0$, then given (a) it follows that $x \circ x'$; and if $y'' \neq 0$, then given (b) it follows that $y \circ y'$. Given that result and (c), it follows that $x \circ x' \vee y \circ y'$.

Here too, our strategy for converting this trivalent view into the implicature view will be to define a two-place existential pluralization operator, ∃∃-PL, and derive the following picture:

(97) a. $[\![\text{DP}_1\ [^{\star\star}\text{V}]\ \text{DP}_2]\!](w) = 1$ *iff* $[\![\mathcal{E}xh_C\ \text{DP}_1\ [\exists\exists\text{-PL}_D\ \text{V}]\ \text{DP}_2]\!](w) = 1$
b. $[\![\text{DP}_1\ [^{\star\star}\text{V}]\ \text{DP}_2]\!](w) = 0$ *iff* $[\![\text{DP}_1\ [\exists\exists\text{-PL}_D\ \text{V}]\ \text{DP}_2]\!](w) = 0$

To achieve this we define ∃∃-PL as follows, and further assume that the sister of plural DPs can be pluralized by applying either ∃-PL or ∃∃-PL.

(98) a. $[\![\exists\exists\text{-PL}]\!] = \lambda D_{\langle e,et\rangle}.\lambda P_{\langle e,\langle e,st\rangle\rangle}.\lambda x_e.\lambda y_e.\lambda w_s.\exists\langle x',y'\rangle \in D \cap Part(\langle x,y\rangle)$
$[\exists\langle x'',y''\rangle \in Part(\langle x,y\rangle)[\langle x'',y''\rangle \circ \langle x',y'\rangle \wedge P(x'')(y'')(w) = 1]]$

I leave it to the reader to verify that the desired truth conditions are derived for the sentences in (94) given the LFs in (99) (when no alternatives are pruned and D contains all pairs of pluralities of boys and girls).[64] As usual, by pruning alternatives in (94a) we can get non-maximal readings.

(99) a. LF of (94a): $\mathcal{E}xh_C$ [the girls] [[∃∃-PL$_D$ [danced with]] [the boys]]
b. LF of (94b): ($\mathcal{E}xh_C$) NEG [the girls] [[∃∃-PL$_D$ [danced with]] [the boys]]

8.3 Plural quantification and Homogeneity removal

Homogeneity disappears when *all* attaches to a definite plural, and so does Non-maximality. As long as we restrict ourselves to distributive predication this is not an issue since *all* could remove both by being a simple universal quantifier over atomic parts of the plurality it applies to. But explaining how *all* can both remove Homogeneity and at the same time allow for non-distributive interpretations as in *all the kids lifted the piano* is quite a complex matter. The issue at stake is not only why *all* removes Homogeneity, but more generally what differentiates plural non-quantificational DPs (e.g., *John and Mary*, *the kids*), which give rise to Homogeneity, from plural quantificational DPs (e.g., *all the kids*, *five kids*) which do not (see Križ 2015). Since my main focus has been providing a theory of Homogeneity with distributive predicates I will only sketch some several directions as to how Homogeneity removal could be carried out within the system developed here, focusing mainly on *all*.

On an intuitive level, what we want is to let *all* (and other quantifiers) do whatever $\mathcal{E}xh$ does, namely turn the existential meaning of plural predication into a universal one. One way to spell this out would be to say that *the kids* and *all the kids* have the same type and meaning, but *all* forces a (local) application of $\mathcal{E}xh$ which allows no pruning.[65] Such a story would be reminiscent of views where *exactly five kids* has

[64] The application of ∃∃-PL to *danced with* is mandated by our assumption that the sister of plural DPs are pluralized. This assumption further requires an ∃-PL to apply to the whole VP (which is a sister of *the girls*); it is omitted here since it's semantically vacuous.

[65] More adequately, what *all* should force is exhaustification only over the alternatives generated by the domain variable of the pluralization operator which attaches to the sister of the DP it associates with. Otherwise we would get undesired results for a sentence like *not all the kids read the books*: instead of the desired meaning 'not every kid read some book' we'd get 'not every kid read every book' (here I assume

the same type and meaning as *five kids*, but *exactly* forces the derivation of an upper bound inference (see Landman 1998; Spector 2014).

