Approaches to scalar inferences

Bob van Tiel[1]

[1]Radboud University Nijmegen, The Netherlands

To appear in:

Antonio Fábregas, Laia Mayol, Clàudia Pons-Mol, and Yanina Prystauka (Eds.)

The Cambridge handbook of linguistic interfaces

Version 21 August 2025 Word count: 9,897

Abstract

Scalar inferences (e.g., the inference from 'some' to 'not all') have traditionally been explained as a variety of Gricean conversational implicature. In recent years, however, this idea by grammaticalist accounts, which attribute scalar inferences to a covert syntactic operator that enriches semantic meaning. This chapter surveys the debate between these pragmatic and semantic approaches, focusing on five topics that have stood at the center of the debate: embedded scalar inferences, obligatory inferences, scalar inferences in non-cooperative contexts, universal free choice inferences, and Hurford's constraint. Reviewing theoretical arguments and experimental findings, I argue that the empirical and conceptual evidence continues to favour a broadly Gricean explanation. In doing so, the chapter illustrates how scalar inferences sit at the heart of the semantics–pragmatics interface, and reveals how meaning arises from the dynamic interaction between compositional structure and pragmatic reasoning.

1. Introduction

Sentences such as (1) have at least two possible interpretations.

(1) It is likely to rain.

According to its *one-sided* interpretation, (1) merely conveys that the probability of rain exceeds some contextually-determined threshold; say, 50%. According to its *two-sided* interpretation, the sentence additionally conveys that the probability of rain lies below 100%.

To analyse this ambiguity, it is often assumed that expressions such as 'likely' evoke a *lexical scale*. Lexical scales are sets of expressions that are ordered in terms of logical entailment, as well as satisfying certain additional criteria, such as pertaining to the same semantic domain, being structurally equally complex, and belonging to the same part of speech (e.g., Atlas & Levinson, 1981; Gazdar, 1979; Soames, 1982). For (1), the relevant scale is <likely, certain>. Here, 'certain' is logically stronger than 'likely' since if an event is certain to occur, it is thereby necessarily also likely to occur—though not the other way around.

Using the notion of a lexical scale, we can thus formulate the following generalisation: *a sentence with the weaker scalar expression in an unembedded position may imply that the corresponding sentence with the stronger scalar expressions is false*.

This type of implication is often called a *scalar inference*. Some authors prefer the term *scalar implicature*, but I will avoid this term since it is a matter of debate as to whether the implication counts as a conversational implicature—precisely what is at issue in this chapter. Nor is it clear whether the implication is an inference rather than, e.g., an aspect of the lexical meaning of scalar expressions. To remain theory-neutral, Geurts and van Tiel (2013) proposed the term *upper-bounded construal*. However, given that 'scalar inference' is the prevailing label in the literature, I will use that term here, but it should be understood in the same theory-neutral way as Geurts and van Tiel's term of art.

Much of the research on scalar inferences has concentrated on only a handful of lexical scales, notably <some, all> and <or, and>. However, the class of lexical scales is both large and diverse; a sample is given in Table 1. Over the past years, a substantial literature has emerged that seeks to capture and explain variability across lexical scales—a phenomenon called *scalar diversity* (e.g., Gotzner et al., 2017; Hu et al., 2023; Pankratz & van Tiel, 2021; Ronai & Xiang, 2024; Sun et al., 2018; van Tiel et al., 2016).

[INSERT TABLE 1 ABOUT HERE]

A central question in the literature is *why* scalar inferences arise. Broadly speaking, these theories fall into one of two groups: either they attribute scalar inferences to *pragmatic* reasoning, or they argue that scalar inferences are incorporated into the *semantic* meaning of sentences containing scalar expressions.

This chapter provides an overview of the debate between these two perspectives, focusing on the Gricean and grammaticalist theories as representative of the pragmatic and semantic traditions, respectively. The next two sections outline the main assumptions of each theory. I then zoom in on five topics that have been at the forefront of the debate between these two theories: embedded scalar inferences, obligatory

scalar inferences, scalar inferences in non-cooperative settings, universal free choice inferences, and Hurford's constraint.

2. The Gricean theory of scalar inferences

The Gricean theory rests on the central assumption that scalar inferences are a species of *conversational implicature*. Conversational implicatures are inferences that can be rationally derived on the basis of the literal meaning of an utterance and the assumption that the speaker is *cooperative*, i.e., does their best to further the goal of the conversation (Grice, 1975).

To illustrate, consider (1) again, here repeated as (2a). Someone who utters this sentence could have been more informative—and hence cooperative—by saying (2b).

(2) a. It is likely to rain.
b. It is certain to rain.

The *alternative* in (2b) is more informative because it is *logically stronger*, i.e., true in a smaller set of situations than (2a). If the speaker is cooperative, why did they not use (2b)? In many contexts, the most plausible answer is that they do not believe (2b) to be true. This *primary* scalar inference can be strengthened if it is assumed that the speaker knows whether or not the alternative is true. If this *competence assumption* holds, it follows that the speaker believes (2b) to be false (e.g., Sauerland, 2004; Soames, 1982; Spector, 2006; van Rooij & Schulz, 2004).

The Gricean theory of scalar inferences was originally developed by Horn (1972), and further refined by, among others, Gazdar (1979), Atlas and Levinson (1981), Soames (1982), Horn (1989), and Hirschberg (1991). See Geurts (2010) for a recent defense of the Gricean theory. In many respects, the Gricean theory aligns with the relevance-theoretic treatment of scalar inferences (e.g., Carston, 1998; Noveck & Sperber, 2007), and much of what I will have to say about the former also applies to the latter. More recently, the Gricean framework has been formalized using tools from *game theory* (e.g., Franke, 2011) and *Bayesian inference* (e.g., Frank & Goodman, 2012).

The Gricean theory is appealing for a number of reasons. First, it makes use of principles that are independently needed and grounded in plausible assumptions about cooperative behaviour (e.g., the assumption that one ought to provide reliable information that is pertinent to the goal of the conversation). It is generally agreed that, when there is no evidence to the contrary, such independently motivated explanations are to be preferred over explanations that stipulate a lexical or syntactic ambiguity—such as the grammaticalist theory that will be discussed in Section 3.

Second, the Gricean theory explains why scalar inferences are *explicitly cancellable*, i.e., why they can be rejected by the speaker without contradiction. To illustrate, compare (3):

(3) a. *It is likely to rain; in fact, it is impossible.
b. It is likely to rain; in fact, it is certain.

In the second sentence of (3a), the speaker rejects the literal meaning of the first sentence (that the probability of rain exceeds, say, 50%), which results in infelicity. By contrast, in (3b), the speaker rejects the scalar inference (that it is not certain to rain), which remains perfectly felicitous. Hence, unlike the literal meaning, scalar inferences can be cancelled without contradiction

Cancellability is a hallmark of conversational implicatures (Grice, 1975). Hence, the fact that scalar inferences have this property supports the idea that they are a species of conversational implicature.

