Generating and selecting alternatives for scalar implicature computation: The Alternative Activation Account and other theories

Nicole Gotzner and Radim Lacina (Osnabrück University)

Keywords: Alternatives, focus, scalar implicature, contextual restriction, asymmetric entailment, antonyms

1. Introduction
Scalar implicature is one of the key phenomena involving reasoning about alternatives. Fifty years of research has looked at the mechanisms underlying scalar implicature computation focusing on whether those are pragmatic or semantic in nature (see Sauerland, 2012; Chemla & Singh, 2014 and Gotzner & Romoli, 2022, for an overview). Closely intertwined with this is the key question of the current paper about the nature of alternatives that serve as the basis for scalar implicature computation.

While different in their architecture, most formal theories assume that scalar implicatures involve three computational steps: (1) The computation of the literal meaning of an utterance, (2) the activation of alternatives and (3) the negation of relevant alternatives.

The key controversy in the literature has surrounded step (3) and the question of how relevant alternatives are being negated: either via a Gricean mechanism in which the listeners reason about alternative utterances the speaker could have said (Grice, 1975; Horn, 1972; Sauerland, 2004) or a grammatical exhaustification mechanism (e.g., Chierchia, Fox, Spector, 2012; Fox & Katzir, 2011). According to Horn’s view, scalar implicatures involve dedicated scales of weak and strong expressions that are stored in the lexicon. In the grammatical framework, different constraints on alternatives have been proposed, also allowing for logically independent alternatives (e.g., in the case of embedded implicatures, Chemla & Spector, 2011). Chierchia (2013) and others explicitly incorporate a notion of activated alternatives in their mechanism for scalar implicature but do not spell out the mechanism underlying activated alternatives. Typically, the set of alternatives is assumed to be given or inferable based on the attested scalar implicatures (Trinh, 2019; Gotzner and Romoli, 2022). Thus, little is known about which alternatives are activated and when during the comprehension process.

Inferring relevant or activated alternatives from attested inferences creates a problem of circularity. In this paper, we argue that we need to separate the mechanisms by which alternatives are being selected from the inferential mechanism itself (see also Gotzner, 2017; 2019). To do so, we draw on work from psycholinguistics on the mechanisms of generating and restricting alternative sets. Work on focus processing has integrated theoretical assumptions about alternatives with insights about semantic activation spreading (Gotzner,
2017; Husband & Ferreira, 2015). On this view, generating alternatives also involves domain-general mechanisms that determine which alternatives are activated and become salient. We adapt this so-called Alternative Activation Account for scalar implicature with the following subsequential steps: (1) Domain-general mechanisms determine the salience and lemma-level activation of a broad cohort of expressions. (2) Dedicated grammatical and pragmatic mechanisms constrain which alternatives are used during scalar implicature computation.

Our main argument is that a cohort of alternatives is being activated as a byproduct of how the brain organizes concepts in semantic networks (e.g., Swinney et al., 1979). Only a subset of these elements are spelled out (receive phonological activation) during implicature computation, i.e., are relevant alternatives. We review the experimental literature in line with this view. We specifically focus on different priming studies that tested the activation of scalar and non-entailed alternatives during scalar implicature computation.

Our paper is structured as follows. First, we present the Alternative Activation for focus alternatives. Then, we develop a corresponding account for scalar implicature while discussing the nature of alternatives assumed in mainstream theories. We review psycholinguistic studies on how listeners generate alternatives during language comprehension. Then, we present evidence for scalar diversity in scalar implicature computation. We discuss theoretical proposals which accommodate this variability by assuming that some scales do not activate alternatives but instead involve a different implicature mechanism. Our concluding remarks provide directions for future research, including novel methods and test cases to answer core questions about the nature of alternatives.

2. Theoretical proposals

2.1. Alternative Activation Account for focus

Grammatical accounts of scalar implicature have been built on the basis of Rooth’ alternative semantics (1985; 1992) and several accounts assume the alternatives for focus and scalar implicature to be the same (see especially Fox & Katzir, 2011). In the area of focus, concrete proposals have been made for how alternatives are generated and restricted during real-time processing.

One key finding from psycholinguistic studies is that focus activates alternatives that are of the same semantic type as the focused element (see Gotzner & Spalek, 2019 for an overview). The evidence for this comes from lexical priming where the reaction times for deciding whether a target is a word in a given language is faster when this word is related to a preceding word with focal stress (e.g., Braun & Tagliapietra, 2010; Husband & Ferreira, 2015). For example, when participants are presented with a sentence like Mary ate bananas,
they are faster to react to the target CHERRIES compared to SOCKS. The extent of this priming effect is modulated by focal stress, preceding focus particles like only, the discourse context and verb selectional restrictions and the time that elapses between stimulus presentation. For example, there is evidence that the unrelated word SOCKS does receive some level of activation in a sentence frame that forms an ad hoc category for buyable things, that is Mary bought bananas (Gotzner, 2015; Jördens, et al., 2020).