While I find this direction appealing, one may entertain more direct ways to achieve this effect. One possibility is to hard-wire it into the semantics of *all*, and assume that *all* has a semantic contribution which mimics what $\mathcal{E}xh$ does. The goal is then to let *all* in the structure *all NP [*$\exists$*-PL$_D$ VP]* quantify universally over all assignments of values for the domain variable D which are subsets of $Part(⟦\text{NP}⟧)$. To achieve this we can assume that *all* takes two arguments: one of type e (the NP), and another of type $\langle et, \langle e, st\rangle\rangle$. The latter is supplied by abstracting over the domain variable, as in (100a).[66] The meaning of *all* can then be defined as in (100b), and the overall result is as in (100c): the sentence will be true as long as $⟦^\star$ lifted the piano$⟧$ is true of the plurality consisting of all the kids.[67]

(100) a. [All the kids] [λD $\exists$-PL$_D$ lifted the piano]

b. $⟦\text{all}⟧ =$
$$\lambda x_e.\lambda P_{\langle et,\langle e,st\rangle\rangle}.\lambda w_s.\forall D' \subseteq Part(x)[D' \neq \varnothing \rightarrow P(D')(x)(w) = 1]$$

c. $⟦(100a)⟧(w) = 1$ *iff*
$$\forall D' \subseteq Part(⟦\text{the kids}⟧)[D' \neq \varnothing \rightarrow \exists y \in D' \cap Part(⟦\text{the kids}⟧)[\exists z \in Part(⟦\text{the kids}⟧)[z \circ y \wedge z \text{ lifted the piano in } w]]]$$
iff $\forall y \in Part(⟦\text{the kids}⟧)[\exists z \in Part(⟦\text{the kids}⟧)[z \circ y \wedge z \text{ lifted the piano in } w]]$

Finally, there is a simpler way to think about Homogeneity removal. The two previous directions relied on the assumption that pluralization is carried out using $\exists$-PL no matter whether the DP is quantificational or not. Homogeneity removal would however be immediately explained if we dropped this assumption and stipulated instead that pluralization of the sister of plural quantificational DPs is carried out by $\forall$-PL, a covert universal counterpart of $\exists$-PL—essentially, a (bivalent) * operator:

that *the books* could take non-surface scope below negation so that its sister is headed by $\exists$-PL and there's no $\exists\exists$-PL in the structure; if $\exists\exists$-PL applies to the verb, following §8.2, a more serious issue arises; see fn. 70).

[66] One would have to make sure that the domain variable abstracted over is the right one in case there's more than one pluralization operator in the structure (see fn. 65). An indexing mechanism along the lines of Križ and Spector (2017) might be utilized to this end.

[67] Similar solutions can be utilized for Homogeneity removal by other upward monotone plural quantifiers as in *five kids lifted the piano*. We can posit a covert existential quantifier $\exists_{all}$ which quantifies over pluralities of five kids and incorporates the Homogeneity removal effect we implemented in the semantics of *all*, as in (ia). The sentence ends up true as long as there is a plurality of five kids that $⟦^\star$ lifted the piano$⟧$ is true of. As pointed out by a reviewer, this line of thought does not extend well to non upward-monotone quantifiers, due to the existential quantification introduced by $\exists_{all}$.

(i) a. LF: [$\exists_{all}$ [five kids]] [λD $\exists$-PL$_D$ lifted the piano]

b. $⟦\exists_{all}⟧ =$
$$\lambda Q_{\langle e,st\rangle}.\lambda P_{\langle et,\langle e,st\rangle\rangle}.\lambda w_s.\exists x \in Q[\forall D' \subseteq Part(x)[D' \neq \varnothing \rightarrow P(D')(x)(w) = 1]]$$

c. $⟦(\text{ia})⟧(w) = 1$ *iff*
$$\exists x \in ⟦\text{five kids}⟧[\forall y \in Part(x)[\exists z \in Part(x)[z \circ y \wedge z \text{ lifted the piano in } w]]]$$

We can also define the semantics of floating *all* (e.g., in *the kids all lifted the piano*) based on (100b) with the minimal difference that it first takes the $\langle et, \langle e, st\rangle\rangle$-type argument and then the e-type argument.

