Only the (informationally) stronger survive: A probe recognition study with scale-mates and antonyms

Radim Lacina & Nicole Gotzner*

Abstract. Speakers often use scalar words such as *warm* in a pragmatically strengthened way that results in the conveyed meaning being *warm but not hot*. In these inferences, known as scalar implicatures, meaning alternatives have been postulated to play a crucial role. Upon encountering *warm*, the informationally stronger alternative *hot* has been shown to be active in online sentence processing. Antonyms (*cool*) have also been shown to be activated in the same way even though they are standardly assumed to not be involved in scalar implicature derivation. In the current study, we focus on the question of whether both strong scalar alternatives and antonyms are represented in the final mental model of the discourse following scalar implicature derivation. We ran two probe recognition experiments, testing strong scalars and antonyms. We found an interference effect for strong scalars, indicating their representation, but not one for antonyms. Thus, we provide evidence that only the strong scalars survive in the eventual representation of the pragmatic meaning of a sentence.

Keywords. scalar implicatures; alternatives; probe recognition; antonyms; informational strength

1. Introduction. In human language, there are many words whose meaning is tied to a particular underlying scale with their contribution being that they describe a certain degree of some property or other (Kennedy 2007). Often, there are families of related words that differ in the degree of the property which they ascribe. Through their semantic relations, they then also give rise to pragmatic inferences communicated by speakers and derived by listeners. Take the following example:

(1) When I put my foot in, the bath water was warm.

When uttering (1), the speaker most likely wishes to convey two meanings at once. Firstly, they say that the temperature of the water crossed a certain contextually determined threshold, in other words, that it was at least warm. This would correspond to the literal meaning contributed by the adjective *warm* (Kennedy & McNally 2005). Secondly, there is the added pragmatic inference that while the water was warm, its temperature was not so high as to be considered hot. These inferences have become known as scalar implicatures (Horn 1972). We can tell that this other

*This research was supported by the German Research Foundation (DFG) through its Emmy-Noether programme, a grant awarded to the second author (Nr. GO 3378/1-1). We would like to thank the audience of ELM3 for their valuable comments on this research and Stavroula Alexandropoulou for collaborating with us on an inter-experimental participant-sharing scheme employed in Experiment 1. Authors: Radim Lacina (radim.lacina@uni-osnabrueck.de), Institute of Cognitive Science, Osnabrück University; Department of Czech Language, Faculty of Arts, Masaryk University, & Nicole Gotzner (nicole.gotzner@uni-osnabrueck.de), Institute of Cognitive Science, Osnabrück University. Author contributions: Radim Lacina: Conceptualisation, Methodology, Resources, Software, Data curation, Investigation, Formal analysis, Visualisation, Writing – original draft, Writing – review and editing; Nicole Gotzner: Conceptualisation, Methodology, Writing – review and editing, Supervision, Funding acquisition, Project administration.

meaning is in fact pragmatic, because this scalar implicature passes the test of cancellability as seen in the following example (Grice 1975):

(2) When I put my foot in, the bath water was warm. In fact, it was hot!

Notice that the conjunction of the two sentences comprising (2) is not a contradiction. What this means is that the negation of the proposition that the water was hot is not entailed by (1). One can clearly see the difference when we substitute *warm* for its antonym *cool*. As the reader can ascertain for themself, uttering (3) is at least pragmatically odd:

(3) #When I put my foot in, the bath water was cool. In fact, it was hot!

In the current study, we are concerned with the eventual representation of these scalar implicatures with special attention paid to the unmentioned stronger scalar alternative *hot*, which is derived in the case of (1), following an online derivation process. We are also interested in how this contrasts with cases such as (3) where the relationship between two scalar words is that of antonymy. The remainder of this article is structured as follows. We begin with an introduction to the theoretical treatments of scalar implicatures followed by a discussion of the current literature on the online derivation of scalar implicatures and the role of alternative meanings in this process. We then present two probe recognition experiments where we contrast stronger scale-mate alternatives with antonymic ones and show that only the former are retained in the representation of implicated meaning whereas the latter, the antonyms, are not.

1.1. THEORETICAL TREATMENTS OF SCALAR IMPLICATURES. Most accounts of scalar implicatures coming from the fields of formal semantics and pragmatics see alternative meanings as crucial to the derivation of the enriched meaning component (for an overview of various theories, see Sauerland 2012). Horn (1972) sees the key to implicature derivation in ordered lexical scales. According to this approach, words such as *warm* and *hot* or *some* and *all* are represented in the lexicon ordered by informational strength. For example, *hot* is said to be the informationally stronger scale-mate to *warm*, since they have an asymmetric entailment relationship. When we describe some object as being hot, the truth of that sentence necessitates that the same object is at least warm at the same time. However, when we reverse the expressions, something's being warm does not mean that it must also be hot.