Husband & Ferreira (2015) spell out psychological mechanisms of activation and selection of relevant alternatives and show that they kick in sequentially during processing. In particular, they found that at the offset of a focused word both semantic associates of BANANA, CHERRIES and YELLOW, become activated while at a later point (750 ms offset) only relevant alternatives remain activated. Concurrently with this literature, Gotzner (2015; 2017) made a proposal in which both domain-general mechanisms as well as grammatical and pragmatic mechanisms serve the generation and restriction of alternative sets. While Husband and Fereira propose psychological mechanisms such as spreading activation and inhibition in a semantic network, Gotzner integrated such mechanisms with grammatical and pragmatic mechanisms assumed in semantics and pragmatics. We call this view the Alternative Activation Account. This account separates a step of initial broad activation spreading from a second step in which only relevant alternatives that are being negated during the inferential process remain:

1. Domain general mechanisms generate a broad set of alternatives including all semantic associates (words/concepts)
   (1) Mary ate [bananas]→ CHERRIES, SOCKS, YELLOW
2. Grammatical and pragmatic mechanisms single out relevant alternatives (negated alternatives)
   (2) Mary ate [bananas]→ CHERRIES

The Alternative Activation Account is supported by a neuroimaging study by Spalek and Oganian (2019) showing that semantic associates activate different brain areas from proper focus alternatives. Specifically, only relevant alternatives activate areas that have been invoked in discourse processing. Since all the aforementioned studies on focus alternatives tested noun alternatives are single words, it is not clear what level of
representation alternatives have. Alternatives could be concepts, lemmas, sub-constituents, phrases or entire utterances. In the following, we will revisit the theoretical literature to make different suggestions about this question for the case of scalar implicature.

2.2. Extending the Alternative Activation Account to Horn scales

Similar to focus, theoretical accounts of scalar implicature assume that more alternatives are initially generated and then restricted. Several theoretical accounts in the grammatical tradition following Fox & Katzir (2011) even assume the alternatives for focus and for scalar implicature to be the same. Thus, we propose to extend the Alternative Activation Account to cases of scalar implicature.

Starting off from the same assumption about online semantic and pragmatic processing, we can straightforwardly extend the Alternative Activation Account to scalars as follows. In the first step, a broad cohort of alternatives is activated including all associates of a scalar expression. Thus, an expression like beautiful should also activate its antonym ugly. Due to the entailment condition (e.g., Horn, 1972), antonyms should be deactivated in step 2 so that only the stronger scale-mates remain as part of the proper set of alternatives to beautiful. That is, only gorgeous should be negated during the inferential process. Thus, the main prediction arising from the Alternative Activation Account is that alternatives beyond scale-mates (of one polarity) should initially be activated as a result of how the brain organizes words in a semantic network.

1. Initial activation of all associates

(3) Mary is beautiful -> UGLY, GORGEOUS

2. Restriction to stronger scale-mates (relevant alternatives)

(4) Mary is beautiful -> GORGEOUS

Let us briefly discuss the nature of relevant alternatives from the view of the theoretical scalar implicature literature. The majority of the literature has assumed that scalar implicatures arise via specialized Horn scales that are ordered with respect to asymmetric entailment (Horn, 1972). These asymmetric entailment relations can be shown with linguistic tests using the scalar particle even:

---

1 A study by Lacina, Sturt and Gotzner (2023) has extended the so-called probe recognition paradigm to verbal alternatives. This study provides evidence that larger constituents become activated in focus processing, however, the study did not find a critical interaction with focus particles. So, it remains unclear whether listeners entertain verbal phrases as part of generating alternatives or as a general by-product of memory retrieval.
(5) He ate some, even all of the cookies
(6) *He ate all, even some of the cookies
(7) *He ate all, even none of the cookies

The tests show that the statement with all entails the one with some but not vice versa. All also does not entail its antonym none and the opposite is not true either. To save the asymmetric entailment condition, negative terms are assumed to be represented on a separate Horn scale (see Hirschberg, 1985). Note also that the ordering relation is reversed for negative scales, for example none is stronger than few. One of the main reasons to postulate entailment scales was the symmetry problem (Kroch 1972; see Breheny et al., 2017 for a recent overview). Symmetry is said to arise when there is a particular expression that has two alternatives whose disjunction it is equivalent to and these alternatives are in contradiction (Fox & Katzir, 2011). Let us exemplify the problem with the some-all scale and its corresponding implicature:

(8) John ate some of the cookies.