Third, the Gricean theory explains why scalar inferences are absent when scalar expressions occur in *downward-entailing environments*, such as the scope of negation. To illustrate, compare (4):

(4) a. It is not likely to rain.
b. It is not certain to rain.

In (4a), the alternative with ‘certain’—that is, (4b)—is not logically stronger, since the scalar expression occurs the scope of negation. Consequently, the Gricean theory

correctly predicts that, normally, an utterance of (4a) does not imply that (4b) is false; that is, it does not imply that it is certain to rain.

The hedge with 'normally' is necessary, here, since the inference from (4a) to the negation of (4b) *can* arise. However, this interpretation requires prosodic emphasis on the scalar expression, as well as an explicit contrast with the stronger scalemate, as illustrated in (5). (Small caps indicate prosodic stress.)

(5) It is not LIKELY to rain; it is CERTAIN.

In this example, the semantic meaning of 'likely' is *narrowed* to exclude 'certain'. Proponents of the Gricean view argue that such narrowing involves a semantic process distinct from the pragmatic mechanism that typically gives rise to scalar inferences (e.g., Geurts, 2010; Geurts & van Tiel, 2013; Horn, 2006). This distinction is supported by examples such as (6) (from Horn, 1985, and Matsumoto, 1995).

(6) a. This is not a CAR; it is a FERRARI.
b. It was WARM yesterday, and it was A LITTLE BIT MORE THAN WARM today.

In these cases, too, the meanings of 'car' and 'warm' are semantically narrowed relative to 'Ferrari' and 'a little bit more than warm', respectively, even though these narrowings do not correspond to scalar inferences (e.g., using 'car' does not normally imply that it is not a Ferrari).

Returning to the contrast in (4), note that the sentence with 'likely' is actually *logically stronger* than the one wit 'certain'. Thus, the Gricean theory predicts that uttering (4b) may imply that (4a) is false—that is, that it is likely to rain. This prediction appears correct: the inferential pattern associated with lexical scales indeed reverses in downward-entailing contexts.

Despite these explanatory successes, the Gricean theory has faced significant challenges. Before turning to the arguments against it, I will first outline its grammaticalist competitor.

3. The grammaticalist theory

The grammaticalist theory explains scalar inferences by postulating a hidden syntactic operator, which is sometimes written as *O*. This operator can be appended to any sentence node in the syntactic structure of a simple or complex sentence. The semantic effect of *O* is similar to that of the expression 'only': it indicates that the sentence it modifies is true, but that all alternatives that are relevant and logically stronger are false. Thus, when *O* is appended to (1), the sentence is read essentially as (7), which implies the scalar inference that it is not certain to rain.

(7) It is only likely to rain.

The grammaticalist theory was introduced by Chierchia (2004), building on earlier work by Landman (1998). It has subsequently been developed by, among others, Chierchia (2006), Fox (2007), and Chierchia et al. (2012).

At first sight, however, the grammaticalist theory lacks several explanatory advantages of the Gricean approach. First, unlike the Gricean theory, it is fundamentally stipulative. Proponents often draw informal parallels between the *O*-operator and Gricean reasoning, but these connections are merely associative. Gricean reasoning concerns a hearer's inference about a speaker's communicative intentions—not the presence of silent operators in the syntactic representation.

Second, the grammaticalist theory cannot straightforwardly explain *ignorance inferences* that arise when the speaker is not competent about the stronger alternative. As noted earlier, the scalar inference from (1) that, according to the speaker, it is not certain to rain relies on the assumption that the speaker knows whether or not it is certain to rain. If this competence assumption is implausible, the utterance gives rise to the ignorance inference that the speaker does not know whether it is certain to rain.

Proponents of the grammaticalist theory acknowledge that ignorance inferences fall outside of their purview, and argue that these inferences should be explained in terms of Gricean reasoning (e.g., Chierchia et al., 2012; Fox, 2007). The grammaticalist theory thus relies on the Gricean theory to explain ignorance inferences. Although this

may seem an important limitation, note that the Gricean theory, too, needs to invoke a secondary semantic mechanism to account for cases of narrowing.

Third, since the grammaticalist theory holds that scalar inferences are semantically entailed rather than pragmatically implicated, it remains unclear why they should be explicitly cancellable. To address this issue, Chierchia (2004) simply stipulates that scalar inferences can be overridden by semantic or contextual factors. Again, this stipulation undermines the explanatory appeal of the grammaticalist theory.

At the same time, the force of this argument should not be overstated, for at least two reasons. First, it has been convincingly shown that conversational implicatures are not always cancellable, notwithstanding Grice's claim to the contrary (e.g., Mayol & Castroviejo, 2013; van Kuppevelt, 1996; Weiner, 2006). Second, there are meaning aspects other than conversational implicatures that can be cancelled, too. A case in point involves syntactic ambiguities. To illustrate, consider (8).

(8) Every student speaks two languages.

This sentence can mean either (i) that every student speaks two languages but not necessarily the same ones, or (ii) that there are two specific languages that every student speaks. The second interpretation can be felicitously cancelled, e.g., by continuing with 'But they do not all speak the same two languages' (e.g., Burton-Roberts, 1984; Sadock, 1978). Hence, if it is assumed that sentences with scalar expressions are syntactically ambiguous between a parse with and without the *O*-operator, it may be entirely expected that they are cancellable.

A fourth limitation of the grammaticalist theory is that it does not straightforwardly explain the absence of scalar inferences in downward-entailing environments. To illustrate, consider (4a) again, here repeated as (9).

(9) It is not likely to rain.

Here, the *O*-operator has two scope sites where it can be appended: (i) the embedded sentence ‘it is likely to rain’, and (ii) the matrix sentence ‘It is not likely to rain’. If the operator is appended to the embedded sentence, the resulting interpretation can be paraphrased as in (10).

(10) It is not the case that rain is likely but not certain.

Crucially, this interpretation is compatible with a situation in which rain is certain. As discussed earlier, this is indeed a possible interpretation, but only in cases of narrowing, i.e., if ‘likely’ is accentuated and contrasted with ‘certain’. Without such marking, (9) is normally taken to exclude the possibility that rain is certain. Hence, the grammaticalist theory needs to include a proviso stating that the *O*-operator is normally not to be appended to sentences that occur in downward-entailing environments.

This issue raises a more general question, namely to which sentence nodes the *O*-operator is to be appended. Various answers have been suggested:

i. The *O*-operator is always appended, but remains inert if the stronger alternative is not contextually relevant (e.g., Magri, 2009).
ii. The *O*-operator is always appended unless this results in a logically weaker interpretation of the entire sentence (e.g., Chemla & Spector, 2011).
iii. The *O*-operator is only appended if this results in a logically stronger interpretation of the entire sentence (e.g., Chierchia et al., 2012).