Under this view, other scale-mates as well as other related words are irrelevant when it comes to scalar implicature derivation. For example, antonyms such as *cool* or *cold* do not play a role in this process and are even not considered to be on the same scale as those adjectives of opposite polarity (see Horn 1972; for the split-scale assumption). However, there is some recent evidence pointing in the direction that despite what this assumption might suggest, antonyms could be involved in the derivation of implicatures. Peloquin & Frank (2016) found that in their computational study of implicatures, including antonyms in their model improved its ability to predict human judgements of pragmatic meaning. Skordos & Papafragou (2016) found that the presence of the non-entailed alternative *none* boosted the derivation of *some* to *some but not all* implicatures in 5-year-old children. Baker et al. (2009) found that when non-entailed alternatives were included in a QUD (question-under-discussion) preceding a sentence with a weak implicature-allowing scalar,

this also led to an increase of the acceptance of the implicated meaning.

In the case of relative adjectives, of which *warm* and *hot* are prime examples (Kennedy & McNally 2005), Alexandropoulou & Gotzner (2022) found that the pragmatic inferences with these adjectives were supported by the presence of an antonymic contrast in the incremental decision visual world study they ran. Absolute adjectives (e.g., *breezy*) did not show the pattern and people were equally likely to arrive at the implicature meaning with or without this contrast. This suggests that, at least with some types of scales, the cueing the antonymic meaning supports implicature derivation. Peloquin & Frank (2016) report their computational modelling study of implicature derivation and argue that also including expressions other than just the strong scalar, such as antonyms, improved how well the their model fit the human judgement data.

All of this research suggests that alternatives are crucial in the derivation of scalar implicatures, but also that what alternatives exactly are involved and how is still unknown and a matter for current research, albeit with preliminary suggestive evidence that alternatives beyond the informationally stronger ones could also play a role. These questions might in turn be helped by empirical methods that seek to examine what alternatives are present during *language processing* and when. We review these findings below.

1.2. THE ACTIVATION OF SCALAR ALTERNATIVES. Recently, researchers have started examining the role of alternatives in the online processing of scalar implicatures. Similarly to earlier research in the related domain of focus, where it had been shown that alternative meanings are active in the process of comprehension (Braun & Tagliapietra 2010, Husband & Ferreira 2016, Gotzner et al. 2016), lexical decision experiments were run to test this. De Carvalho et al. (2016) ran a masked priming experiment focusing on the lexical association between weak and strong terms such as *some* and *all*. What they reported was a pattern of asymmetric priming. Weak scalar terms (*some*) activated their stronger scale mates (*all*) to a larger degree than the strong ones activated the weak. The researchers interpreted this as evidence for the psychological reality of lexically encoded Horn scales (Horn 1972). They, however, did not test scalar words in contexts where they could plausibly support implicatures in the minds of comprehenders, since they presented their scalar stimuli as isolated lexical items.

Ronai & Xiang (2023) moved this research forward into the domain of sentence processing and asked whether there is a difference in the activation of the strong term by its weak scale-mate when the latter is embedded in a sentence that could give rise to an implicature as opposed to when presented as an isolated lexical item. They used sentential frames such as the following:

(4) Zack's carpet was dirty/patterned.

In their sentential experiment, native speakers of American English saw sentences such as (4) either with a related weak scalar prime (*dirty*) or an unrelated one (*patterned*). They then completed a lexical decision task on the word *filthy*, which was the strong scale-mate to *dirty*. Ronai & Xiang (2023) found that it was only when the weak scalar primes were embedded within a sentential context that priming occurred. On the other hand, them presenting these words in isolation did not cause participants to react to the strong scalar targets faster.

The researchers concluded this to be evidence of their priming effect being indicative of pragmatic-specific activation and argued that strong scalar terms were being involved as mean-

ing alternatives in the process of online implicature derivation.

A study by Lacina et al. (2024) examined two questions regarding the activation of scalar alternatives in processing. They were interested in whether (a) the priming was specific to sentential contexts with a potentially derivable implicature and would disappear when the context would not allow for it and (b) whether alternatives other than just the informationally stronger term were active in the process. Firstly, they introduced constituent negation to sentences such as (4) with related primes being *not dirty*. Negation is said to reverse entailment relations, which in the case of (4) results in the presented target word *filthy* no longer being informationally stronger in this context. Consequently, the scalar implicature of *dirty, but not filthy* is no longer derived. What the researchers found was that in these cases, there was no longer any priming from *dirty* to *filthy*. Next, they examined whether antonyms were also activated during the process. They replaced the weak scale-mate primes (*dirty*) with their antonyms (*clean*) while keeping the same targets (*filthy*). They found that antonyms activated the targets, similarly to weak scalars.