(101) $[\![\forall\text{-PL}]\!] = \lambda D_{et}.\lambda P_{\langle e,st\rangle}.\lambda x_e.\lambda w_s.$
$\forall y \in D \cap Part(x)[\exists z \in Part(x)[z \circ y \wedge P(z)(w) = 1]]$

Positing such an operator may seem inherently at odds with the account of Homogeneity developed in this paper: by assuming a universal counterpart of ∃-PL we run the risk of deriving a conjunctive alternative to *the kids laughed* which would obliterate the derivation of Maximality implicatures (recall the discussion in §4.5). Note however that what's untenable on our view is not the existence of such operators in and of itself, but rather the possibility of deriving universal alternatives based on them. To avoid this situation we can entertain the following view: Every plural DP requires a pluralization operator to head its sister (as we assumed so far); while quantificational DPs require ∀-PL to head their sister, non-quantificational DPs require ∃-PL to head their sister (in fn. 52 I mentioned the possibility that there might be some cross-linguistic variation with respect to which DP-types license which pluralization operator). Homogeneity removal is simply the result of this sort of agreement between quantificational DPs and ∀-PL.[68]

As a result, only (102a) would be a licit LF of *the kids laughed* and only (102c) would be a licit LF of *all the kids laughed*.[69] We would then be able to explain why none of the LFs featuring ∀-PL in (102) could be derived as alternatives to (102a): under the reasonable assumption that only licit LFs can serve as alternatives, (102b) cannot be an alternative at all; and (102c) is, as before, ruled out as an alternative to (102a) due to complexity.

(102) a. ✓[The kids] [∃-PL laughed]
b. ✗[The kids] [∀-PL laughed]
c. ✓[All the kids] [∀-PL laughed]
d. ✗[All the kids] [∃-PL laughed]

More generally, sentences with non-quantificational plural DPs would not give rise to alternatives where the pluralization operator heading the DP's sister is ∀-PL, because we would only be able get a licit LF when we replace ∃-PL with ∀-PL if we also replace the non-quantificational DP with a quantificational DP at the same time; but the latter would be more complex than the former (at least semantically, see fn. 26), and hence ruled out as an alternative.[70]

[68] It is straightforward to define ∀∀-PL, a two-place counterpart of ∀-PL. Quantificational DPs could allow for co-distributive readings if we assumed that they require either ∀-PL or ∀∀-PL to head their sister (much like our assumption in §8.2 that non-quantificational DPs require either ∃-PL or ∃∃-PL to head their sister). See however fn. 70 for some complications with this view.

[69] For the purposes of this discussion we can assume that *all* is semantically vacuous and only serves as a way to license ∀-PL. This would be similar in spirit to the view of Homogeneity removal by *all* in Brisson (1998, 2003), though here the effect of *all* is changing the type of the pluralization operator involved, while for her it changes the type of the covers which can serve as restrictors of the pluralization operator.

[70] A reviewer points out that explaining how *all* can remove Homogeneity and at the same time allow for co-distributive interpretations is quite a complex matter, given the view of Homogeneity with two-place predicates from §8.2. Consider a sentence with a non-quantificational plural in object position and a quantificational plural in subject position, as in (i). Given our assumptions from §8.2, ∃∃-PL could attach to the verb *read*, and the result (simplified) would be the predicate in (ia) (there is also another possible LF, mentioned in fn. 65, which involves no ∃∃-PL; this LF would however not derive co-distributivity).

9 Conclusion

I argued in favor of an analysis of Homogeneity which parallels the Implicature account of Free Choice. I have shown that when such a perspective is taken non-maximal readings are predicted to arise, and their context dependency can be reduced to the context sensitivity of implicature calculation. At the same time, this analysis can explain otherwise intriguing asymmetries between positive and negative sentences in the acquisition of Homogeneity, and with Non-maximality and Gappiness.