Sentence (8) has the implicature that John ate some but not all of the cookies due to the negation of the statement with the alternative all. However, the expression some but not all is in principle a symmetric alternative to some. The problem is that if some but not all is part of the alternative set, a contradiction would follow. The symmetry problem is how to include all as an alternative while excluding the symmetric alternative some but not all. In Horn’s view, symmetry is broken by stipulating that alternatives are lexical scales ordered by asymmetric entailment. This means that some but not all is excluded from the set of relevant alternatives.

Neo-Gricean globalist accounts such as the one by Horn (1972) are globalist in the sense that entire utterances are being negated. For example, to derive a scalar implicature for sentence (8) one would need to negate the sentence John ate all of the cookies. This assumption was challenged by cases of embedded implicatures where a triggering expression embedded under another scalar generates a strengthened reading (Chierchia, 2006). To take an example, a sentence like All girls found some of their marbles, has a reading in which all girls found some but not all of their marbles (e.g., Chemla & Spector, 2011; Potts, Lassiter, Levy & Frank., 2015; Gotzner & Benz, 2018). This reading cannot be derived by globally negating the alternative All girls found all of their marbles in a standard way (but see Chemla & Spector, 2011). As a reaction, Chierchia (2006) and others developed the grammatical view of implicatures, which incorporates implicatures locally at
the triggering site. This account uses sub-constituents as alternatives, for example when encountering some the lexical expression all is triggered and negated by a silent operator (see Potts et al., 2015 and Benz & Gotzner, 2021 for non-grammatical accounts to derive embedded implicatures).

While some versions of the grammatical account still use specialized scales (e.g., Chierchia, 2013), other accounts use alternatives that are constrained by complexity considerations (e.g., Fox and Katzir, 2011; Trinh & Haida, 2015). What is common to grammatical accounts is that they use phrases or sub-constituents as the basis of the inferential process. This view prima facie seems to be more compatible with the incremental nature of language processing. But on a Neo-Gricean globalist account, it is also plausible that listeners initially entertain a broader set of alternatives beyond proper scale-mates during real-time processing. On this view, a later pragmatic process would involve the reasoning about entire alternative utterances the speaker could have used (see for example, Chemla & Singh, 2014 for an overview of this debate). Importantly, all of the existing accounts would exclude antonyms or more complex alternatives from the relevant set of alternatives to avoid the symmetry problem.

To summarize, the main novel prediction arising from the Alternative Activation Account is that alternatives beyond scale-mates (of one polarity) should be activated during the inferential process. This would be compatible with several accounts of scalar implicature, that is accounts of how alternatives are being negated. Hence, experimental work needs to find a way to distinguish activated and relevant alternatives that are negated during pragmatic processing. We will discuss some suggestions for doing so after reviewing the existing body of experimental research on which alternatives become activated during processing.

### 3. Psycholinguistic evidence for the activation of alternatives for scalar implicature

A recent body of research has investigated the nature of alternatives and their underlying mental representation during sentence comprehension. Firstly, we review the evidence on the activation of informationally stronger alternatives that are predicted to be a necessary component of pragmatic inference processes (e.g., Horn, 1972). Next, we turn our attention to the emerging body of literature that aims to test the prediction of the Alternative Activation Account (Gotzner, 2017) that it is not only the stronger, but also the other, weaker alternatives that are active in the process. Additionally, we discuss related research using structural priming and show how it can be related to the questions surrounding the activation of alternatives as well.

#### 3.1. Evidence for the activation of strong terms
De Carvalho et al. (2016) were the first to investigate whether informationally stronger terms are in fact activated during comprehension. Using a masked priming paradigm and a lexical decision task, they presented their participants either weak scalar terms (some) or strong ones (all) as subliminal primes. The targets were either identical (i.e., some - some, all - all) or the other scale-mate (i.e., some - all, all - some). There was also a control condition where a sequence of consonants not comprising an existing word was presented. When their French participants performed the task, they were found to be the quickest in the identical condition and the slowest when the prime was a non-word sequence. This suggested that the experimental subliminal priming set up of de Carvalho and colleagues (2016) in fact impacted people’s lexical decision times. Crucially, they found a difference between the two scalar word prime conditions. It turned out that weak scalar word primes activated their stronger scale-mate targets more than the latter used as primes did for the weak. In other words, de Carvalho et al. (2016) found a priming asymmetry. This, they argued, was evidence in favor of the psychological reality of lexical scales, as predicted by accounts such as that proposed by Horn (1972). However, what this research did not address was whether this activation is indicative of anything beyond lexical relations between scale mates, as their participants were presumably engaged in lexical processing only and were not deriving any scalar implicatures.