What this research shows is that stronger scalar alternatives are preferentially activated during comprehension in cases where they can support a pragmatic inference and that antonymic meanings are activated in that process at the same time. In the following section, we present an account attempting to capture this pattern.

1.3. THE ROLE OF ALTERNATIVES IN COMPREHENSION. To explain the processing of alternatives in comprehension, the Alternative Activation Account has been posited (Husband & Ferreira 2016, Gotzner 2017). This is view that puts forward a two-step mechanism (see Gotzner & Lacina (in print) for an application to scalar implicatures). It suggests that lexical alternatives are first activated together with pragmatically irrelevant semantically associated items in early processing by means of domain-general mechanisms of activation spreading. In the following step, the proper alternatives (i.e., the ones relevant to implicature processing) are selected and maintained in further processing. This model is illustrated in a simplified manner in Figure 1 below, which takes the example of the scale used in (4) and gives the reader a general idea of what the Alternative Activation Account postulates. The example itself is of the process that begins when a comprehender encounters the weak scalar word *dirty* within an upward-entailing sentential context, where implicatures are expected to be derived, such as the one in (4).

The studies of Ronai & Xiang (2023) and Lacina et al. (2024) have focused on the early stages of processing. They have shown that both strong scalars and antonyms are activated at least during the first stage (Step 1 in Figure 1). What these studies have not addressed, however, is the state of the comprehenders' minds once the process of implicature derivation is complete and what is included in the eventual representation of the enriched meaning constructed. In the following sections, we aim to address this gap in knowledge. We first discuss how this question might be approached using the probe recognition task (Gernsbacher & Jescheniak 1995) and argue for its suitability as a tool to examine representations in the mental models of discourse that comprehenders form.

1.4. THE REPRESENTATION OF ALTERNATIVES. Above, we reviewed how the lexical decision task, having been first used to study focus alternatives, has then been implemented in the domain of scalar implicatures. In the current section, we introduce another method that was first used to study focus alternatives, but with the goal of studying eventual representations—the probe recognition

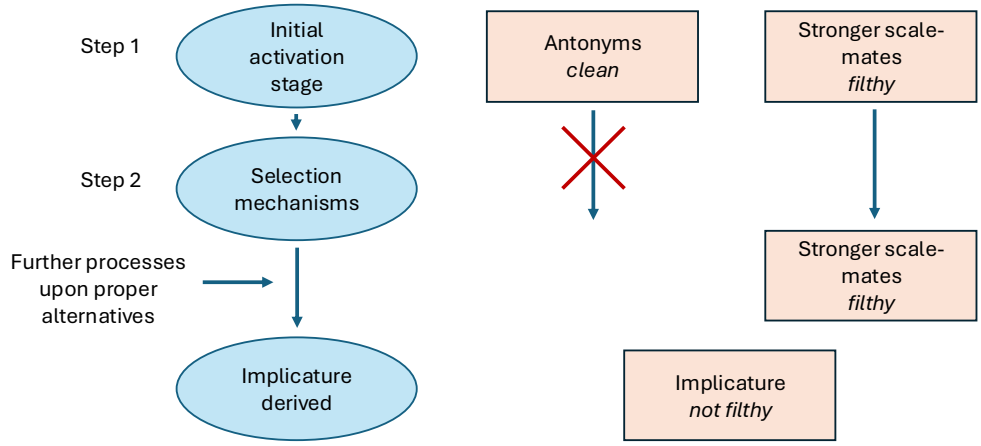

Figure 1: Diagram illustrating the proposed Alternative Activation Account in the case of encountering the weak scalar word *dirty* in an implicature allowing context. During Step 1, both antonyms and strong scalars are activated. What follows is a selectional phase during which only the stronger scale-mate (*filthy*) remains activated. Next, further processes derived the implicature that is equivalent to the negation of the strong term (*not filthy*).

task (Gernsbacher & Jescheniak 1995, Gotzner et al. 2016).

Studies on meaning alternatives in the case of focus have found that both contextually mentioned and unmentioned alternatives are represented in the mental model of the discourse (Gotzner et al. 2013, 2016, Jördens et al. 2020) and that these alternatives are more strongly encoded in memory (Fraundorf et al. 2010, Norberg & Fraundorf 2021, Spalek et al. 2014). The method employed in these studies is the probe recognition task, in which participants judge whether a given probe word appeared in a previously presented stimulus. This task has been argued to tap into the eventual strength of representation in the mental model of the discourse (Gernsbacher & Jescheniak 1995).