Whatever the descriptive merits of these suggestions, they are clearly stipulative rather than being grounded in generally accepted principles of rational behaviour.

Moreover, the question as to when the *O*-operator is to be appended also introduces a dilemma: on the one hand, the grammaticalist theory offers a potentially convincing account of cases of narrowing; on the other hand, most of the proposed

distributional principles (in particular, ii. and iii.) appear to block two-sided interpretations of scalar expressions under negation

In summary, the grammaticalist theory lacks much of the explanatory appeal of the Gricean theory. However, proponents of the grammaticalist theory have argued that there are decisive reasons for abandoning the Gricean theory in favour of its grammaticalist competitor. In the next section, I discuss five of the most prominent arguments.

4. Arguments against the Gricean theory

4.1. Embedded scalar inferences

Much of the original impetus for the grammaticalist theory stemmed from the observation that scalar expressions seem to receive a two-sided interpretation even in embedded contexts. Consider (11), in which 'likely' is embedded under 'believe'.

(11) John believes that it is likely to rain. = BEL_{John}[rain-is-likely]

Standard Gricean reasoning leads to the inference that, according to the speaker, John does not believe that it is certain to rain, i.e., ¬BEL_{John}[rain is certain]. This inference is compatible with a situation in which John is unsure about whether or not it is certain to rain. By contrast, in the grammaticalist theory, it is possible to append the *O*-operator to the embedded sentence, which leads to the *embedded scalar inference* that, according to the speaker, John believes it is false that rain is certain, i.e., BEL_{John}¬[rain is certain]. The latter inference is logically stronger than the one predicted by the Gricean theory, since it excludes the aforementioned possibility that John is agnostic about the possibility that rain is certain. Many people intuit that the embedded scalar inference is indeed accessible, which poses a problem to the Gricean theory.

In response, it has been countered that the embedded scalar inferences in examples such as (11) can be explained within the Gricean theory with minimal

adjustments. In particular, the embedded scalar inference of (11) logically follows from the scalar inference predicted by the Gricean theory if it is assumed that John knows whether or not it is certain to rain, i.e., BEL_{John}[rain is certain] v $\text{BEL}_{\text{John}}\neg$[rain is certain]. Together with the Gricean inference that John does not believe it is certain to rain, the assumption about John's competence entails that he believes that rain is not certain (e.g., Greenhall, 2008; Russell, 2006; Spector, 2006).

Other examples of so-called embedded scalar inferences can be explained in terms of narrowing. A case in point is (12) from Chierchia et al. (2012).

(12) If you take salad OR desert, you pay $20; but if you take BOTH there is a surcharge.

Here, 'or' is interpreted exclusively within the antecedent of a conditional. However, this interpretation requires prosodic emphasis on 'or' and an explicit contrast with 'both'. These markings are indicative of narrowing. According to proponents of the Gricean theory, narrowing draws on a different mechanism from regular scalar inferences.

At the same time, not all purported instances of embedded scalar inferences can be dismissed so easily. Particularly challenging are cases in which scalar expressions occur in the scope of quantifiers, such as the sentences in (13). Here, the quantifiers in the formalisations range over days of the week.

(13) a. Every day of the week, it is likely to rain. = $\forall x$[rain is likely on x]
b. Exactly one day of the week, it is likely to rain. = $\exists!x$[rain is likely on x]
c. On no day of the week, it is likely to rain. = $\neg\exists x$[rain is likely on x]

Consider (13a). Standard Gricean reasoning leads to the inference that, according to the speaker, it is not the case that it is certain to rain on every day of the week, i.e., $\neg\forall x$[rain is certain on x]. By contrast, the grammaticalist theory allows for the *O*-operator to be

appended to the embedded sentence ‘it is likely to rain’, which results in the logically stronger inference that it is not certain to rain on *any* day of the week, i.e., ∀x¬[rain is certain on x].

Next, consider (13b). According to the Gricean theory, this sentence may imply that on exactly one day, rain is likely but not certain, while on all other days it is unlikely to rain, i.e., ∃!x[rain is likely on x] & ¬∃!x[rain is certain on x] (Geurts, 2010, p. 172ff.). By contrast, if within the grammaticalist theory the *O*-operator is appended to the embedded sentence, the sentence implies that on exactly one day, rain is likely but not certain, while on all others days it is *either unlikely or certain* to rain, i.e., ∃!x[rain is likely on x & ¬[rain is certain on x]]. This interpretation is logically weaker than the interpretation predicted by the Gricean theory, and logically independent from the literal meaning, since there are situations that verify the literal meaning but not the interpretation with an embedded scalar inference (e.g., a situation in which it is certain to rain on Monday, and unlikely to rain on all other days), and vice versa (e.g., a situation in which it is likely but not certain to rain on Monday, and certain to rain on all other days).

Finally, consider (13c). The Gricean theory does not predict any scalar inferences here, since the scalar expression occurs in a downward-entailing environment. By contrast, if within the grammaticalist theory the *O*-operator is appended to the embedded sentence, the sentence implies that there is no day on which rain is likely but not certain, so that on all days of the week rain is either unlikely or certain, i.e., ¬∃x[rain is likely on x & ¬[rain is certain on x]]. This interpretation is logically weaker than the literal meaning since there are situations that verify the interpretation with an embedded scalar inference but not the literal meaning (e.g., a situation in which it is certain to rain on every day of the week), but not vice versa.

All current versions of the grammaticalist theory agree that the embedded scalar inference in (13a) is accessible, since this results in a logically stronger interpretation. Whether or not embedded scalar inferences are predicted in the other two examples depends on the constraints that are placed on the distribution of the *O*-operator (see Section 3). By contrast, the Gricean theory holds that, in all cases shown in (13), embedded scalar inferences are restricted to cases of narrowing, as illustrated in (14).

(14) Every day of THIS week, it is LIKELY to rain. But every day of NEXT week, it is CERTAIN to rain.

Many experimental studies have tested the predictions of the two theories, focusing on the scalar expressions ‘some’ and ‘or’ (e.g., Chemla & Spector, 2011; Geurts & Pouscoulous, 2009; Gotzner & Benz, 2018; Potts et al., 2016; van Tiel et al., 2018). To illustrate the main conclusions from this line of research, I focus on studies that make use of the *sentence-picture verification task*. In this type of task, participants are presented with a sentence and a picture, and they have to decide whether or not (or, sometimes, to what extent) the sentence provides an adequate description of the picture. The underlying idea is that if the sentence triggers a particular inference, and if this inference is not satisfied by the picture, participants should reject the sentence.