Let us now turn to discussing the recent study by Ronai and Xiang (2023), which attempted to fill this gap. These researchers were interested in the “psycholinguistic reflexes” of pragmatic inferential processes. More concretely, they set out to test whether in the context of SI-triggering sentences, the stronger alternatives are retrieved from memory and activated. They did this for a variety of scales, which they took from their previous research on the inference rates of these different scales (Ronai & Xiang, 2022). These included scalar words of many different categories, among which were quantifiers such as the pair some and all, the conjunctions or and and, adverbs such as partially and completely, verbs (for example, slowed and stopped), and adjectives (for example warm and hot), which made up more than half of all items used. They included 60 scalar pairs altogether.

They conducted several web-based experiments in which their participants had the task of performing lexical decision on the strong term of each item pair. The prime words were either the weak scale-mate of the given scale (e.g., warm with hot as the target) or an unrelated word (e.g., mercurial). Their Experiment 3 was the crucial one for testing whether the stronger scalar terms are activated in real-time comprehension. Ronai and Xiang (2023) embedded their weak scalars and unrelated words within sentential contexts that could give rise to a scalar implicature (The weather is warm/mercurial). The sentence was presented in the rapid-serial visual presentation mode and was followed by 650ms of a blank screen, at which point the target word appeared.
What the results of this experiment showed was that strong terms (hot) were indeed activated by the presence of their weaker scale-mates (warm), evidenced by the fact that comprehenders reacted to the strong scalar target quicker when the weak scalar was the prime compared to when it was the unrelated word. In order to ascertain that this was in fact the work of pragmatic inferential processes and not just a result of lexical-based priming, they conducted their Experiment 2, in which the prime words were presented in isolation without any sentential context. Here, no priming was observed. What Ronai and Xiang (2023) take these results to mean is that weak scalar terms activate their stronger scale-mates only in SI-triggering contexts, while this activation cannot be assigned to simple lexical factors unrelated to pragmatics, and thus, that comprehenders in fact entertain the stronger terms during online scalar implicature derivation processes².

What these two studies taken together show is that strong scalar terms are activated in the minds of comprehenders and that this effect is specific to sentential contexts, which could in principle give rise to scalar implicatures, and that they are primed asymmetrically when compared to their weaker counterparts, or in other words that priming seems to be stronger when going up the scale in terms of informational strength as opposed to downwards. This evidence points towards the psychological reality of scalar alternatives and to them being employed in a similar way to focus ones.

There is another strand of research that operates with the activation of alternatives within their both experimental and explanatory framework. This is the examination of scalar implicature derivation by means of structural priming methods, ultimately derived from language production studies (Branigan & Pickering, 2017). Here, visual stimuli or linguistic material (often both) are presented as primes with the target being a sentence (often accompanied by a visual scene), which could give rise to a scalar implicature (e.g., Bott & Chemla, 2016). Different tasks and measures such as the truth-value task or the covered box paradigm are used to assess whether the participant in question derived a scalar implicature in a given trial. In the study of Rees and Bott (2018), participants were found to be more likely to derive an implicature (e.g., some of the shapes are circles) when the prime sentence preceding this trial contained the stronger alternative (e.g., all of the shapes are squares). Rees and Bott (2018) interpreted these results as supporting a particular model of implicature derivation, in which the activation of the stronger alternative is key. They proposed a theory in which scalar implicature derivation is initiated only when the alternative is made sufficiently salient, i.e., activated. These data and theoretical suggestions were further corroborated by Bott and Frisson (2022), who showed that when their participants

---

²Ronai and Xiang (2023) also tested whether the inclusion of the particle only had an effect on the activation of strong scalars, since the particle had been found to create additional interference in the case of focus (e.g., Gotzner et al., 2016). However, no such effect was found.
were shown the alternative in the same paradigm, the time it took them to derive the implicature in the subsequent trial decreased.

All in all, studies using both types of priming seem to coalesce around the idea that the activation of the stronger alternative within the minds of comprehenders is key to deriving scalar implicatures.

3.2. Evidence for the role of non-entailed alternatives
While the research reviewed above focused on the question whether informationally stronger terms are active during online scalar implicature derivation, an element that most of the theoretical accounts consider to be necessary, there has also been recent research asking whether other, informationally weaker alternatives also appear within the process or even participate in it.