In this focus work by Gotzner et al. (2016), their probe recognition experiments revealed that focus alternatives were reacted to more slowly compared to unrelated words, and that this effect was increased by the presence of focus-sensitive particles such as *only* or *also*. Their reasoning was that this was due to a competition process between the alternatives and the focused element, making it more difficult to either confirm the presence of mentioned alternatives or reject that of unmentioned ones. Such competition could have only occurred in case that these alternatives were actually being represented as being parts of the discourse. Mere semantic associates, as opposed to plausible alternatives, did not show an interference effect in the probe recogition task (Gotzner

& Spalek 2017). Thus, the interference effect could be taken as evidence for the representation of a particular alternative.

Employing this method would, therefore, allow us to answer the question of what scalar alternatives are retained following their initial activation and subsequent pruning during implicature. For those alternatives that we predict to be necessary for this process, namely stronger scale-mates, we should expect an interference effect caused by their competition with the weak scale-mate that is the implicature trigger. Other related words, such as antonyms, however, are predicted not to give rise to this interference under both standard theoretical accounts (Horn 1972) and the Alternative Activation Account, since they ought to be eliminated following their pruning after an initial activation stage during processing. In other words, they should be absent from the final representation.

2. Experiments 1 & 2. In this research, we aim to answer the question of whether scalar alternatives are represented in the mental model of the discourse at a point when implicature derivation is complete. Furthermore, we are interested in what kinds of alternatives these are—whether only the informationally stronger alternatives, which are relevant in the derivation, are represented or whether it is also antonyms, which theory considers irrelevant, that are in the final mental model. Lacina et al. (2024) found that in earlier processing, antonyms are activated. We therefore ask whether this activation is maintained.

We ran two web-based probe recognition experiments, one with weak scale-mate primes and one with antonymic ones in order to answer these questions. Our predictions were based on the Alternative Activation Account (Gotzner & Lacina in print). Given that strong scalar terms are said to be relevant for implicature derivation (Horn 1972) and that their negation forms a conjunction with the literal meaning of the sentence giving rise to the implicature, we expect the strong term to be included in the eventual representation in the mental model of the discourse. As for antonyms, these do not form any core meaning of the derived implicature under standard theoretical treatments (Horn (1972) and following work). While there is evidence that antonyms are active and potentially relevant during the early stages of implicature derivation (see the discussion of Alexandropoulou & Gotzner 2022, Lacina et al. 2024, Peloquin & Frank 2016; above), they are predicted to not be in the final discourse representation.

As argued for above, we operationalise the strengthened representation of a particular element in the mental model of the discourse by its interference effect in the probe recognition task. We expect the strong scalar targets to be rejected slower when preceded by their weaker scale-mate as opposed to by an unrelated word. The same effect should not occur with antonymic primes.

3. Methods.

3.1. DATA AVAILABILITY. Our experimental stimuli as well as the statistical analysis script are freely available for downloading on the Open Science Framework platform. They may be reached using the following hyperlink: https://osf.io/khfgb/.

3.2. PARTICIPANTS. We recruited 79 participants for Experiment 1 and 80 for Experiment 2 (159 in total) on the Prolific platform. They were monolingual native speakers of American English, born in the US and American nationals, between the ages of 18 and 35, and participants with a 100 approval rate on Prolific. Experiment 1 was run together with another web-based eye-tracking

experiment not reported here. Participants were invited to take part in the eye-tracking experiment first. In case they failed certain web-camera calibration checks, they were redirected to Experiment 1, here discussed. Participants for Experiment 2 were recruited directly.

The demographics were the following. In Experiment 1, 45 women, 32 men and one person of diverse gender took part. One participant chose the option of preferring to not give their gender information. In Experiment 2, these numbers were 35 women, 42 men, two people of diverse gender, and one person did not disclose their gender. The mean age of participants in Experiment 1 was 27.9 (SD = 4.8) and in Experiment 2, it was 28.1 (SD = 4.3).