[INSERT FIG. 1 ABOUT HERE]

Fig. 1 shows two example trials from Geurts and Pouscoulous (2009, Exp. 3). The sentence with ‘all’ is associated with the embedded scalar inference that none of the squares are connected to all of the circles. This inference is violated by the picture, since the topmost square is in fact connected to all of the circles. Deriving the embedded scalar inference of the sentence with ‘exactly two’ leads to the conclusion that there are exactly two squares that are connected to some but not all of the circles. The picture verifies this interpretation, since exactly two squares are connected to some but not all of the circles, while the others are connected to either none or all of them. Hence, participants who compute the embedded scalar inference are expected to reject the sentence with ‘all’ and accept the sentence with ‘exactly two’. Table 2 shows the rates of responses that are indicative of embedded scalar inferences in a representative sample of four studies using the sentence-picture verification task.

[INSERT TABLE 2 ABOUT HERE]

As Table 2 illustrates, embedded scalar inferences were generally found to be quite weak: for the most part, the proportion of responses indicative of embedded scalar inferences fell well below 50%. For example, returning to Fig. 1, Geurts and Pouscoulous (2009, Exp. 3) found that none of their participants rejected the sentence with 'all', and none of them accepted the sentence with 'exactly two'. This speaks against the grammaticalist idea that embedded scalar inferences occur "freely and systematically" (Chierchia et al., 2012, p. 2297).

A second conclusion is that the rates of embedded scalar inferences depended to a large extent on the embedding quantifying expression: they were the lowest for 'every' (or 'all') and the highest for 'exactly one'. This observation is striking, since all versions of the grammaticalist theory agree that embedded scalar inferences should occur in the former case, while there is disagreement about their occurrence in the latter case, since in that case the resulting interpretation is logically independent from the literal meaning.

One relevant observation here is that 'exactly one' is inherently contrastive in that it highlights both a witness set (i.e., a unique individual that satisfies the predicate) and a complement set (the remaining individuals that do not). From a Gricean perspective, it is tempting to suppose that this inherent contrast facilitates a narrowed reading of the scalar expression. In line with this idea, van Tiel et al. (2018) report that embedded scalar inferences in the case of 'exactly one' are particularly robust if there is a salient visual contrast within the displays, e.g., if the squares that are connected to some but not all of the circles are visually clearly set apart from the squares that are connected to all of them.

A third conclusion that can be drawn from the experimental literature is that the rates of embedded scalar inferences vary not just across quantifying expressions but also across studies. Again, this variability can be explained, at least in part, on the basis of experimental factors that highlight visual contrasts. For example, Chemla and Spector (2011) presented participants repeatedly with the same sentences and images.

Based on a close analysis of their data, Geurts and van Tiel (2013) report that participants were initially unlikely to derive the embedded scalar inference of sentences with ‘all’, but became significantly more likely to do so after repeated presentation. Again, this observation suggests that embedded scalar inferences require a salient contrast.

A final conclusion is that there is a marked discrepancy between embedded and unembedded scalar inferences. To illustrate, Geurts and Pouscoulous (2009, Exp. 1) carried out an *inference task*. In this task, participants were presented with items consisting of a sentence and a candidate inference. They had to decide whether the inference followed from the sentence. Two example items are given in (15) and (16).

(15) Fred heard some of the Verdi operas.
?⇝ Fred did not hear all of the Verdi operas.

(16) All students heard some of the Verdi operas.
?⇝ None of the students heard them all.

Geurts and Pouscoulous found that over 90% of their participants endorsed the unembedded scalar inference in (15). By contrast, the embedded scalar inference in (16) was endorsed by only 27% of the participants. This finding indicates an asymmetry in the robustness of embedded and unembedded scalar inferences.

Some researchers have concluded from the observation that embedded scalar inferences are relatively uncommon in sentence-picture verification tasks that other, “more sensitive” experimental tasks are needed to detect them (Clifton & Dube, 2010; Franke et al., 2017; Gotzner & Benz, 2018). For example, rather than asking whether a sentence is an adequate description of a picture, Clifton and Dube (2010) asked participants whether they prefer a sentence such as the one with ‘all’ in Fig. 1 to describe a picture that satisfies the embedded scalar inference or a picture that does not. Van Tiel (2014) criticises this approach on the count that a preference for a picture should not be equated to evidence for a genuine inference. Whatever the merits of this

argument, it is telling that researchers feel the need to abandon the sentence-picture verifiction paradigm that has been so central in the experimental study of linguistic meaning (e.g., Clark & Chase, 1972; Crain & Thornton, 1998).

In summary, the experimental literature has shown that embedded scalar inferences are generally weak in quantificational environments, and that there is a marked empirical asymmetry between embedded and unembedded scalar inferences. How do these two conclusions bear on the predictions made by the Gricean and grammaticalist theories? Recall that, according to the Gricean theory, embedded scalar inferences should be marked relative to unembedded scalar inferences. By contrast, the grammaticalist theory predicts that embedded scalar inferences should be systematically available, at least when the scalar expression is embedded under an upward-entailing quantifier such as 'all'. The asymmetry between embedded and unembedded scalar inferences, as well as the sensitivity of embedded scalar inferences to semantic or perceptual factors that reinforce contrasts, appears to be more in line with the Gricean theory than with the grammaticalist theory.

Initially, the issue of embedded scalar inferences stood at the forefront of the debate between the Gricean and grammaticalist theories. However, for a while now, research on embedded scalar inferences has come to a standstill. Nonetheless, there is still ample room for research on the topic. Two issues in particular stand in need of further investigation. First, it will be interesting to see whether current results generalise to other embedding environments. While Geurts and Pouscoulous (2009) investigated a relatively wide array of quantifiers, the sample has become more restricted since then. Especially given the sensitivity of embedded scalar inferences to properties of the embedding quantifiers (e.g., the inherent contrastiveness of 'exactly one'), it will be interesting to see whether current results generalise to other embedding environments. For example, based on Table 2, one may expect a difference between 'exactly one' and 'exactly two', which would reinforce the idea that contrast is needed for embedded scalar inferences. Second, almost all of the literature on embedded scalar inferences has been concerned with the scalar expressions 'some' and 'or'. It will be interesting to see whether current results generalise to other scalar expressions (e.g., van Tiel et al., 2016).

4.2. Obligatory scalar inferences

To illustrate the problem of obligatory scalar inferences, consider (17).

(17) ?Some triangles have three sides.

It is common knowledge that all triangles have three sides. Nonetheless, the sentence triggers the scalar inference that not all triangles have three sides, which potentially results in oddness or infelicity when uttered out of the blue. Hence, the scalar inference is obligatory in that it is derived even though it conflicts with what is thought to be common knowledge (e.g., Magri, 2009).

It is not immediately obvious why obligatory scalar inferences would be problematic for the Gricean theory. Indeed, sentences such as (17) have been a staple of experimental research on scalar inferences for a long time, and they were never considered intrinsically problematic to the Gricean theory (e.g., Bott & Noveck, 2004). Indeed, on closer analysis, there is a perfectly Gricean explanation for the infelicity of sentences such as (17).