Acknowledgments First and foremost, I would like to thank my PhD thesis supervisors, Luka Crnič and Danny Fox, for endless discussions on this work. I also greatly benefited from talking to Itai Bassi, Elitzur Bar-Asher Siegal, Gennaro Chierchia, Kai von Fintel, Martin Hackl, Irene Heim, Roni Katzir, Manuel Križ, Daniel Margulis, Roger Schwarzschild and Benjamin Spector, among others. Feedback I got at SALT 28 and the LF Reading Group at MIT was very helpful. My work on this project was funded by the Language, Logic and Cognition Center at the Hebrew University of Jerusalem, as well as the Jack, Joseph and Morton Mandel School for Advanced Studies in the Humanities, which also funded my stay at MIT in the academic year 2017–2018 during which much of this work was developed.

References

Aloni, Maria. 2007. Free choice, modals, and imperatives. *Natural Language Semantics* 15(1): 65–94. doi:10.1007/s11050-007-9010-2.

Aloni, Maria. 2016. FC disjunction in state-based semantics. Talk given at Logical Aspects of Computational Linguistics 2016, Nancy, France.

Alonso-Ovalle, Luis. 2005. Distributing the disjuncts over the modal space. In L. Bateman and C. Ussery (eds.), *North east linguistic society (NELS)*, vol. 35, 1–12.

(i) All the girls read the books.

a. $\lambda x.\lambda y.\lambda w.\exists x' \sqsubseteq x[\exists y' \sqsubseteq y[y' \text{ read } x' \text{ in } w]]$

While the existential quantification over parts of the first argument of the verb is precisely what we want in the basic semantics, since this would give us the desired existential quantification over books, for the second argument we would want something stronger. But since ∃∃-PL lumps both arguments together, it would be quite difficult to only strengthen the quantification over the second argument while leaving the quantification over the first argument intact. Indeed, all the views of Homogeneity removal considered in this section predict that it doesn't matter whether we have *all the girls* or *the girls* in subject position in (i), we should expect the same truth conditions (that is, for LFs where ∃∃-PL applies to the verb). This is of course wrong.
One of the directions pursued in this section faces a problem also with cases where there is a non-quantificational plural in subject position and a quantificational plural in object position, as in (ii): Given the entry for *all* in (100b), it would be impossible to abstract over the domain variable introduced by ∃∃-PL which is of type $\langle e, et\rangle$ and get the right kind of argument for *all the books* (that is, even if *all the books* does not remain in situ).

(ii) The girls read all the books.

Since this section is only meant to sketch directions for accounts of Homogeneity removal I cannot seriously pursue here a solution to these problems. A possible direction is to move to an event semantics framework where ∃∃-PL (and possibly ∀∀-PL, see fn. 68) does not apply directly to the verb, but rather to syntactically realized thematic roles; such a view would allow us to determine the quantificational force for each argument position independently of the others. See Bar-Lev (2018, appendix A.3) for a framework of this sort utilized in order to account for what Križ calls Upward Homogeneity.

Bar-Lev, Moshe E. 2018. Free choice, homogeneity, and innocent inclusion. PhD dissertation, Hebrew University of Jerusalem.

Bar-Lev, Moshe E. 2019. Specification and homogeneity in plural predication. Manuscript.

Bar-Lev, Moshe E. and Danny Fox. 2017. Universal free choice and innocent inclusion. In Dan Burgdorf et al. (eds.), *Semantics and linguistic theory (SALT)*, vol. 27, 95–115.

Bar-Lev, Moshe E. and Danny Fox. 2019. Free choice and scope. Manuscript.

Bar-Lev, Moshe E. and Danny Fox. 2020. Free choice, simplification, and innocent inclusion. *Natural Language Semantics* 1–49.

Bar-Lev, Moshe E. and Daniel Margulis. 2014. Hebrew *kol*: a universal quantifier as an undercover existential. In U. Etxeberria et al. (eds.), *Proceedings of sinn und bedeutung 18*, 60–76.

Barker, Chris. 2010. Free choice permission as resource-sensitive reasoning. *Semantics and Pragmatics* 3(10): 1–38. doi:10.3765/sp.3.10.