3.3. MATERIALS. We used the 60 items implemented in Experiment 3 of Ronai & Xiang (2023) and Experiment 2 of Lacina et al. (2024). Take the following example item:

(5) Zack's carpet was dirty. [related weak scale-mate, Experiment 1]

(6) Zack's carpet was clean. [related antonym, Experiment 2]

(7) Zack's carpet was patterned. [unrelated, Experiments 1 and 2]

The target word that participants reacted to was the same across experiments and conditions and was always the strong term relative to the weak scale-mate prime, in this case *filthy*. Each experiment contained one of the two related items, weak scale-mates in Experiment 1 and antonyms in Experiment 2. Both experiments shared the same unrelated items. This means that the correct response in the probe recognition task was always *no* in the experimental items, since the target word did not appear in the stimulus. Additionally, we created 60 fillers, where the correct answer was *yes*. These fillers were again taken from the study of Ronai & Xiang (2023). Our modification consisted in changing the probes from non-words to words appearing in the sentence part of the filler in order to fit the different requirement of the probe recognition task.

3.4. PROCEDURE. When participants arrived at the website with the experiments, either directly from Prolific in the case of Experiment 2 or via redirection from another experiment in the case of Experiment 1, they first read a consent form. After this, they received the instructions for the experiment. Their task was to read sentences in the rapid serial visual presentation mode (RSVP) followed by probe words. They were asked to indicate whether a given probe word appeared anywhere in the preceding sentence. They were instructed to use *j* for *yes* and *f* for *no*. Following practice items, the experiment itself commenced. Sentences presented in the RSVP mode (Potter 2018) were displayed at the rate of 350ms per word. After the last word of each sentence, 2000ms of a blank screen elapsed. After this, the probe word associated with the stimulus sentence appeared. Overall, the participants rated 60 experimental items and 60 filler items. The ratio of the correct *yes* and *no* responses was 1:1. Items were distributed based on the Latin Square design.

3.5. ANALYTICAL STEPS. We first excluded all trials of those participants whose accuracy on the combined set of experimental and filler items was below 90%. In Experiment 1, this amounted to four participants. In Experiment 2, we excluded two according to the same criterion. We then took all the trials of the remaining participants and filtered out all the trials with incorrect responses. The response times of these trials were then log-transformed for the analysis. We ran three linear mixed effects models. In all models, we included the maximal random effects structure allowed appropriate for the data that converged (Barr et al. 2013). There was one model

per experiment, where we included the single fixed effect of RELATEDNESS (related or unrelated) that was treatment coded with the related condition being 1 and the unrelated 0. The final model was the joint analysis one, where we first created a combined dataset and added the factor of EXPERIMENT with sum coded conditions of weak scalar (as 1, Experiment 1) and antonym (as -1, Experiment 2). In this analysis, the condition of RELATEDNESS was sum coded too with related being 1 and unrelated -1. We included the fixed effects of the two factors and their interaction. Here, in the random effects structure, we did not attempt a model with random slopes for the effect of EXPERIMENT or its interaction with RELATEDNESS for participants as this was a between-subject factor.

4. Results. The reader may consult the combined graphical representation of the response time results of Experiments 1 and 2 in Figure 2. Below, we first give the results of the individual models for each experiment followed by the joint analysis.