Recall that, according to the Gricean theory, utterance interpretation involves trying to establish the speaker's motivation for producing a certain utterance. In some cases, there are multiple plausible explanations. For example, in (18)—from van Kuppevelt (1996, p. 402)—A's motivation for saying they have two children can be either (i) to convey that they are eligible for the state benefit, or (ii) to inform B of how many children they have.

(18) B: If you have at least two children, you get a fixed amount of state benefit.
A: I have two children. In fact, I have four.

The cancellation of the scalar inference from ‘two’ to ‘no more than two’ essentially serves to rule out the second explanation, and to indicate that the speaker said they have two children to indicate that they qualify for the state benefit.

By contrast, in the case of (19), A does not have a good reason for saying they have two children if they have four, given that the goal of the conversation is unambiguously to establish how many children A has. As a consequence, A’s cancellation of the scalar inference becomes infelicitous.

(19) B: How many children do you have?
A: I have two children. ?In fact, I have four.

Analogously, in the case of (17) above, the speaker does not have a good reason for saying ‘some’ instead of ‘all’ (or using the shorter ‘Triangles have three sides’). The only viable explanation is that they unreasonably believe that not all triangles have three sides. Hence, the scalar inference cannot be made to disappear, even if it clashes with what was assumed to be common knowledge.

At a more general level, conversational implicatures vary dramatically in terms of their robustness, ranging from fickle and associative to being on a par with entailments. This variability is predicted by the Gricean idea that utterance interpretation involves explaining the speaker’s behaviour, since explanations of other people’s behaviour also range from tentative to nearly unavoidable.

In summary, in spite of Magri’s claims to the contrary, obligatory scalar inferences can be straightforwardly accounted for within the Gricean theory.

4.3. Non-cooperative contexts

Fox (2014) argues that the Gricean theory is incapable of explaining scalar inferences in situations in which the interlocutors’ goals are in conflict. To make this concrete, he sketches the following scenario: imagine a TV show which revolves around 100 boxes

whose content is hidden. Five of these boxes contain a million dollars; the others are empty. Contestants have to choose a box, and they win whatever is inside. Imagine in this scenario that the host of the TV show utters (20).

(20) There is a prize in box 20 or 25.

According to Fox, one would naturally infer that there is no prize in both boxes. He argues that this scalar inference from 'or' to 'not both' cannot be explained by the Gricean theory, since in the scenario at hand the assumption that the host should be informative (i.e., the Maxim of Quantity) is "deactivated". In other words, the host is not expected to be maximally informative as this would defeat the purpose of the TV show.

By contrast, Fox argues that the exclusivity inference can be explained within the grammaticalist theory, since it does not rely on any pragmatic maxims. Instead, the *O*-operator is blindly inserted into the syntactic structure of (20), and the sentence comes out as implying that there is no prize in both boxes.

Clearly, Fox is right that the scalar inference of (20) cannot be premised on the assumption that, if the host believed that both boxes contain a prize, they should have said so. However, the particular context of a competitive TV show supplies another assumption that can serve as premise in the argument, namely that the host should leave the candidate guessing. Since this assumption is common knowledge, it may be inferred that there is no prize in both of the boxes (see Westera, 2022, p. 26 for a similar suggestion).

In line with this explanation, it would be very odd for the host to cancel the scalar inference by continuing with 'In fact, there is a prize in both boxes'. Similarly, it would be odd for the host to utter (21).

(21) ?There is a prize in box 20.

In both cases, the host's behaviour would be at odds with the purpose of the TV show. Hence, the scalar inference of (20) can be derived within the Gricean theory, but based on different assumptions about how the speaker should behave than the conversational maxims that Grice formulated in 'Logic and conversation'.

At a more general level, Grice's maxims hold when the purpose of a talk exchange is to ensure a "maximally effective exchange of information" (Grice, 1975, p. 47). However, in many talk exchanges, this is not the intended purpose. In contrast with Fox's claim, it does not follow that, in such talk exchanges, the maxims are "deactivated", but rather, different maxims hold that may give rise to different non-conventional implicatures.[1] Aside from TV shows, this holds for, e.g., witness interrogations (cf. Asher & Lascarides, 2013; van Tiel & Geurts, 2025), rap battles, small talk, and job interviews. In none of these cases can the purpose of the talk exchange be characterised as a maximally effective exchange of information (e.g., an applicant is not expected to extensively discuss their negative qualities). Nonetheless, these exchanges are governed by norms that interlocutors are expected to follow, and whose (apparent) non-adherence may trigger certain inferences. To further illustrate, consider the exchange in (22).

(22) A: How are you doing?
B: I'm fine.

If A and B are close friends, A may infer from B's answer that nothing of importance happened to B recently (e.g., they were not fired from work). However, if this exchange happened during small talk, this inference would not be valid since—like game shows on TV—small talk is not characterised by maximally effective exchange of information.

[1] A reviewer suggests, alternatively, to take seriously the rider on Grice's formulation of the Maxim of Quantity as requiring the speaker to make their contribution optimally informative *given* what is "required (for the purpose of the exchange)". In the case of the TV show, the purpose of the exchange is to allow the candidate to make a consequential choice, which licenses the inference that there is no prize in both of the mentioned boxes.

In summary, Fox's argument relies on a simplified view of the Gricean theory, according to which the conversational maxims either hold or are deactivated. We have seen that there are situations in which different maxims hold, depending on the purpose of the exchange. Further research is needed to flesh out this proposal in more detail. In particular, it seems to be the case that, even in non-cooperative contexts, interlocutors tend to operate under the assumption that communication is aimed at a maximally efficient exchange of information (e.g., Cummins, 2025; Cummins & Franke, 2021). In any case, it is clear that the Gricean theory provides a conceptual apparatus that is capable of explaining scalar inferences in non-cooperative contexs.

4.4. Universal free choice inferences

The final two arguments against the Gricean theory center around the word 'or'. It is often assumed that 'or' is equivalent to the inclusive disjunction from propositional logic (e.g., Gamut, 1991). If this view is correct, sentences with 'or' should fail to license conclusions about the truth of the individual disjuncts. In line with this prediction, it cannot be inferred from (23) that Kim ate pizza, nor can it be inferred that Kim ate pasta, though it can be inferred that they had one of the two.

(23) Kim ate pizza or pasta.

While this characterisation is accurate for (23), it does not generalise across all occurrences of 'or'. For example, the standard view makes the wrong predictions for sentences in which 'or' is embedded under an existential modal quantifier. Thus, (24) implies that you are allowed to eat pizza, and that you are allowed to eat pasta (e.g., Kamp, 1973). These inferences are called *free choice inferences*.

(24) You are allowed to eat pizza or pasta.

The problem of free choice inferences has provoked a wide range of possible solutions. Some of these locate the source of free choice inferences in the semantics of the modal quantifier (e.g., Barker, 2010) or the lexical item 'or' (e.g., Zimmermann, 2000). Others explain free choice inferences as a variety of scalar inferences (e.g., Fox, 2007; Kratzer & Shimoyama, 2002).