Barner, David and Asaf Bachrach. 2010. Inference and exact numerical representation in early language development. *Cognitive psychology* 60(1): 40–62.

Barner, David, Neon Brooks and Alan Bale. 2011. Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition* 118(1): 84–93.

Bassi, Itai and Moshe E. Bar-Lev. 2016. A unified existential semantics for bare conditionals. In R. Truswell et al. (eds.), *Proceedings of sinn und bedeutung 21*, 125–142.

Beck, Sigrid. 2001. Reciprocals are definites. *Natural language semantics* 9(1): 69–138.

Bowler, Margit. 2014. Conjunction and disjunction in a language without 'and'. *Semantics and Linguistic Theory* 24: 137–155. doi:10.3765/salt.v24i0.2422.

Brisson, Christine. 2003. Plurals, all, and the nonuniformity of collective predication. *Linguistics and philosophy* 26(2): 129–184.

Brisson, Christine M. 1998. Distributivity, maximality, and floating quantifiers. PhD dissertation, Rutgers.

Buccola, Brian and Benjamin Spector. 2016. Modified numerals and maximality. *Linguistics and philosophy* 39(3): 151–199.

Bylinina, Lisa and Rick Nouwen. 2018. On "zero" and semantic plurality. To appear in Glossa.

Caponigro, Ivano, Lisa Pearl, Neon Brooks and David Barner. 2012. Acquiring the meaning of free relative clauses and plural definite descriptions. *Journal of semantics* 29(2): 261–293.

Champollion, Lucas. 2016a. Covert distributivity in algebraic event semantics. *Semantics and Pragmatics* 9: 15–1.

Champollion, Lucas. 2016b. Overt distributivity in algebraic event semantics. *Semantics and Pragmatics* 9: 16–1.

Champollion, Lucas. to appear. Distributivity, collectivity and cumulativity. In Lisa Mathewson et al. (eds.), *Wiley's companion to semantics*, Wiley.

Chemla, Emmanuel. 2009. Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics* 2(2): 1–33. doi:10.3765/sp.2.2.

Chemla, Emmanuel and Lewis Bott. 2014. Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition* 130(3): 380 – 396. doi:10.1016/j.cognition.2013.11.013.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. Oxford University Press.

Crnič, Luka, Emmanuel Chemla and Danny Fox. 2015. Scalar implicatures of embedded disjunction. *Natural Language Semantics* 23(4): 271–305. doi:10.1007/s11050-015-9116-x.

Dalrymple, Mary, Makoto Kanazawa, Sam Mchombo and Stanley Peters. 1994. What do reciprocals mean? In *Semantics and linguistic theory 4*, 61–78.

Dotlačil, Jakub. 2010. Anaphora and distributivity: A study of same, different, reciprocals and others. PhD dissertation, Utrecht University.

Dowty, David. 1987. Collective predicates, distributive predicates and *all*. In *Proceedings of the 3rd ESCOL*, 97–115.

von Fintel, Kai. 1994. Restrictions on quantifier domains. PhD dissertation, University of Massachusetts at Amherst.

von Fintel, Kai. 1999. Npi licensing, strawson entailment, and context dependency. *Journal of Semantics* 16(2): 97–148.

Flor, Enrico, Nina Haslinger, Hilda Koopman, Eva Rosina, Magdalena Roszkowski and Viola Schmitt. 2017. Cross-linguistic evidence for a non-distributive lexical meaning of conjunction. In Alexandre Cremers, Thom van Gessel and Floris Roelofsen (eds.), *Proceedings of the 21st amsterdam colloquium*, 255–264.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland and Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 71–120. London: Palgrave Macmillan UK. doi:10.1057/9780230210752_4.

Fox, Danny. 2017. Exhaustivity and the presupposition of questions. Talk given at Sinn und Bedeutung 22.

Fox, Danny and Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1): 87–107. doi:10.1007/s11050-010-9065-3.