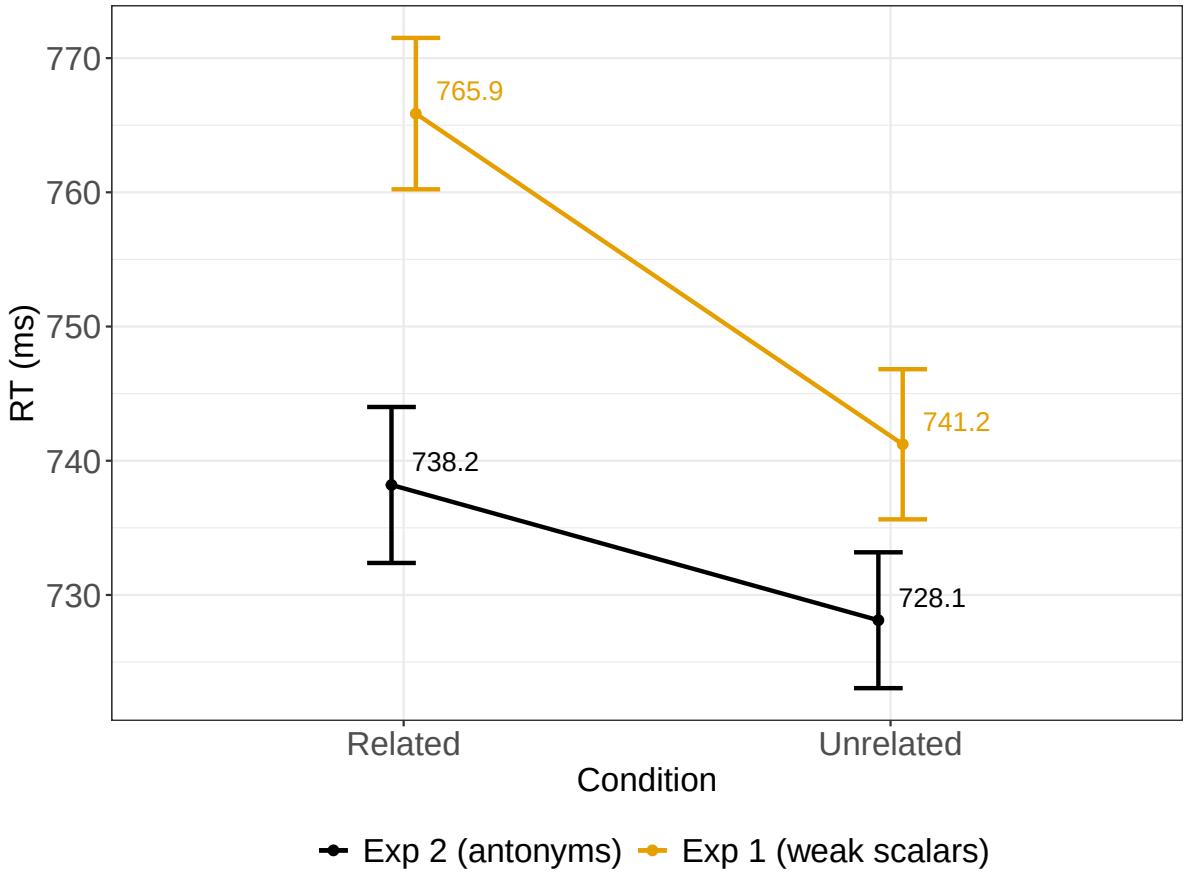

Figure 2: Mean reaction times in ms by condition with associated standard errors in Experiments 1 and 2

4.1. EXPERIMENT 1: WEAK SCALARS. The final model that converged had a random structure containing random intercepts for participants and items as well as random slopes for the effect

of RELATEDNESS for both. The model revealed that a significant effect of RELATEDNESS was present ($\beta = 0.0337$, $SE = 0.0099$, $df = 52.11$, $t = 3.416$, $p = 0.00124$**). This effect was in the positive direction, meaning that related weak scalar primes were associated with longer response times to the targets compared to when these targets were preceded by sentences with unrelated critical words.

4.2. EXPERIMENT 2: ANTONYMS. The model run on the data from Experiment 2 included random intercepts for participants and items as well as random slopes for the effect of RELATEDNESS for participants. As for the single effect of RELATEDNESS, this proved insignificant ($\beta = 0.0087$, $SE = 0.0076$, $df = 75.78$, $t = 1.144$, $p = 0.256$). Related antonymic primes, therefore, were not found to cause any differences in the rejection of the targets compared to unrelated ones.

4.3. JOINT ANALYSIS. The random effects structure of the joint analysis model that ended up converging without errors included random intercepts for participants and items as well as random slopes for the main effect of RELATEDNESS for both. The main effect of EXPERIMENT was insignificant ($\beta = 0.0126$, $SE = 0.0170$, $df = 151$, $t = 0.738$, $p = 0.46155$). There was a significant main effect of RELATEDNESS: $\beta = 0.0106$, $SE = 0.0031$, $df = 58.7$, $t = 3.418$, $p = 0.00115$**. Crucially, the interaction between RELATEDNESS and EXPERIMENT was significant: $\beta = 0.0061$, $SE = 0.0027$, $df = 146.3$, $t = 2.260$, $p = 0.02528$*.

5. General discussion. We investigated the representation of alternatives in the mental model of the discourse following the comprehension of scalar implicature-allowing sentences. We asked whether both strong scalars and antonyms were represented therein and ran two probe recognition experiments in this pursuit. Based on the Alternative Activation Account (Gotzner 2017, Gotzner & Lacina in print), we predicted that strong scalars would be present and antonyms would be absent.

Previous studies found that stronger scale-mates, which have been proposed to be necessary for implicature computation in the theoretical literature (Horn 1972, Sauerland 2012), are in fact activated in early stages of processing following exposure to a scalar word within an implicature-allowing sentence (Ronai & Xiang 2023). Further, it was found that even antonyms, which are standardly assumed not to be involved in the implicature derivation process, are activated at the same time (Lacina et al. 2024).