Working within the Rational Speech Act framework, Peloquin and Frank (2016) attempted to model participants’ judgments regarding scalar implicatures on different scales. They tested several models that included either only the stronger alternatives, as predicted by the classical treatments discussed above (e.g., Horn, 1972), or other alternatives, such as antonyms and middling terms (e.g., okay for the scale <good, excellent>). They found that contrary to the predictions, the models that included these other alternatives provided a better fit for participants’ judgments. Peloquin and Frank (2016) took this as evidence that other informationally not stronger alternatives might also play a role in people’s reasoning about pragmatic meanings and deriving scalar implicatures.

Additionally, Degen and Tanenhaus (2016) showed that in scalar implicature processing with quantifiers (i.e., some), including exact numbers (one, two) as alternatives that could be entertained by the comprehender, people judged the quantifier some as less appropriate within the given discourse, thus giving these researchers evidence that alternatives that are not stronger and are even from a different category could be playing a role in the process.

Another empirical suggestion for the relevance of non-entailed alternatives comes from the domain of the acquisition of pragmatics. In the research of Skordos and Papafragou (2016), which examined the derivation of scalar implicatures by preschool children and the role of alternatives therein, it was found that when their children participants were explicitly provided the non-entailed alternative none, they were more likely to interpret the quantifier some in a pragmatic way. In fact, the strength of this priming was similar to when the strong term all was cued. Such findings are surprising given standard accounts which assume that negative and positive quantifiers are represented on separate Horn scales.

To directly test whether non-entailed alternatives could play a role in scalar implicature derivation, a recent study by Lacina, Ronai, Alexandropoulou and Gotzner (2023)
probed the activation of antonyms and of scale-mates under negation. They took the method and stimuli from the above-discussed research by Ronai and Xiang (2023) as their starting point. The main question being answered in their research was whether the priming observed by Ronai and Xiang (2023) was indeed indicative of scalar implicature derivation and whether even in situations where the target word, i.e., the strong scalar term hot in the previous study, would also be activated when it ceased to be informationally stronger due to manipulations to the sentence context and the prime word.

Negation is a downward entailing operator. When a sentence includes constituent negation of, say, a predicative adjective, i.e., not warm, no scalar implicature to the effect of warm but not hot is expected to arise (e.g., Horn, 1972). This is due to the fact that hot is no longer informationally stronger within the negated context. The meaning not hot is not implicated here, but entailed. In their first experiment, Lacina and colleagues (2023) used this to test for the activation of strong scalar target words (hot). What they found was that negating the weak scalar term canceled the priming effect for the scale-mate. In other words, when negated, warm no longer activated hot.

Next, they conducted two experiments where the weak scale-mate prime (warm) of Ronai and Xiang (2023) was replaced with its antonym (cool), while the target word was the same (hot). In one experiment, they embedded these antonyms into the same sentential frames that were used in the study of Ronai and Xiang (2023) in their Experiment 3. In the other experiment, they presented the same antonymic primes as isolated lexical items, mirroring the lexical Experiment 2 from Ronai and Xiang (2023). What they found out was that the antonyms (cool) activated the strong terms of opposite polarity (hot) both when inside a sentential context and when presented as isolated words.

Finally, the researchers also conducted an experiment where antonymic primes were used but constituent negation was also added. Their participants were therefore exposed to sentences containing not cool and were to react to the target hot. Here, no priming was observed, similar to the negated weak scale-mate experiment. In order to test whether the lack of activation observed in the two negated experiments was only due to the presence of negation rather than due to any implicature or informational strength factors, Lacina et al. (2023) conducted a joint analysis of their experiments with sentential context and of Experiment 3 from Ronai and Xiang (2023). They found that negation had a different impact on the lexical decision times to the targets depending on whether the prime was a weak scale mate or its antonym. Negation interacted with the weak scalar primes, but not with the antonymic ones. This in turn suggested that while antonyms might activate the strong scalar targets, negation does not seem to be effecting this activation in the same way that it does when the target is informationally stronger and is a the locus of a possible scalar implicature. Consequently, we interpret these results to be in line with the Alternative Activation Account
discussed above, since these data suggest that associative links do activate a slew of alternatives, some informationally stronger, some weaker, yet that a selection process follows that, depending on the context and grammatical restrictions, proceeds to pick out only those alternatives that are directly relevant to implicature derivation.