Fox, Danny and Benjamin Spector. 2018. Economy and embedded exhaustification. *Natural Language Semantics* 26(1): 1–50. doi:10.1007/s11050-017-9139-6. https://doi.org/10.1007/s11050-017-9139-6.

Gajewski, Jon Robert. 2005. Neg-raising: Polarity and presupposition. PhD dissertation, Massachusetts Institute of Technology.

Gillon, Brendan S. 1987. The readings of plural noun phrases in english. *Linguistics and philosophy* 10(2): 199–219.

Gillon, Brendan S. 1990. Plural noun phrases and their readings: A reply to lasersohn. *Linguistics and Philosophy* 13(4): 477–485.

Goldstein, Simon. 2018. Free choice and homogeneity. Unpublished manuscript.

Gualmini, Andrea, Stephen Crain, Luisa Meroni, Gennaro Chierchia and Maria Teresa Guasti. 2001. At the semantics/pragmatics interface in child language. In *Semantics and linguistic theory*, vol. 11, 231–247.

Heim, Irene. 1994. Plurals. Lecture notes.

Horn, Laurence Robert. 1972. On the semantic properties of logical operators in english. PhD dissertation, University of California, Los Angeles.

Kamp, Hans. 1974. Free choice permission. *Proceedings of the Aristotelian Society* 74(1): 57–74. doi:10.1093/aristotelian/74.1.57.

Karmiloff-Smith, Annette. 1981. *A functional approach to child language: A study of determiners and reference*. Cambridge studies in linguistics 24. Cambridge University Press.

Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30(6): 669–690. doi:10.1007/s10988-008-9029-y.

Katzir, Roni. 2014. On the roles of markedness and contradiction in the use of alternatives. In Salvatore Pistoia Reda (ed.), *Pragmatics, semantics and the case of scalar implicatures*, 40–71. Springer. doi:10.1057/9781137333285_3.

Klinedinst, Nathan and Daniel Rothschild. 2012. Connectives without truth tables. *Natural Language Semantics* 20(2): 137–175. doi:10.1007/s11050-011-9079-5. https://doi.org/10.1007/s11050-011-9079-5.

Kobayashi, Filipe Hisao and Vincent Rouillard. 2019. Tying free choice in questions to distributivity. Manuscript, MIT.

Kratzer, Angelika. 2007. On the plurality of verbs. *Event structures in linguistic form and interpretation* 269–300.

Kratzer, Angelika and Junko Shimoyama. 2002. Indeterminate pronouns: The view from japanese. In Yukio Otsu (ed.), *The third tokyo conference on psycholinguistics*, 1–25. Hituzi Syobo.

Krifka, Manfred. 1986. Nominalreferenz und zeitkonstitution: Zur semantik von massentermen, pluraltermen und aspektklassen. PhD dissertation, University of Munich.

Krifka, Manfred. 1995. The Semantics and Pragmatics of Polarity Items. *Linguistic Analysis* 25: 209–257.

Krifka, Manfred. 1996. Pragmatic strengthening in plural predications and donkey sentences. In *Semantics and linguistic theory*, vol. 6, 136–153.

Križ, Manuel and Benjamin Spector. 2017. Interpreting plural predication: Homogeneity and non-maximality. Ms., Institut Jean Nicod.

Križ, Manuel. 2015. Aspects of homogeneity in the semantics of natural language. PhD dissertation, University of Vienna.

Križ, Manuel. 2016. Homogeneity, non-maximality, and all. *Journal of Semantics* 33: 1–47.

Križ, Manuel. 2017. In soviet russia, alcohol is dependent on you. In Clemens Mayr and Edwin Williams (eds.), *Festschrift for martin prinzhorn*, 173–180. Wiener Linguistiche Gazette.

Križ, Manuel and Emmanuel Chemla. 2015. Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics* 23(3): 205–248.

Landman, Fred. 1998. Plurals and maximalization. In *Events and grammar*, 237–271. Springer.

Landman, Fred. 2011. Boolean pragmatics. Unpublished manuscript.