The results of our current probe recognition study showed that strong scalar items (*filthy*) were more strongly represented in the model when their weaker scale-mates were present in the preceding stimulus as opposed to when the prime was an unrelated word (Experiment 1). This was evidenced by an interference effect reflected in increased response times. Antonymic primes (*clean*), on the other hand, did not lead to an increased strength of representation for the same targets (Experiment 2). There, we saw no difference with the semantically unrelated condition. Our joint analysis then showed that this difference was present when the two datasets were directly compared—the type of related prime, i.e., weak scalar or antonym, influenced the size of the interference effect with weak-scalar primes associated with higher response times.

What this pattern shows is, we argue, that the informational strength relations between what is encountered in the course of comprehension and potential alternatives has an impact on what ends up being present in the representation of the combined literal and pragmatic meaning in real-time

comprehension. Strong scalar terms such as *hot* or *filthy* are relevant for the representation of the enriched meaning gained by implicature derivation, whereas antonyms (*cool*, *clean*) are not.

This interpretation is based on the ideas presented in Gotzner et al. (2016), which examined the influence of the particle *only* on focus alternatives and the lack of this effect with mere associates reported in Gotzner & Spalek (2017). The suggestion is that those elements that are alternatives show an interference effect, since they are in competition with the element present in the sentence. These are then harder to distinguish from each other. The reasoning is the same in the scalar implicature case. Given that we assume that the final representation of sentences with scalar words such as *My soup was warm* includes the negated strong term *hot*, which could be said to be roughly equivalent to the representation of *My soup was warm but not hot*, we expect comprehenders to experience difficulty in indicating that *hot* was not in fact present in the sentence they had just read.

The present results relate to those reported in the study of Lacina et al. (2024), who found that a the point of 650ms after encountering a scalar words, antonyms are still activated. Their data could not disentangle two competing options—antonyms being selected for and used in implicature derivation or only activated due to the first step of domain-general activation spreading in the lexical-semantic network. What the current data add here is evidence in favour of the latter, namely that antonyms are not directly involved in the process of implicature derivation and are only the remnants of the first activation step. Thus, our data are in line with the predictions of the Alternative Activation Account (see Figure 1 for a schematic illustration). What our study targeted was the stage at which the implicature derivation processed is finished and antonyms have already been eliminated during Stage 2.

One potential criticism of our conclusions might be that what our results reflect is simply the difference in similarity between the weak scalars and the targets and the antonyms and the same targets. It could be that antonyms are, on average, less related to the targets compared to the weak scalar primes, causing a difference to appear that is not related to any pragmatic processes, but a much lower-level contrast between the primes.

In order to deal with this issue, we conducted an additional analysis where we included similarity scores to test whether our results could be put down to this issue. Firstly, we took the similarity scores between the related (weak scalar or antonym) primes and the targets as well as the scores between the same targets and the unrelated primes as reported in Lacina et al. (2024). We ran a post-hoc analysis where we included the similarity scores between the prime and target as a fixed effect reported below. In order to explore the influence of semantic similarity on our results and whether there was any difference in its effect on the priming caused by antonymic and weak scalar primes, we ran a nested effects model on the related conditions subset of the data. We nested the effect of SIMILARITY within the experiment factor (i.e., within weak scalars or antonyms). The model included random intercepts for participants and items and the random slope for SIMILARITY in the case of participants. There was no main effect of EXPERIMENT (i.e., type of related prime): $\beta = -0.0216$, $SE = 0.0280$, $df = 141.9311$, $t = -0.771$, $p = 0.4422$. As for the effect of SIMILARITY within weak scalar and antonym primes, the model revealed that it only had an effect within Experiment 1 where weak scalars were the primes ($\beta = 0.1380$, $SE = 0.0627$, $df = 66.3464$, $t = 2.203$, $p = 0.0311$*). This effect was insignificant within Experiment 2 ($\beta = 0.0042$, $SE = 0.0536$, $df = 65.9823$, $t = 0.078$, $p = 0.9382$).

What this suggests is that the gradient effect of semantic similarity was only in place when the relationship between the prime and target was that of the weaker and stronger scale-mate. This did not occur when the relationship in question was that of antonymy. We might interpret this result as suggesting that in the case of weak-scalar primes, the strong scalar targets are in fact being represented in the mental model of the discourse and therefore, comprehenders experience difficulty rejecting them in the probe recognition task. The effect of similarity can then be understood conditionally: when the element is represented, the degree of difficulty is then modulated by semantic closeness with those more associated being harder to reject. In the case of antonymic primes, however, the targets are presumably not represented in the model, since they are not required for any pragmatic derivations. Here, similarity has no effect, since there is no competition due to representation that could give rise to a slow-down effect. This is also in line with the main result of Experiment 2, which found that antonymic primes did not differ from unrelated ones in their influence on the rejection of target words.