What we have attempted to show and highlight in this section is that alternatives beyond the stronger, entailed scale-mates play a role in the process of scalar implicature derivation, which goes against the predictions of most standard accounts of scalar implicature. On the contrary, it appears that modeling data suggest that including non-entailed alternatives improves accuracy with regards to matching human data and that experimental studies show that non-entailed alternatives such as antonyms are active in the process of real-time scalar implicature derivation. The Alternative Activation Account presents us with an explanation—it might be that non-entailed alternatives are activated during the process via the semantic network. Those should be discarded later in order for the relevant alternatives to be worked upon in further stages to achieve the final pragmatic meaning.

Relating these findings back to the theories of scalar implicature outlined in Section 2.2, one should note that all accounts would exclude antonyms from the set of relevant alternatives. At this point, we cannot distinguish whether antonyms are activated as part of the broader activation spreading mechanism (step 1 in the Alternative Activation Account) or whether they are part of the relevant set of alternatives. This would mean that antonyms are being explicitly negated during the inferential process or that there is another scalar implicature mechanism that is not sensitive to lexical Horn scales. In the following, we review evidence that casts doubt on the role of lexical alternatives in scalar implicature and we briefly sketch some novel scalar implicature mechanisms.

4. Different alternatives or different implicature mechanisms?
While the theoretical and experimental literature reviewed so far indicated that alternatives play a crucial role in pragmatic reasoning, there are also studies that cast doubt on this assumption. Especially when considering scales beyond the quantificational ones involving some and all, alternatives may be invoked differently or not at all. In the following section, we first review evidence suggesting that alternatives are invoked differently across triggering expressions and then we sketch recent proposals that could explain these results via a scalar implicature mechanism that does not involve lexical alternatives.

4.1. Scalar diversity and the role of semantics in scalar implicature computation
The first study showing variation across triggering expressions was carried out by Doran, Baker, McNabb, Larson and Ward (2009). They manipulated whether a contextual question
evoked alternatives on a scale and how many scale-mates were mentioned. This manipulation was only effective for adjectival scales. For example, when the stronger scale-mate *gorgeous* was mentioned in the contextual question, participants were most likely to endorse the scalar implicature *Kate is pretty but not gorgeous*. In fact, the highest rates were observed in the condition where both the stronger scale-mate *gorgeous* and a non-entailed alternatives *average-looking* was mentioned. For quantifiers like *some* the endorsement of the scalar implicature was equally high in all contextual conditions. This could mean that quantifiers like *some* more automatically activate their stronger scale-mate than adjectives. The adjective results, in turn, might suggest that non-entailed alternatives play a role in the inferential process, as argued in Section 3.2.

Building on Doran et al.’s study, van Tiel, Miltenburg, Zevakhina and Geurts (2016) constructed a set of Horn scales involving different grammatical classes and demonstrated large variability in scalar implicature endorsement rate across triggering expressions - an effect that became known as *scalar diversity*. To measure the extent to which activated alternatives predict inference rates, van Tiel et al. used several predictors such as the likelihood of mentioning the strong term in a cloze task and relative frequencies in the COCA corpus. Under the hypothesis that inference rates vary as a function of whether alternatives are available to the listener, one would expect that cloze probabilities predict inference rates. However, no such effect was not observed. What did predict inference rates was the boundedness of the strong term and the distance between scalar expressions. That is, when the strong term denoted an endpoint (e.g., *certain* representing 100% percent on a degree scale), inference rates were relatively higher. Semantic distance was measured via a rating of the relative strength of corresponding expressions and the stronger a scale-mate was judged, the more likely participants were to endorse a scalar implicature. While these factors did explain some of the variability, a large amount still remained unexplained.

In a study using the same items, Ronai and Xiang (2022) tested whether the variability in scalar implicature disappears entirely when *only* is used. Since *only* requires alternatives due to its semantics, one would expect that with any scalar trigger, participants endorse the negation of its stronger alternative. For example, *only some* should mean *not all* on its semantic meaning. Contrary to expectation, *only* did not fully eliminate the scalar diversity effect. Providing a contextual question with the strong scale-mate also left some of the variability unexplained. Interestingly, a combination of both manipulations led to almost 100% endorsement rates across all scalar triggers. This indicates that for certain scalars, listeners do not automatically generate the strong alternative and that simply providing it in the context, does not necessarily lead listeners to negate the strong alternatives. In line with this, modeling work by Hu et al. (2022) finds that a corpus measure quantifying the uncertainty over evoked alternatives predicts the scalar diversity effect. That is, the more
listeners are uncertain about which alternatives are crucial, the less they endorse a pragmatic inference.