To illustrate the latter explanation, a speaker who says (24) could have been more informative by uttering one of the sentences in (25).[2]

(25) a. You are allowed to eat pizza.
b. You are allowed to eat pasta.

These sentences come with the exhaustivity inferences that you are not allowed to eat pasta (25a) or that you are not allowed to eat pizza (25b). Why did the speaker not produce either of these alternatives? Presumably because they do not believe that you are allowed to eat pizza but not pasta, and they do not believe that you are allowed to eat pasta but not pizza. When these inferences are combined with the literal meaning of (24), it follows that I am allowed to eat pizza, and that I am allowed to eat pasta.

Free choice inferences such as those triggered by (24) can be accounted for within both the Gricean and grammaticalist theory. However, this changes if (24) is embedded, e.g., under a universal quantifier, as in (26).

(26) Everyone is allowed to eat pizza or pasta.

As Chemla (2009) shows experimentally, (26) triggers the inferences that everyone is allowed to choose between eating pizza and pasta (see also van Tiel, 2012). These

[2] For perspicuity and reasons of space, I here ignore the stronger alternatives where 'allowed' is replaced with its scalemate 'required'. See, e.g., Fox (2007) for more technical discussion.

universal free choice inferences cannot straightforwardly be accounted for by the Gricean theory. To illustrate, consider the alternatives in (27).

(27) a. Everyone is allowed to eat pizza.
b. Everyone is allowed to eat pasta.

An utterance of (27a) may imply that no one is allowed to eat pasta; an utterance of (27b) that no one is allowed to eat pizza. Consequently, someone who utters (26) may imply that it is not the case that everyone is allowed to eat pizza but not pasta, and that it is not the case that everyone is allowed to eat pasta but not pizza. Crucially, these inferences are compatible with a situation in which some people are allowed to eat pizza but not pasta, while everyone else is allowed to eat pasta but not pizza. Hence, these inferences fail to license the observed conclusion that everyone is allowed to choose between having pizza or pasta.

By contrast, universal free choice inferences can be accounted for in the grammaticalist theory by attaching the required *O*-operators to the embedded sentence (see Fox, 2007, for technical details). Hence, universal free choice inferences appear to provide a strong argument in favour of the grammaticalist theory.

At the same time, this argument should be interpreted with some caution. In particular, it is possible to derive universal free choice inferences within the Gricean theory if a homogeneity assumption is in place, e.g., in the case of (26) if it is assumed that everyone must have the same food options. If universal free choice inferences are indeed premised on such homogeneity assumptions, one would expect that they behave more like "default" inferences (e.g., the inference from "He entered the house" that he entered through the door rather than the chimney) than like genuine conversational implicatures. That is, one might expect that they can easily be overridden. Thus, it will be interesting to see whether participants accept (26) as a description of a situation in which some people are allowed to eat pizza but not pasta, while everyone else is allowed to eat pasta but not pizza. If so, this would support the

idea that universal free choice inferences are different in nature from other types of scalar inferences, and that they rely on defeasible background assumptions.

It should also be noted that there are compelling observations suggesting that free choice inferences are not a variety of scalar inferences in the first place. Perhaps the most prominent observation involves 'or' taking wide scope, as in (28).

(28) You are allowed to eat pizza, or you are allowed to eat pasta.

This sentence also tends to imply that you are allowed to choose between eating pizza and pasta. However, in this case, the free choice inferences cannot be straightforwardly accounted for within either the Gricean or grammaticalist theory, because the sentence competes, inter alia, with the more informative alternative (29), and negating (29) is tantamount to negating the desired free choice inferences.

(29) You are allowed to eat pizza, and you are allowed to eat pasta.

In summary, universal free choice inferences pose an interesting challenge to the Gricean theory, but further investigation is needed to uncover just how serious the challenge really is.

4.5. Hurford's constraint

In a short article, Hurford (1974) observed that, generally, sentences of the form 'A or B' are infelicitous if 'A' asymetrically entails 'B', as illustrated by the infelicity of (30).

(30) ?John lives in Paris or France.

However, as Hurford observes, this generalisation does not apply if 'A' contains a scalar expression and 'B' its stronger scalemate. Thus, even though the disjuncts in (31) stand in an asymmetric entailment relation, the sentence is felicitous.

(31) It is likely or certain to rain.

Chierchia et al. (2012) argue that this descriptive pattern can only be explained within the grammaticalist theory. According to this theory, the *O*-operator can be appended to the disjunct 'it is likely to rain', which thus comes to mean 'it is likely but not certain to rain'. Once the operator is appended, the sentence is not longer an exception to Hurford's generalisation, since the scalar inference of the first disjunct breaks the entailment relation. Within the Gricean theory, it is not possible to derive a scalar inference on the basis of first disjunct, since the speaker never said that it is likely to rain.

Of course, this explanation is premised on the assumption that Hurford's generalisation is a *constraint* on the felicitous use of 'or'. Given this assumption, the grammaticalist theory accounts for the felicity of apparent counterexamples to this constraint, such as (31). However, as Westera (2020) points out, it is not necessary to view Hurford's generalisation as a proper constraint. If this assumption is dropped, the challenge shifts from explaining why sentences such as (31) are felicitous to explaining why sentences such as (30) are *in*felicitous.

Westera addresses this challenge based on a different generalisation about the use of 'or', namely that for a sentence of the form 'A or B' to be felicitous, both 'A' and 'B' should be interpretable as *relevant alternatives*, i.e., as suitable answers to a single accessible question under discussion (e.g., Simons, 2001). To illustrate, (31) from Simons (2001) is infelicitous in an out-of-the-blue context because the two disjuncts are not readily interpretable as answers to an accessible question under discussion.

(32) ?Either there is dirt in the fuel line, or it is raining in Tel-Aviv.

As Simons points out, this sentence may become felicitous if such a question under discussion is construed. For example, imagine that the car in question is in Jerusalem, and that rain in Tel-Aviv causes humidity in Jerusalem, which occasionally causes similar car problems as dirt in the fuel line. In that—somewhat contrived—situation, (32) may be felicitously uttered to answer the question 'Why did the car break down?'.

An interesting correlate of the relatedness constraint is that disjuncts should provide answers at the same *level of granularity*. Hence, sentences such as those in (33) are infelicitous because the two disjuncts answer the same question under discussion (e.g., 'What is the temperature outside') but at different levels of granularity.

(33) a. ?It is either warm or 3.2 degrees Celsius.
b. ?John is either tall or 1 meter and 63 centimeters.
c. ?There were either hundreds of people or 31 people.

Note that these sentences do not violate Hurford's generalisation, since the disjuncts do not stand in an entailment relation. Hence, their infelicity cannot be explained in terms of the derivation of scalar inferences.