6. Conclusion. This study examined the strength of representation of scalar terms in the mental model of the discourse when these were preceded either by their weak scale-mates or by their antonyms. We found an interference effect, indicative of increased representation strength, only for the weak scalar primes (Experiment 1) and not antonymic ones (Experiment 2). The interpretation here pursued is that this is due to the fact that only strong scalar terms maintained in the final product of implicature derivation, as per most standard theories (Sauerland 2012), yet antonyms are irrelevant and are therefore not preferentially represented. This is in line with the Alternative Activation Account (Gotzner 2017, Gotzner & Lacina in print) which proposes an initial activation phase in which both strong scalars and antonyms are present with a subsequent narrowing to only the relevant alternatives.

References

Alexandropoulou, Stavroula & Nicole Gotzner. 2022. Negation, polarity, scale structure: Different inferences of absolute adjectives. In *Proceedings of Sinn und Bedeutung*, vol. 26, 35–54.

Baker, Rachel, Ryan Doran, Yaron McNabb, Meredith Larson & Gregory Ward. 2009. On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics* 1(2). 211–248.

Barr, Dale J, Roger Levy, Christoph Scheepers & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Braun, Bettina & Lara Tagliapietra. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25(7-9). 1024–1043.

De Carvalho, Alex, Anne C Reboul, Jean-Baptiste Van der Henst, Anne Cheylus & Tatjana Nazir. 2016. Scalar implicatures: The psychological reality of scales. *Frontiers in psychology* 7. 1500.

Fraundorf, Scott H, Duane G Watson & Aaron S Benjamin. 2010. Recognition memory reveals just how contrastive contrastive accenting really is. *Journal of Memory and Language* 63(3). 367–386.

Gernsbacher, Morton Ann & Jörg D Jescheniak. 1995. Cataphoric devices in spoken discourse. *Cognitive psychology* 29(1). 24–58.

Gotzner, Nicole. 2017. *Alternative sets in language processing: How focus alternatives are represented in the mind*. Springer.

Gotzner, Nicole & Radim Lacina. in print. Generating and selecting alternatives for scalar implicature computation: The Alternative Activation Account and other theories. In *Alternatives in Grammar and Cognition*, Palgrave Macmillan.

Gotzner, Nicole & Katharina Spalek. 2017. Role of contrastive and noncontrastive associates in the interpretation of focus particles. *Discourse Processes* 54(8). 638–654.

Gotzner, Nicole, Katharina Spalek & Isabell Wartenburger. 2013. How pitch accents and focus particles affect the recognition of contextual alternatives. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35, .

Gotzner, Nicole, Isabell Wartenburger & Katharina Spalek. 2016. The impact of focus particles on the recognition and rejection of contrastive alternatives. *Language and Cognition* 8(1). 59–95.

Grice, Herbert P. 1975. Logic and conversation. In *Speech acts*, 41–58. Brill.

Horn, Laurence Robert. 1972. *On the semantic properties of logical operators in English*. University of California, Los Angeles.

Husband, E Matthew & Fernanda Ferreira. 2016. The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience* 31(2). 217–235.

Jördens, Kim, Nicole Gotzner & Katharina Spalek. 2020. The role of non-categorical relations in establishing focus alternative sets. *Language and Cognition* 12(4). 729–754.

Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1). 1–45.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 345–381.

Lacina, Radim, Stavroula Alexandropoulou, Eszter Ronai & Nicole Gotzner. 2024. Scalar alternative activation in implicature processing: A lexical decision study with antonyms and negation. *PsyArXiv* 10.31234/osf.io/r3q79.

Norberg, Kole A & Scott H Fraundorf. 2021. Memory benefits from contrastive focus truly require focus: evidence from clefts and connectives. *Language, Cognition and Neuroscience* 36(8). 1010–1037.

Peloquin, Benjamin N & Mike Frank. 2016. Determining the alternatives for scalar implicature. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 38, .

Potter, Mary C. 2018. Rapid serial visual presentation (RSVP): A method for studying language processing. In *New methods in reading comprehension research*, 91–118. Routledge.

Ronai, Eszter & Ming Xiang. 2023. Tracking the activation of scalar alternatives with semantic priming. In *Experiments in Linguistic Meaning*, vol. 2, 229–240. Https://doi.org/10.3765/elm.2.5371.

Sauerland, Uli. 2012. The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass* 6(1). 36–49.

Skordos, Dimitrios & Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153. 6–18.

Spalek, Katharina, Nicole Gotzner & Isabell Wartenburger. 2014. Not only the apples: Focus sensitive particles improve memory for information-structural alternatives. *Journal of Memory and Language* 70. 68–84.