In all of the aforementioned studies, the scalar implicature endorsement rates were relatively low for adjectival scales. However, it is implausible that this is due to the nature of this grammatical category, as a study by Gotzner, Solt and Benz (2018) indicates. The authors constructed a set of 70 adjective scales that were balanced with respect to the scale structure underlying the semantics of adjectives (Kennedy and McNally, 2005). The authors reasoned that quantifiers may not differ from adjectives per se as a grammatical class but instead that certain adjectives have a similar semantics as the quantificational scales (e.g., *<possible, certain>* ) and others involve a completely different semantics. Gotzner et al. hypothesized that these semantic factors should play a key role in scalar implicatures and classified their test set according to the types of standards their scale structure invokes, their polarity and extremeness. The factors relating to the semantics of adjectives explained around 67 percent of the variability in the endorsement of scalar implicature. While van Tiel et al. (2016) only looked at the boundedness of the stronger scale-mate, the study by Gotzner et al. (2018) took into account additional semantic factors concerning both scale-mates (weak and strong) and used linguistic tests from the formal literature on adjective semantics to classify expressions. The study by Ronai & Xiang (2022) showed that not all expressions automatically activate their stronger scale-mates. It remains to be determined whether this is a factor that operates independently of the role of adjective semantics.

A study by Alexandropoulou, Herb, Discher and Gotzner (2022) investigated the role of adjective semantics to the incremental processing of scalar implicature. The study compared relative adjectives (e.g., *warm*) that involve context-dependent standards to minimum standard adjectives that impose a lower bound by their semantics (e.g., *breezy*) in a visual world setup (building on Sedivy et al., 1999). Participants clicked on a target picture over a competitor representing the stronger scale-mate were taken to indicate a scalar implicature. As hypothesized, the authors found an interaction of the adjective semantics with the context-dependence of the scalar implicature. Specifically, the presence of a contextual contrast (e.g. a picture of cold water) facilitated the derivation of scalar implicature involving relative adjectives, presumably by setting the threshold of the adjective meaning. In contrast, the scalar implicature was computed independently of the contextual contrast manipulation for minimum standard adjectives. Overall, the study indicated that semantic thresholds and pragmatic upper bounds are computed in parallel.

To summarize, there is evidence that alternatives are evoked to a different extent in the minds of listeners when considering different scalar expressions. What is more, the semantics especially of adjectival scales systematically affects the computation of scalar
implicate. What we can conclude at this point is that the traditional mechanisms for scalar implicature are either incomplete to account for the processing findings or they only apply to a subset of scalar expressions such as quantifiers.

4.2. Semantic and conceptual representations

Most work reviewed so far tries to rescue the assumption that alternatives are evoked in scalar reasoning to some extent, for example by postulating that alternatives are activated to a different extent across triggering expressions. But there is also a way to make sense of the above-mentioned results without appealing to alternatives.

Given the findings that measurement scales play a role in pragmatic reasoning, Gotzner (2022) proposes a different mechanism for scalar implicatures involving adjectival scales without reference to alternatives. Instead the so-called measurement mechanism postulates that listeners reason about positions on an underlying measurement scale. A term like warm would exclude degrees that are above the degree interval denoted by warm. Thus, listeners may not be reasoning about stronger expressions that could have been used but rather about where to order a given expression on a degree scale. Since the meaning of certain adjectives involves relative context-dependent thresholds, the upper bounded meaning (corresponding to a scalar implicature) will also vary as a function of context. This would account for the boundedness and scalar distance findings (van Tiel et al., 2016; Gotzner et al., 2018) and the role of contextual contrast in incremental processing (Alexandropoulou et al., 2022) without appealing to activated alternatives. For example, the effect of a contextual contrast could be akin to the role of comparison classes in determining the threshold of a relative adjective. If a contrast is contextually salient, the bounds are placed lower such that the ranges denoted by different expressions are more distinguishable. Considering the findings mentioned in Section 3.2., which indicated a role of non-entailed alternatives, the measurement mechanism would state that antonyms play a role in the inferential process but not their lexical representation per se. Instead it is the degree scale onto which different expressions may be mapped.

There are several other proposals in the literature which have a resemblance with the measurement mechanism. For example, Magri (2017) proposed to model Hirschberg scales (e.g., Bachelor’s degree, Master’s degree) via the negation of elements that are ranked higher on a partial order. Based on modeling results, Hu et al. (2023) conjecture that listeners may not have the strong scale-mate in mind but rather a concept of a more extreme case. To summarize, different proposals have been made that do not involve lexical alternatives but some other semantic or conceptual basis for scalar-implicature reasoning (see also Buccola et al., 2022 for a view involving conceptual alternatives for quantificational scales).
At this point, none of the existing work can pinpoint exactly the nature of the representation of alternatives. While different inference computation mechanisms may be involved for different scalars, another possibility is that there is only one mechanism but different kinds of alternatives. To make progress on these debates, we outline the key outstanding questions and how they can be addressed experimentally in the next section.