Westera argues that sentences such as (30) above are infelicitous for the same reason as the sentences in (33), namely because the disjuncts provide answers at different levels of granularity, e.g., either at the city-level or at the country-level. By contrast, in the case of (31), the disjuncts provide answers at the same level of granularity, and the sentence is therefore felicitous.

According to Westera's explanation, Hurford's generalisation is in fact a corollary of the more general relatedness constraint, which requires disjuncts to provide competing answers to the same question under discussion. Crucially, this explanation does not invoke any embedded scalar inferences, and is thus congenial to the Gricean theory. In this discussion, we have ignored several important issues (e.g., concerning more complex disjunctions, the role of the order of the disjuncts, and the status of levels

of granularity), but overall these, too, do not appear to fall outside of the purview of the Gricean theory (see, e.g., Krifka, 2024 and Westera, 2020 for more discussion).

5. Conclusion

Scalar inferences occupy a revealing position at the semantics–pragmatics interface. Both the Gricean and the grammaticalist theories aim to explain how hearers enrich sentence meaning beyond what is literally encoded, but they differ in where they locate the underlying mechanism: in reasoning about cooperative communication on the one hand, and in the grammatical architecture of meaning composition on the other. The comparison across the five domains discussed in this chapter—embedded scalar inferences, obligatory scalar inferences, scalar inferences in non-cooperative contexts, free choice inferences, and Hurford's constraint—shows that neither approach has a clear empirical or theoretical monopoly. In some areas, the Gricean theory offers a parsimonious and independently motivated account, while in others, the grammaticalist theory appears better positioned to capture local effects.

The debate between the Gricean and grammaticalist theories exemplifies a more general disagreement about the division of labour between semantics and pragmatics. The two theories agree that both semantic and pragmatic factors modulate the interpretation of scalar expressions. On the one hand, the Gricean theory assigns primacy to pragmatic processes, relegating semantics to cases of narrowing where scalar expressions are marked prosodically or contextually. On the other hand, the grammaticalist theory foregrounds semantic factors, consigning pragmatics to explaining ignorance inferences, such as the inference from 'It is likely to rain' that the speaker does not believe (rather than believing it is false) that rain is certain.

Ultimately, scalar inferences continue to serve as a valuable proving ground for interface theories because they make visible the boundary between semantics and pragmatics. By studying when scalar inferences arise, when they do not, and how they interact with structure and context, we may thus better understand how semantics and pragmatics cooperate to shape meaning. Far from closing the debate, the comparison between Gricean and grammaticalist approaches points toward a more integrative

perspective in which both grammar and pragmatic reasoning play an essential role in explaining how speakers convey—and hearers recover—enriched interpretations.

6. Bibliography

Asher, N., & Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, *6*, 1–62.

Atlas, J. D., & Levinson, S. C. (1981). It-clefts, informativeness and logical form: Radical pragmatics (revised standard version). In P. Cole (Ed.), *Radical pragmatics* (pp. 1–62). Academic Press.

Barker, C. (2010). Free choice permission as resource-sensitive reasoning. *Semantics and Pragmatics*, *3*, 1–38.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457.

Burton-Roberts, N. (1984). Modality and implicature. *Linguistics and Philosophy*, *7*, 181–206.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory* (pp. 179–238). Benjamins.

Chemla, E. (2009). Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics*, *2*, 1–33.

Chemla, E., & Spector, B. (2011). Experimental evidence for embedded implicatures. *Journal of Semantics*, *28*, 359–400.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39–103). Oxford University Press.

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

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 2297–2332). Mouton de Gruyter.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517.

Clifton, C., & Dube, C. (2010). Embedded implicatures observed: A comment on Geurts and Pouscoulous (2009). *Semantics and Pragmatics*, *3*, 1–13.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. MIT Press.

Cummins, C. (2025). Noncooperative communication. *Annual Review of Linguistics*, *11*, 35–52.

Cummins, C., & Franke, M. (2021). Rational interpretation of numerical quantity in argumentative contexts. *Frontiers in Communication*, *6*, 662027.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). Palgrave.

Fox, D. (2014). Cancelling the maxim of Quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics*, *7*, 1–20.

Frank, M. C., & Goodman, N. D. (2012). Prediction pragmatic reasoning in language games. *Science*, *336*, 998.

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

Franke, M., Schlotterbeck, F., & Augurzky, P. (2017). Embedded scalars, preferred readings and prosody: An experimental revisit. *Journal of Semantics*, *34*, 153–199.

Gamut, L. T. F. (1991). *Logic, language, and meaning. Volume I: Introduction to logic*. University of Chicago Press.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. Academic Press.

Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and Pragmatics*, *2*, 1–34.

Geurts, B., & van Tiel, B. (2013). Embedded scalars. *Semantics and Pragmatics*, *6*, 1–37.

Gotzner, N., & Benz, A. (2018). The best response paradigm: A new approach to test implicatures of complex sentences. *Frontiers in Communication*, *2*, 21.

Gotzner, N., Solt, S., & Benz, A. (2017). Scalar diversity, negative strengthening and adjectival semantics. *Frontiers in Psychology*, *3*, 34.

Greenhall, O. (2008). Against Chierchia's computational account of scalar implicatures. *Proceedings of the Aristotelian Society*, *108*, 373–384.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics. Volume 3: Speech acts* (pp. 41–58). Academic Press.

Hirschberg, J. B. (1991). *A theory of scalar implicature* (Unpublished doctoral dissertation). University of Pennsylvania.

Horn, L. R. (1972). *On the semantic properties of logical operators in English* (Unpublished doctoral dissertation). University of California, Los Angeles. (Distributed by the Indiana University Linguistics Club)

Horn, L. R. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, *61*, 121–174.

Horn, L. R. (1989). *A natural history of negation*. University of Chicago Press.

Hu, J., Levy, R., Degen, J., & Schuster, S. (2023). Expectations over unspoken alternatives predict pragmatic inferences. *Transactions of the Association for Computational Linguistics*, *11*, 885–901.

Hurford, J. (1974). Exclusive or inclusive disjunction. *Foundations of Language*, *11*, 409–411.

Jennings, R. E. (1994). The ‘or’ of free choice permission. *Topoi*, *13*, 3–10.

Kamp, H. (1973). Free choice permission. *Proceedings of the Aristotelian Society*, *74*, 57–74.

Kratzer, A., & Shimoyama, J. (2002). Indeterminate pronouns: The view from Japanese. In Y. Otsu (Ed.), *The proceedings of the third Tokyo Conference on Psycholinguistics* (pp. 1–25). Hituzi Syobo.

Krifka, M. (2024). Hurford’s constraint and disjunctions over speech acts. *Proceedings of the Amsterdam Colloquium* (p. 226–233). https://platform.openjournals.nl/PAC/article/view/21843

Landman, F. (1998). Plurals and maximalization. In S. Rothstein (Ed.), *Events and grammar* (pp. 237–271). Kluwer.