Lasersohn, Peter. 1989. On the readings of plural noun phrases. *Linguistic Inquiry* 20(1): 130–134. http://www.jstor.org/stable/4178619.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 522–551.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice theoretical approach. In Rainer Bäuerle, Christoph Schwarze and Arnim von Stechow (eds.), *Meaning, use, and interpretation of language*, 127–144. Walter de Gruyter.

Link, Godehard. 1987. *Generalized quantifiers and plurals*. Springer.

Löbner, Sebastian. 1987. Natural language and generalized quantifier theory. In *Generalized quantifiers*, 181–201. Springer.

Löbner, Sebastian. 2000. Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy* 23(3): 213–308.

Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural language semantics* 17(3): 245–297. doi:10.1007/s11050-009-9042-x.

Magri, Giorgio. 2014. An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening. In *Pragmatics, semantics and the case of scalar implicatures*, 99–145. Palgrave Macmillan.

Malamud, Sophia A. 2012. The meaning of plural definites: A decision-theoretic approach. *Semantics and Pragmatics* 5(3): 1–58. doi:10.3765/sp.5.3.

Meyer, Marie-Christine. 2016. Generalized free choice and missing alternatives. *Journal of Semantics* 33(4): 703–754. doi:10.1093/jos/ffv010.

Oikonomou, Despina. 2016. Imperatives are existential modals: Deriving the strong reading as an implicature. In *Semantics and linguistic theory 26*, 1043–1062. doi:10.3765/salt.v26i0.3949. http://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/26.1043.

Rothstein, Susan. 2017. *Semantics for counting and measuring*. Cambridge University Press.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and philosophy* 27(3): 367–391. doi:10.1023/B:LING.0000023378.71748.db.

Schwarz, Florian. 2013. Maximality and definite plurals—experimental evidence. In Emanuel Chemla, Vincent Homer and Grégoire Winterstein (eds.), *Proceedings of sinn und bedeutung 17*, 509–526. http://semanticsarchive.net/sub2012/.

Schwarzschild, Roger. 1994. Plurals, presuppositions and the sources of distributivity. *Natural Language Semantics* 2(3): 201–248.

Schwarzschild, Roger. 1996. *Pluralities*. Springer Science & Business Media.

Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar and Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 24(4): 305–352. doi:10.1007/s11050-016-9126-3.

Spector, Benjamin. 2013. Homogeneity and plurals: From the strongest meaning hypothesis to supervaluations. Talk presented at Sinn und Bedeutung 18.

Spector, Benjamin. 2014. Plural indefinites and maximality. Talk given at the UCLA linguistics colloquium.

Spector, Benjamin. 2018. Homogeneity and non-maximality within the rational speech act model. In *Proceedings of the amsterdam colloquium*, vol. 21, 435–444.

Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic inquiry* 19(3): 425–449.

Stanley, Jason and Zoltán Gendler Szabó. 2000. On quantifier domain restriction. *Mind & Language* 15(2-3): 219–261.

Starr, William. 2016. Expressing permission. In *Semantics and linguistic theory*, vol. 26, 325–349. doi:10.3765/salt.v26i0.3832.

Tieu, Lyn, Manuel Križ and Emmanuel Chemla. 2015. Children's acquisition of homogeneity in plural definite descriptions. Ms., LSCP.

Tieu, Lyn, Jacopo Romoli, Peng Zhou and Stephen Crain. 2016. Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics* 33(2): 269–298. doi:10.1093/jos/ffv001. http://dx.doi.org/10.1093/jos/ffv001.

Westerståhl, Dag. 1984. Determiners and context sets. In *Generalized quantifiers in natural language*, 45–71. Foris Dordrecht.

Willer, Malte. 2017. Simplifying with free choice. *Topoi* doi:10.1007/s11245-016-9437-5. https://doi.org/10.1007/s11245-016-9437-5.

von Wright, Georg Henrik. 1968. *An essay in deontic logic and the general theory of action*. Amsterdam: North-Holland.

Zhou, Peng, Jacopo Romoli and Stephen Crain. 2013. Children's knowledge of free choice inferences. In *Semantics and linguistic theory*, vol. 23, 632–651.

Zimmermann, Thomas Ede. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8(4): 255–290. doi:10.1023/A:1011255819284.