5. Future research on key test cases
We began this paper by showing that alternatives are represented in the mind of listeners in real-time processing and that initially expressions beyond scale-mates are invoked. We accounted for these findings via the Alternative Activation Account, which assumes two steps for generating and selecting alternatives (Gotzner, 2017). On this view, a broad cohort of semantically-related expressions is activated and subsequently pragmatic and grammatical mechanisms narrow down this broad set to the relevant alternatives. Several existing studies have shown that non-entailed alternatives become activated during processing but it is unclear which precise role they play during the inferential process.

According to the Alternative Activation Account, only relevant alternatives should be detected at later processing stages. Thus, it is crucial to test the evolving set of alternatives over the course of real-time processing. For this purpose, the paradigms of focus can be borrowed employing stimulus onset asynchrony (Husband & Ferreira, 2015) and memory (Gotzner, 2019) manipulations for activated alternatives and corresponding inference computation.

The second main point of this article was that theoretical accounts make different assumptions about the nature of alternatives: they could be entire sentences, phrases, lemmas or concepts. While the assumption that listeners reason about entire alternative utterances is implausible given current findings, none of the existing work can tease apart the other existing proposals since only nouns were tested in lexical tasks. In case alternatives are phrases, how can they be constructed online? Future processing studies need to test whether listeners activate alternatives beyond lexical expressions and to what extent grammatical restrictions guide the selection of alternatives. A recent study, Lacina, Sturt and Gotzner (2023) showed that probe recognition can be used for phrasal alternatives in the case of broad focus. Such paradigms can be used to tackle grammatical constraints and phrases of different levels of complexity, for example to tackle the symmetry problem (see Breheny et al., 2017 for a recent overview).

To test new accounts that assume lemma-level or conceptual alternatives, visual paradigms can be used. The existing studies in this domain (e.g., Tieu et al., 2019; Chung, Bade, Blanc-Cuenca, Mascarenas, 2022) do seem to indicate that scalar implicature reasoning can be done without linguistic stimuli and the involvement of language. However,
it cannot be excluded that participants also activate some lexical representations. Hence, distinguishing conceptual from lemma-level or some other form of lexical representation will require researchers to design new paradigms that do for example inhibit aspects of the representation.

Furthermore, two recent proposals make reference to measurement scales (Gotzner, 2022) and ranked orders (Magri, 2017) and discard the role of alternatives for certain scales. To test these accounts, cross-linguistic studies on languages that do not involve degrees or that involve different alternative sets (e.g., Stateva et al., 2019; Dionne and Coppock, 2022) could shed light. The other main prediction of the measurement mechanism is that scalar implicature should be possible when even access to lexical alternatives is interrupted. For example, production interference paradigms or transcranial magnetic stimulation (TMS) used to inhibit areas of the brain responsible for lexical access could be used to test this assumption. Finally, testing scalar implicature for novel expressions in an artificial learning paradigm (e.g., Buccola et al., 2022) or in developing children are promising avenues to tease apart lexical alternatives from other degree-based representations.

The key piece missing in all existing work is evidence that alternatives play a causal role in reasoning. The structural priming paradigms (Bott and Frisson, 2022) are currently the closest to providing an answer to this question. However, a more direct online measure is needed to study the evolving set of alternatives and the inference making process in tandem.

6. Concluding remarks

In this paper, we have reviewed recent work on the nature of alternatives underlying scalar implicature computation. Contrary to existing theoretical proposals, studies have shown that listeners entertain alternatives beyond scale-mates during processing. These findings are expected based on domain-general mechanisms that subserve semantic networks in the brain. The literature indicates that sentence processing is incremental and so is the generation of alternatives. For this reason, the key questions about the nature of alternatives and the mechanisms underlying scalar implicature computation need to be addressed in real-time processing and probed at different time points.

We have highlighted that different scalar expressions may invoke alternatives to a different extent. It is an open question whether this means that different mechanisms underlie the computation of scalar implicature for different scalar triggers. Therefore, future work needs to use paradigms that distinguish activated alternatives from those alternatives that an inference is made on.
References


Gotzner, N. (2022). All you ever wanted to know about scales but were afraid to ask. Invited talk, 22.9.- 23.9.2022. Biennial XPRAG, Pavia.


Rooth, M. E. (1985). *Association with focus (montague grammar, semantics, only, even)*. University of Massachusetts Amherst.