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicature. *Natural Language Semantics*, *17*, 245–297.

Matsumoto, Y. (1995). The conversational condition on Horn scales. *Linguistics and Philosophy*, *18*, 21–60.

Mayol, L., & Castroviejo, E. (2013). How to cancel an implicature. *Journal of Pragmatics*, *50*, 84–104.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguistics and Philosophy*, *15*, 509–559.

Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of ‘scalar inferences’. In N. Burton-Roberts (Ed.), *Advances in pragmatics* (pp. 184–212). Palgrave.

Pankratz, E., & van Tiel, B. (2021). The role of relevance for scalar diversity: A usage-based approach. *Language and Cognition*, *13*, 562–594.

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2016). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, *33*, 755–802.

Recanati, F. (1989). The pragmatics of what is said. *Mind & Language*, *4*, 295–329.

Ronai, E., & Xiang, M. (2024). What could have been said? Alternatives and variability in pragmatic inferences. *Journal of Memory and Language*, *136*, 104507.

Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics*, *23*, 361–382.

Sadock, J. (1978). On testing for conversational implicature. In P. Cole (Ed.), *Syntax and semantics. Volume 9: Pragmatics* (pp. 281–297). Academic Press.

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

Sauerland, U. (2010). Embedded implicatures and experimental constraints: A reply to Geurts & Pouscoulous and Chemla. *Semantics and Pragmatics*, *3*, 1–13.

Simons, M. (2001). Disjunction and alternativeness. *Linguistics and Philosophy*, *24*, 597–619.

Singh, R. (2008). On the interpretation of disjunction: Asymmetric, incremental, and eager for inconsistency. *Linguistics and Philosophy*, *31*, 245–260.

Soames, S. (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry*, *13*, 483–545.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques* (Unpublished doctoral dissertation). University of Paris VII.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (2nd ed.). Blackwell.

Sun, C., Tian, Y., & Breheny, R. (2018). A link between local enrichment and scalar diversity. *Frontiers in Psychology*, *9*, 2092.

van Kuppevelt, J. (1996). Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy*, *19*, 393–443.

van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information*, *13*, 491–519.

van Tiel, B. (2012). Universal free choice? In A. Aguilar Guevara, A. Chernilovskaya, & R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung 16* (pp. 627–638). MIT Working Papers in Linguistics.

van Tiel, B. (2014). Embedded scalars and typicality. *Journal of Semantics*, *31*, 147–177.

van Tiel, B., Noveck, I., & Kissine, M. (2018). Reasoning with 'some'. *Journal of Semantics*, *35*, 757–797.

van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, *33*, 137–175.

Weiner, M. (2006). Are all conversational implicatures cancellable? *Analysis*, *66*, 127–130.

Westera, M. (2020). *Hurford disjunctions: An in-depth comparison of the grammatical and the pragmatic approach*. SemanticsArchive. https://semanticsarchive.net/Archive/jk5YzYyO/

Westera, M. (2022). Attentional pragmatics: A new pragmatic approach to exhaustivity. *Semantics and Pragmatics*, *15*, 10.

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, *8*, 255–290.

Adjective		
<big, huge>	<happy, ecstatic>	<likely, certain>
<low, empty>	<okay, great>	<tough, impossible>
<ugly, hideous>	<warm, hot>	<wet, sopping>
Verb		
<continue, intensify>	<damage, destroy>	<eat, devour>
<imply, say>	<match, surpass>	<praise, extol>
<push, shove>	<stabilise, reduce>	<tolerate, enjoy>
Noun		
<discomfort, illness>	<envy, resentment>	<hunger, famine>
<increase, surge>	<precision, rigidity>	<respect, reverence>
<shock, disbelief>	<surprise, shock>	<suspicion, conviction>

Table 1: Examples of lexical scales.

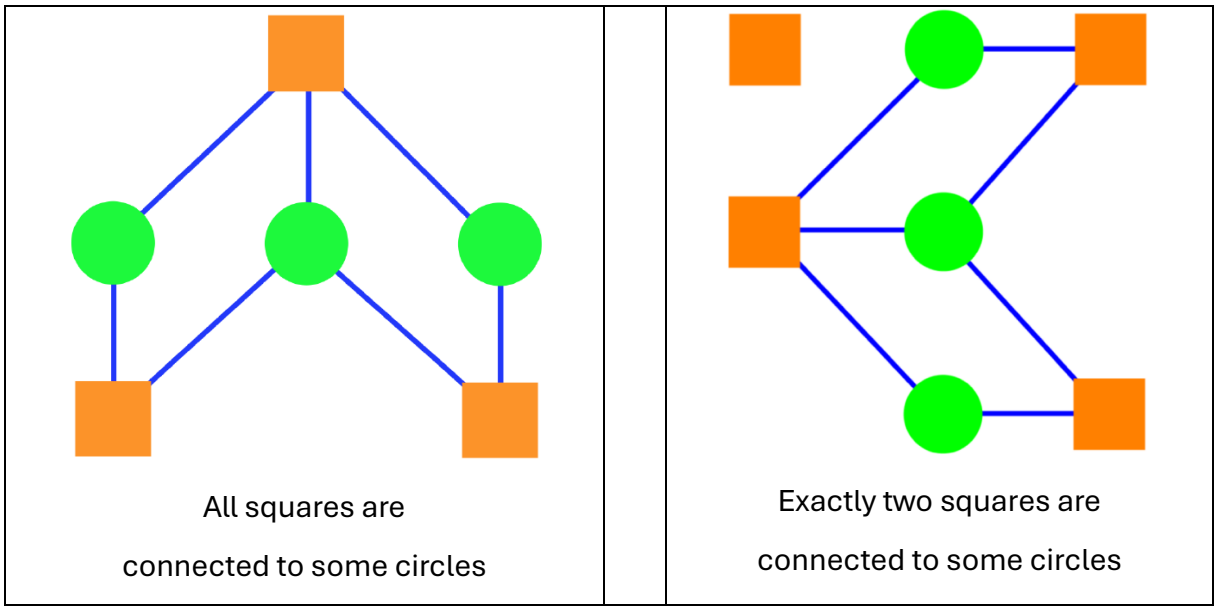

Figure 1: Example items from Geurts and Pouscoulous (2009, Exp. 3).

Study	*Embedding*	*%*
Geurts and Pouscoulous (2009, Exp. 3)	'all'	0
	'more than one'	0
	'exactly two'	0
	'not all'	0
	'not more than one'	0
Chemla and Spector (2011)	'every'	31
	'exactly one'	73
Potts et al. (2016, Exp. 1)	'every'	8
	'exactly one'	55
	'no'	29
Van Tiel et al. (2018, Exp. 1)	'every'	14
	'exactly one'	30
	'no'	25

Table 2: Rates of responses that are indicative of embedded scalar inferences in four studies using the sentence-picture verification task.